

**This dissertation has been
microfilmed exactly as received 68-13,555**

**CROWLEY, Leon Vernon, 1936-
AN INVESTIGATION OF THE MARKING PRACTICES
IN SECONDARY SCHOOLS.**

**The University of Oklahoma, Ed.D., 1968
Education, administration**

University Microfilms, Inc., Ann Arbor, Michigan

THE UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

AN INVESTIGATION OF THE MARKING PRACTICES

IN SECONDARY SCHOOLS

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

DOCTOR OF EDUCATION

BY

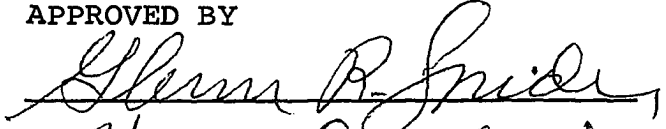
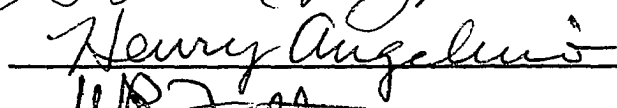
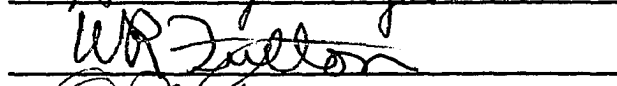
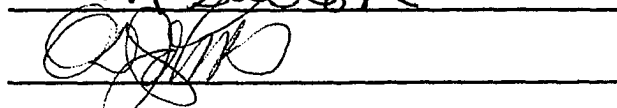
LEON V. CROWLEY

Norman, Oklahoma

1968

AN INVESTIGATION OF THE MARKING PRACTICES
IN SECONDARY SCHOOLS

APPROVED BY

DISSERTATION COMMITTEE

ACKNOWLEDGMENTS

The writer wishes to express his sincere appreciation to Dr. Glenn R. Snider, the Chairman of this doctoral committee, for his assistance and encouragement during the writing of this dissertation and throughout the doctoral program; to Dr. Henry Angelino, Dr. William R. Fulton, and Dr. O. D. Johns for their assistance as members of the committee.

He also wishes to express his gratitude to Dr. Bill Lillard and Dr. Jesse Lindley of the Oklahoma City Public Schools, for their interest, encouragement, and for providing the facilities which made this study possible.

A special note of thanks is due Severlan Vaughn for his contribution toward the completion of this study.

Sincere gratitude is extended to the writer's parents for their encouragement and assistance throughout the undergraduate years.

And finally, to my wife, Jayne, whose understanding and willingness to sacrifice made this day possible, I shall always be grateful and indebted.

TABLE OF CONTENTS

LIST OF TABLES	Page vi
Chapter	
I. INTRODUCTION	1
Background and Need for the Study	1
Purpose of the Study	10
Statement of the Problem	11
Hypotheses to be Tested	12
Statistical Analysis	13
Delimitations	14
Definition of Terms	15
Population	16
Method of Study	17
Procedure	17
Development and Validation of the Instrument	18
Organization of the Study	21
II. REVIEW OF RELATED LITERATURE	22
Period From 1910-1920	22
Period From 1920-1930	24
Period From 1930-1940	26
Period From 1940-1950	28
Period From 1950 to Present	30
Purposes of Marking	31
Basis of Marks	32
Factors Involved in Assigning Marks	33
Variability and Reliability of Marks	36
Improving Marking Practices	43
Marking Systems	46
Summary of Review of Literature	50

	Page
III. DESIGN AND PROCEDURE	53
Design of the Study	53
Procedure of the Study	56
Development and Validation of the Instrument	57
Percent of Returns in this Study and Related Literature	61
Statistical Methods Used	62
IV. PRESENTATION AND ANALYSIS OF DATA	65
An Analysis of the Teacher Sample	66
An Analysis of the Questionnaire Items ..	74
V. SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS ...	158
Summary of Findings	159
Classification Data	159
Questionnaire Data	160
Conclusions	166
Recommendations	169
BIBLIOGRAPHY	171
APPENDIX	179

LIST OF TABLES

Table	Page
1. Secondary School Teaching Level	67
2. Sex of Respondents	68
3. Number of College Level Courses Taken by Teachers in Testing, Measurement and Evaluation	69
4. Highest Degree Held by Teachers	71
5. Respondents' Years of Teaching Experience	72
6. Age Category of Respondents	73
7. Primary Basis Used in Evaluating Achievement ..	75
8. Consideration Given to Judgment of Character and Citizenship	78
9. Consideration Given to Improvement	80
10. Consideration Given to Achievement	84
11. Consideration Given to Educational Evaluation .	87
12. Consideration Given to Absolute Measurement ...	89
13. Consideration Given to Objectives That Are Stated in Terms of Desired Behavior Outcomes .	93
14. Consideration Given to the Achievement of Course Objectives	96
15. Consideration Given to the Variation of Individual Pupil Ability	100

Table	Page
16. Consideration Given to Classroom Behavior	102
17. Consideration Given to Pupil Effort	104
18. Consideration Given to Unexcused Absences and Tardiness	107
19. Consideration Given to Quiz Scores	109
20. Consideration Given to the Place of Homework in Determining Grades	111
21. Consideration Given to Test Scores	113
22. Consideration Given to the Use of Marks as Persuasive Devices	116
23. Consideration Given to Classroom Performance .	119
24. Consideration Given to Work Done in a Previous Grading Period	123
25. Frequency With Which a Grade is Recorded in the Class Record Book	125
26. Consideration Given to Final Examinations	128
27. Frequency With Which Homework is Graded for Quality	130
28. Frequency With Which Short Quizzes are Administered	134
29. Frequency With Which Longer Tests are Administered	137
30. Percent of Students With Which Grades Were Discussed on an Individual Basis	139
31. Consideration Given to Homogeneous Grouping ..	141
32. Consideration Given to Special Weighting Procedures	144

Table	Page
33. Methods Used When Determining the Distribution of Student Grades	146
34. Formulas Used When Weighting Such Factors as Homework, Tests, Projects, Etc.	149
35. Teacher Attitudes Toward School Grades	151
36. The Attitude Held by Students Toward Grades Which is Most Conducive to Deriving the Maximum Benefit From the Teachers' Subject	154
37. Importance of Evaluation as a Part of the Job of the Teacher	156

AN INVESTIGATION OF THE MARKING PRACTICES
IN SECONDARY SCHOOLS

CHAPTER I

INTRODUCTION

Background and Need for the Study

The focus in public school education today is on curriculum improvement and on a concomitant extension of educational services to the end that every student will enjoy a full opportunity to achieve maximum intellectual and/or vocational success. Since 1957, various crash programs have been devised to improve education. Although emphasis appears initially to have been placed on secondary education, the impact of these programs has also been felt in the elementary schools.

The curriculum improvement effort is a result of conscious endeavor on the part of the educators to create more effective organizations for the attainment of educational objectives. Educators today are motivated by a level

of thought and a degree of perception into human and social needs as never before in the history of American Education. The pursuit of equality of educational opportunity is an eager pursuit and a realistic objective of many alert school systems.

With the secondary school curriculum in a state of transition, many problems at the local level face the administrator as he begins to implement programs designed in accord with the pursuit of equality of educational opportunity. How can course content be reorganized? What are the most effective grouping procedures to adopt? How shall experimental work, as well as the effectiveness of basic program change be evaluated? How shall student achievement be evaluated?

The problem of establishing a consistent and uniform grading procedure has been persistently troublesome at all levels of education. Month after month, articles appear in popular magazines and professional journals criticizing current practices or suggesting some new approach. Colleges of education, public schools, and private schools are constantly experimenting with new systems of marking, or sometimes of not marking. And the problem seems to remain.

One of the reasons why marking is a difficult problem is suggested in the following explanation by Thorndike:

Measurement which involves human capacities and acts are subject to special difficulties due chiefly to:

- (1) The absence or imperfection of units in which to measure.
- (2) The lack of constancy in the facts to be measured, and
- (3) The extreme complexity of the measurements to be made.¹

A second reason why problems of marking are difficult according to Ebel, is because marking systems tend to become issues in educational controversies. Ebel noted in his research the following:

The rise of progressive education in the third or fourth decade of this century, with its emphasis on the uniqueness of the individual, the wholeness of his mental life, freedom and democracy in the classroom, and the child's need for loving reassurance, led to criticisms of the common standards of achievement for all pupils, implicit in many marking systems. In the sixth and seventh decades renewed emphasis on what is called "basic education" and on the pursuit of academic excellence, has been accomplished by pleas for more formal evaluations of achievement and more rigorous standards of attainment.²

A third reason why marking systems present problems is that since 1957 the dramatic change in learning theory, the constant change in the content and structure of the curriculum, and the gradual refinement of educational objectives have forced educators to seek and create new approaches

¹E. L. Thorndike, An Introduction to the Theory of Mental and Social Measurement, 2nd ed.; (New York: Teachers College, Columbia University, 1913), Chapter 2.

²Robert L. Ebel, Measuring Educational Achievement, (New Jersey: Prentice-Hall, Inc., 1965), p. 397.

for the evaluation of student achievement.

Most teachers and administrators at all levels of education seem to agree that although marks are out of step with present day concepts of education, they are indispensable tools in the present structure of our educational system. Justifications for marks in secondary schools range all the way from keeping parents informed as to what is going on in school to the rather superficial need for a procedure whereby those pupils who are "eligible" to participate may be identified. Such justification lends itself to superficiality and a much more fundamental basis for marking should be sought.

In an article entitled, "Evaluating and Reporting Pupil Progress," Rothney made the following observation about the importance of marks:

Marks are the coin of the school realm. They continue to be the measure of school success - the keys that open doors of educational institutions for entrance and for exit. The marks given by some classroom teachers have great value. Others may mean nothing more than that the student has official permission to forget what he has learned. In either case they seem likely to continue to be the principal basis for honor awards, promotions and placement in school. For a long time to come parents will accept them as the basic evaluative device.¹

¹John W. M. Rothney, "Evaluating and Reporting Pupil Progress," Research Bulletin of the National Education Association, VII (January, 1956), p. 4.

So common has the practice of giving marks become that it is universally accepted as necessary. The importance and complexity of this practice was aptly stated by Johnson:

Parents the country over have been educated in a system in which marks have been emphasized. To parents, marks are virtually pay checks for their efforts as well as those of the children. Marks are the passports from one grade to another, from one subject to the next. They determine rank in class. They have a great bearing upon whether one goes to college or is admitted to the college of his choice. All in all, marks play a most important role in the general scholastic scheme.¹

Concerned by the increasing importance of marks and rank to the welfare of secondary school students, the National Association of Secondary School Principals employed Terwilliger to make a survey of marking practices. Commenting on the importance of marks Terwilliger had this to say:

Larger graduating classes accompanied by a smaller increase in number of college and work opportunities have created imbalances that have had the effect of making marks and rank more important than they have ever been.²

In view of the above evidence it might be concluded that marks, like test scores, being quantitative, lend

¹L. R. Johnson, "Are There Better Ways of Evaluating, Recording, and Reporting Pupil Progress in the Junior and Senior High Schools?" The Bulletin of the National Association of Secondary School Principals, XXXIV (March, 1950), p. 75.

²James S. Terwilliger, "Self Reported Marking Practices and Policies in Public Secondary Schools," The Bulletin of the National Association of Secondary School Principals, I (March, 1966), p. 3.

themselves to use for quick and economical decision making by admission officers and personnel managers. The terms "top tenth" and "top half" have spine chilling meanings to students, parents and educators.

Whether all this is good or bad may be argued at great length. The argument, however, is beside the point of this inquiry. The fact remains that parents expect a "report" from the school as to how well the pupil is doing. Schools have used a shorthand form of reporting in order to reduce clerical time and teacher effort. The schools send home letter marks, or percentile marks, or some other graphical representation which symbolizes a relative degree of successful behavior or accomplishment to the average parent. The importance and complexity of pupil marks was aptly stated by Austin when he said:

If we change, the plan is very upsetting.
If we abandon the plan, it is devastating.
We may change the symbols but we had better
keep reporting. Society expects it.¹

As dramatic changes in the content and structure of the secondary curriculum have taken place, it becomes apparent that the efforts of teachers in grading student

¹David B. Austin, "The Need for Marking Systems at the Secondary School Level," The High School Journal, XXXVI (April, 1953), p. 194.

achievement must be coordinated. Variability in the marking practices of individual teachers is of great concern to many educators since it leads to a basic problem. This problem is reflected in the question, "What generality, if any, is there in the meaning of marks?"

Recent studies¹ seem to indicate that individual schools have attacked the problem independently of each other. Thus while some degree of consistency within each school was achieved, there still existed variability in the marking practices of teachers in secondary schools. Commenting on the meaning of marks Ahmann and Glock, in their book on Evaluating Pupil Growth, made the following point:

What does a B in general science mean? Does it show that the pupil did A- work in quizzes, C- work in laboratory, B- work in class participation? Or, more important still, does it mean that the pupil achieved it on A- level with respect to one educational objective, a B- level with respect to a third. We have no way of knowing. If marks are used, there must be additional data in the report that provides diagnostic information.

The mark is also influenced by other factors, one of which is the nature of the pupil population. Too often pupils who have been graduated magna cum laude from their high school find to their dismay that they do not have exceptional ability when they face competition in the university. Another factor is the peculiarity of the individual teacher who does the marking. Some teachers give consistently high marks while other teachers pride themselves in never being so easy as to

¹Terwilliger, op. cit., p. 4.

award an A or a 95.

Marks, therefore, are far from being as meaningful as many people think they are. They are being over-interpreted; they reveal far less about the pupil than commonly supposed, and their meaning is ambiguous.¹

Why does it matter that marking practices vary widely, even within the same schools? Terwilliger, an authority on marking and grading, cited the following reasons:

It matters because marks are supposed to be a common currency among schools and colleges, a common language. We should be able to trust that an "A" is an "A" is an "A."

It matters because thousands of young people fail in higher education every year, and a substantial percentage of these failures are due to faulty selection - faulty because the screening devices and selection procedures are inaccurate, and marking and ranking practices so varied. Every failure is a loss to an individual and family, a loss to the college and society. Every student who failed held the place of a boy or girl who might have succeeded.

It matters because we, in the nation's secondary schools, are in a large part responsible for the unnecessary heartbreaks, the unnecessary loss to our society.²

The preceding reference seems to indicate that variability in the marking practices of individual teachers can have a far reaching impact on the student both in the establishment of a self-image on a spectrum between success

¹J. Stanley Ahmann and Marvin D. Glock, Evaluating Pupil Growth, (Boston: Allyn and Bacon, 1963), pp. 566-567.

²Terwilliger, op. cit., p. 4.

and failure and in providing a predictive factor for success at the next level of instruction. Variability in the assessment of marks can limit or extend future educational opportunities as well as encourage or discourage individual initiative.

The numerous shortcomings of our current marking practices does not eliminate the basic fact that we have marks and will no doubt continue to have marks in most of our secondary schools for a long time to come. They are the best means we have found so far to quickly and succinctly report a large number of facts concerning a large number of young people to a large number of parents.

Promise of future improvements in marking practices may be found in support of continuing analysis. Wrinkle¹ stressed the value of a periodic review of any marking system by the student it affects. An attitude favoring continuous reassessment of the marking system is of considerable value during a period when pressures on education have considerable impact on the purposes of marking. The importance of marks to college admission and work opportunities have heightened the concern of students and parents as greater

¹William L. Wrinkle, Improving Marking and Reporting Practices, (New York: Rinehart & Company, Inc., 1947), p. 30.

numbers of high school graduates find themselves excluded from the college or the job of their choice. Under these pressures, school marking practices have come under searching scrutiny.

The article on "Marks and Marking" in the latest edition of Encyclopedia of Educational Research closes with this observation:

Although research has uncovered some limitations and suggested some promising directions in marking procedures, no commonly accepted system has emerged from a half a century of inquiry. Perhaps the development of such a system awaits wider agreements on the goals of instruction and the purposes of marking.¹

Certainly, the rationale for improving marking practices is the student himself. Educators make decisions every day in which grades play a crucial part. These are not trivial decisions. The marking practices of our teachers must provide an evaluation of student achievement which will allow such decisions to be made as fairly and as intelligently as possible.

Purpose of the Study

The purpose of this study was to investigate the marking practices of secondary school teachers and to explore

¹"Mark and Marking Systems," Encyclopedia of Educational Research, ed. Chester W. Harris, (3rd ed: New York: The Macmillan Company, 1960), p. 789.

the sources of variability in those marking practices. This investigation attempted to collect data that would give a clear picture of teacher marking practices in selected academic areas and also of the diversity and agreement which exists in the marking practices of individual teachers. Two basic questions which this study attempted to answer were, "What are the marking practices of secondary teachers?" and "What are the major sources of variability in the marking practices of secondary teachers?"

An investigation of marking practices and their possible sources of variation in the secondary schools would be extremely useful to school administrators in planning a more efficient organization for the attainment and the evaluation of educational objectives. School counselors and teachers should find the results of the study helpful in their efforts to provide consistent and equitable approaches to the evaluation of student achievement.

Statement of the Problem

The problem of this study was to determine the marking practices of teachers in the secondary schools of Oklahoma City during the school year 1966-67 and to analyze the

sources of variability in those marking practices. More specifically, it was intended to:

1. Develop an instrument which would determine the marking practices used by individual teachers in the assessment of pupil marks.
2. Analyze both the variability and agreement which exists in the practices used by individual teachers in arriving at or assigning marks to pupils.
3. Ascertain the selected variables affecting the marking practices used by individual teachers in arriving at or assigning marks to pupils.

Hypotheses to be Tested

HO₁ There is no statistically significant difference, other than what might occur by chance, among teachers of the four academic areas of English language arts, mathematics, science, and social studies with regard to marking practices.

HO₂ There is no statistically significant difference, other than what might occur by chance, between probationary teachers and nonprobationary teachers with regard to marking practices.

HO₃ There is no statistically significant difference, other than what might occur by chance, in the marking practices

of probationary and nonprobationary teachers among the four academic areas of English language arts, mathematics, science, and social studies.

Statistical Analysis

All items in the instrument were expressed in descriptive units on a continuum. The descriptive units in each item were assigned arbitrary code numbers. Each item in the instrument was treated by the use of analysis of variance - multiple classification. Analysis of variance - multiple classification is a statistical test of significance of the difference between two or more groups simultaneously.

Wert, Neidt and Ahmann pointed out the appropriateness of this statistical treatment for data such as those to be obtained in this study.

When designing experiments in education and psychological research, the possibility of classifying the data in more than one manner is invariably considered. Indeed, it is often not only possible but also advisable to design studies so that results permit the testing of hypotheses concerning separate subdivisions of the data. It is in keeping with efficient experimental methods to incorporate logical multiple classification in such research problems.

In addition, the subdivision of the data into two or more classifications allows the investigator to control certain characteristics known to influence the result of the experiment, or perhaps only suspected of such influences. In this manner possible sources of bias can be controlled and the demands of sound

experimental design for experiments free of bias can be met, at least in part, by meaningful classification of the data. Such classification also sensitizes the test of significance by enabling the investigator to identify more of the sources of variation in his investigation.¹

After coding the data and punching the information on IBM cards, statistical computations were performed on the IBM 1401 Computer. The procedures used in programming the statistical analysis were those suggested by IBM Fortran IV.

Delimitations

This study was limited to teachers of secondary schools in the Oklahoma City Public School System. The study was further limited to teachers of the following subjects: English language arts, social studies, mathematics, and science. It was believed that to investigate the marking practices of secondary teachers, the subjects normally viewed as the heart of the academic program of the school would yield more reliable data than could be obtained through a broad general survey of all subject matter areas.

¹James E. Wert, Charles O. Neidt, and J. Stanley Ahmann, Statistical Methods in Educational and Psychological Research, (New York: Appleton-Century-Crafts, Inc., 1954), p. 188.

Definition of Terms

Educational Evaluation. A process in which a teacher commonly uses information derived from many sources in order to arrive at a value judgment.

Grade. The evaluation that a teacher makes of pupil progress or achievement. Used synonymously with marks.

Secondary Schools. A school division following the elementary school which would include the following subdivisions:

1. Junior High School. A school that enrolls pupils in grades 7, 8, and 9 or grades 7 and 8.
2. High School - Six Year. A secondary school that incorporates six grades, 7 to 12, administered under one principal and having a faculty organized to serve all six grades.
3. Senior High School. A secondary school composed of the upper high school grades, 10 to 12 or 9 to 12.

Marking Practices. The methods used by the individual classroom teacher in arriving at or assigning marks to students.

Marks. The evaluation that a teacher makes of pupil progress or achievement. Used synonymously with grade.

Probationary Teacher. A teacher with less than three years of teaching experience.

Nonprobationary Teacher. A teacher with more than three years of presumed successful teaching experience.

Specific Variable. A statistical study of a population as to probationary and nonprobationary teachers.

Population

The population of this study was composed of all teachers in the secondary schools of Oklahoma City Public Schools who were teaching in the English language arts, social studies, mathematics and science fields for the 1966-67 school year.

The sample was drawn from the total population of 635 utilizing accepted methods for selecting random samples as described by Van Dalen.¹ From the total population of 635 teachers, a sample size of 160 was chosen according to procedures as outlined by Cella.² This technique gave a sample small enough to study in depth and large enough to

¹Deobold B. Van Dalen, Understanding Educational Research, (New York: McGraw-Hill Book Company, Inc., 1962), pp. 249-54.

²Francis R. Cella, Sampling Statistics in Business and Economics, (Norman, Oklahoma: Bureau of Business Research, University of Oklahoma, 1950), pp. 160-63.

be representative of the total population with a sampling error of not more than 10 percent and a 95 percent confidence interval.

The sample used in this study was designed to be a miniature of the population from whence it was selected and presumed to embody the characteristics of the total population defined in the problem.

Method of Study

An experimental design utilizing the action method of research was used in this study. Data were gathered through the use of the questionnaire. Action research is defined by Goode and Hatt as "part of a program aimed at changing existing conditions."¹ The questionnaire, also defined by Goode and Hatt, is "a device for securing answers to questions by using a form which the respondent fills in himself."²

Procedure

This study was developed through the following steps:

1. The literature was surveyed in the area of marking

¹William J. Goode and Paul K. Hatt, Methods in Special Research, (New York: McGraw-Hill Book Company, Inc., 1952), p. 326.

²Goode and Hatt, op. cit., p. 133.

and grading.

2. A questionnaire was developed utilizing in part the suggestions of teachers, administrators and the university advisor.
3. The questionnaire was validated through the personal interview technique and checking questionnaire responses against actual behavior.
4. The questionnaire was mailed to all teachers selected by the random sample.
5. The data from the questionnaire were statistically analyzed and interpreted.
6. Conclusions and recommendations for the development of a consistent and equitable approach to the evaluation of student growth and achievement in the public secondary schools of Oklahoma City were made.

Development and Validation of the Instrument

The instrument used in this study was developed with the close collaboration of teachers, administrators, and the university advisor. The purpose of the questionnaire was to ascertain as objectively as possible the evaluation procedures and practices teachers use in determining a pupil's final grade.

The instrument was developed according to procedures as outlined by Rummel¹ in his book, An Introduction to Research Procedures in Education. After the instrument had been developed, an objective evaluation and comparison of the questionnaire was made using the criteria as set down by Mouly² in his book, The Science of Educational Research. Mouly³ pointed out that the questionnaire can give reliable and valid information if certain requirements are met in its formulation.

The questionnaire was substantially revised many times before the first pre-test study. The pre-test questionnaire was submitted to twenty teachers of the same subject matter area as those who were eventually to receive the final draft. These teachers were to fill out the questionnaire and to indicate their reactions to every phase of its organization. The outcome of the pre-test helped determine flaws in certain items that the investigator was too close to see.

¹J. Francis Rummel, An Introduction to Research Procedures in Education, (New York: Harper and Row, Publishers, 1958), pp. 126-27.

²George J. Mouly, The Science of Educational Research, (New York: American Book Company, 1963), p. 263.

³Ibid., p. 263.

Although an objective evaluation and comparison, to which the questionnaire as an instrument of science must subscribe, had been made, there remained the task of identifying the specific ways in which the validity would be established.

The actual validation of the questionnaire came in the second phase. A more adequate validation required checking the responses which the questionnaire elicited against an external criterion. This was accomplished by following the questionnaire with a personal interview of a 20 percent sample of the respondents to see whether their responses to the questionnaire actually represented their views on the subject discussed.

A further effort was made to validate the instrument by checking questionnaire responses against actual behavior. This was accomplished through personal observation and interviews in selected classrooms for the second phase of validation. In general, the observations made in the classrooms, showed that the marking practices being used followed the same marking practices as given by the teachers in the interviews and on the answered questionnaire.

Organization of the Study

The problem of this study is presented in Chapter I. Chapter II is devoted to a review of pertinent literature related to the study. Chapter III is a detailed discussion of the construction of the instrument used and procedures followed in this study. Chapter IV is a report and analysis of the data secured from the questionnaires. Chapter V is composed of a summary of the findings, some conclusions reached from the study, and recommendations for the development of a consistent and equitable approach to the evaluation of student growth and achievement in the public secondary schools of Oklahoma City.

CHAPTER II

REVIEW OF RELATED LITERATURE

One of the most important aspects of any investigation is the examination of the literature and research related to the problem. An examination and analysis of the writing in the field provides not only the history of developments but also a broad panorama of the field of thinking regarding the subject.

Two phases stand out in bold relief in a half century of research studies on school marks; (a) the period extending from 1910 to 1940, when research was focused mainly on the mechanical and semantic problems of marking; and (b) the period from 1940 to the present, during which a greater interest has centered on improvement of marks in comprehensiveness and communication.

The period from 1910 to 1920

Many of the studies reported during this period dealt with such matters as the variability and unreliability of teacher marks.

Starch and Elliott¹, in their well known study, described wide discrepancies in standards of marking in history and mathematics.

Rugg², writing on teacher marks and marking systems in 1915, showed that individual teachers were setting their own standards and this led to variability, unreliability and inconsistency in arriving at pupil marks. Other writers during this period, such as Dearborn³ and Kelly⁴, gave added evidence of the unreliability and variability of teacher marks.

During the period 1910-1920, school marks were almost wholly reported in percentages with a trend toward the use of a three to seven point system. Rugg⁵ reported that

¹Daniel Starch and E. C. Elliott, "Reliability of Grading Work in History," School Review, XXI (December, 1913), pp. 676-81: "Reliability of Grading Work in Mathematics," School Review, XXI (April, 1913), pp. 254-59.

²Harold Rugg, "Teacher Marks and Marking Systems," Educational Administration and Supervision, VII (January, 1915), pp. 117-42.

³Walter Dearborn, "Marks and Marking Systems," Encyclopedia of Educational Research, ed. Chester W. Harris (3rd ed: New York: The Macmillan Company, 1960), p. 784.

⁴Frederick Kelly, "Marks and Marking Systems," Encyclopedia of Educational Research, ed. Chester W. Harris (3rd ed: New York: The Macmillan Company, 1960), p. 784.

⁵Rugg, op. cit., p. 117.

educators favored a system using from three to seven letters and the adoption of a standard marking system.

Dearborn¹, in 1910, was one of the first to support the idea of using normal probability curve combined with relative ranking to arrive at a mark. In the same year Meyer² established that the use of the concept of the normal distribution resulted in the reduction of the variability of teacher's marks.

The scientific measurement movement and the widespread use of the Army Alpha and Army Beta test during this period had a profound impact not only on testing but on the identification of ability. This concept of measurable differences in individuals had a dramatic effect on marking practices.

The Period from 1920-1930

Prior to 1920, school marks were almost wholly reported in percentages. Beginning with the measurement movement which characterized the 1920's, the shift was made

¹Dearborn, op. cit., p. 784.

²Max Meyer, "Marks and Marking Systems," Encyclopedia of Educational Research, ed. Chester W. Harris (3rd ed: New York: The Macmillan Company, 1960), p. 784.

slowly to a better system, varying from three to seven symbols. Roelf¹, in a 1925 study of junior high schools, found that 23 percent of the schools were using percent marks. Twenty-eight percent of the schools surveyed used symbols expressed or percents, and symbols expressed as descriptive words or phrases were used by 49 percent of the schools.

During the 1920's, efforts were begun to broaden the base of marking systems to include some recognition of the learners efforts to succeed. But, while the modification in practice was gaining acceptance, research studies sturdily defended the justification of distributing marks as the normal curve. The case for use of the normal distribution curve in assigning marks was supported by Thorndike and Bregman² in their study of ninth grade intelligence.

One of the most controversial features of grading at this time was increased use of standardized intelligence tests which gave impetus to individualization of instruction

¹R. M. Roelf, "Trends in Junior High School Reporting," Journal of Educational Research, IXL (December, 1955), pp. 241-49.

²E. L. Thorndike and E. O. Bregman, "On The Form of The Distribution of Intellect in the Ninth Grade," Journal of Educational Research, X (November, 1924), pp. 271-78.

and ability grouping. Researchers introduced the problem of whether or not a full range of marks should be used in each ability section. Ruch¹ suggested that A to C marks should be used for superior grouped classes, B to D marks for average classes, C to F marks for slow classes. In contrast to this, a Los Angeles Committee² recommended a full range of marks in each section.

The Period from 1930 to 1940

The early 1930's was the time of re-examination of marking systems and a recognition that the extensive use of standardized achievement test often overlooked the importance of personal and social development. Williamson³, reported that the trouble was not with the test but with the misuse of these tests and the failure of teachers to define goals.

¹Giles M. Ruch, The Objective or New Type Examination; An Introduction to Educational Measurement, (New York: Scott, Foresman and Company, 1929), p. 478.

²Los Angeles, Committee of the Secondary School Principal's Association, "Marking Slow Pupils," California Quarterly of Secondary Education, I (June, 1926), pp. 386-91.

³E. G. Williamson, "The Cooperative Guidance Program," School Review, XLIII (April, 1935), pp. 273-80.

Drake¹, concerned with the effects of testing on instruction, suggested that there must be a willingness to sacrifice some of the accuracy possible in tests of achievement in order to provide some measure of more intangible behaviors.

Some of the still unsolved problems relating to pupil marking were evident in the literature of this period. Opinions in the 30's favored use of achievement as the basis for academic marks, with other behaviors rated separately. Crooks² found that teachers intended marks to show actual achievement, but that many other factors entered in. Odell³ reported that most educators agreed that absolute achievement alone should be reflected by marks.

It was during this period of time that Hick⁴ identified the principal difficulties involved in the improvement

¹C. E. Drake, "Trends in the Field of Evaluating Secondary Education," Educational Administration and Supervision, XXVI (April, 1940), pp. 241-56.

²A. B. Crooks, "Marks and Marking System: A Digest;" Journal of Educational Research, XXVII (December, 1933), pp. 259-72.

³Charles Odell, Educational Measurements in High Schools, (New York: Century Co., 1930), p. 641.

⁴Arch Hick, "Contributions of Research to the Classification, Promotion Marking and Certification of Pupils," Thirty Seventh Yearbook of the National Society for the Study of Education, Part II (Bloomington, Ill.: Public School Publishing Company, 1938), pp. 187-99.

of marking as being the multiplicity of factors in the situation, the subjective nature of evaluative methods, and the variability of teachers' standards. Wakehom¹ also reported the importance of setting up a marking system which was impartial. Campbell² defined standards for each symbol and recommended specific definitions for the meaning of each mark.

The Period from 1940 to 1950

It was during this time that the full force of the complexity of the problem struck home. The deeper perceptions and understandings of the complex phenomena of human growth and development caused less concern in research with the mechanical aspects of marking and created more concern with the purposes of marking and their relation to educational objectives.

The new concept of evaluation, based on the development of the whole child, called for appraisal in many areas of student growth. Traxler³ reported evidence of the

¹Glen Wakehom, "Humanizing Grades," School and Society, XXXIV (October, 1931), pp. 596-98.

²Laurence Campbell, "So Pupils May Know," School and Society, XXXII (December, 1930), pp. 762-63.

³Arthur Traxler, Techniques of Guidance, (New York: Harper and Brothers, 1957), p. 394.

general practice of attempting to evaluate much more than subject matter achievement. Wrinkle¹ indicated the greatest weakness in marking practices was the fallacious belief that one single mark could reflect the level of student achievement. Rothney² has mentioned the inadequacy of a single mark to summarize the appraisal of any area of student achievement.

Burton³, in his solution to the problem, recommended that the language used to describe the desired behavior be definite enough to set up an outcome which could be evaluated. Traxler⁴ cited a trend toward rating several specific behaviors for each course rather than assigning a single mark.

During this period, increased cooperation between the school and the public was advised by many authors. The

¹Wrinkle, op. cit., p. 120.

²W. M. John Rothney, Evaluating and Reporting Pupil Progress: What Research Says to the Teacher, National Education Association Bulletin No. 7, (Washington, D. C., 1955), p. 33.

³William H. Burton, The Guidance of Learning Activities, (New York: Appleton-Century Company, 1944), p. 601.

⁴Traxler, op. cit., p. 394.

cooperation extended the search for more satisfactory marking practices. Peterson¹ reported how teacher-student-parent cooperation resulted in the improvement of a new marking plan. Wrinkle² advised having students participate in the development of new marking policies and practices and the appraisal of existing ones.

There was concern shown for the possible adverse effect of pupil marking. Smith³ reported that many teachers thought the competitive features of marks were disturbing and often harmful to students. Wrinkle⁴ believed that if the competition aspect of marks could be eliminated, the teachers would be forced to rely on more acceptable methods of instruction.

The Period from 1950 to Present

The research in this period was concerned mainly with: Purposes of Marking; Basis of Marks; Factors Involved

¹Russell S. Peterson, "Marking Achievement in Relation to Ability," National Association of Secondary School Principals Bulletin, XXXVIII, (February, 1954), p. 63.

²Wrinkle, op. cit., p. 120.

³Eugene R. Smith, Appraising and Recording Student Progress, (New York: Harper and Brothers, 1942), p. 550.

⁴Wrinkle, op. cit., p. 120.

in Assigning Marks; Variability and Reliability of Marks; Improving Marking Practices; and Marking Systems.

Purposes of Marking

Continued interest was shown during this period in the relationship between the marking system and the educational philosophy. Morris¹, concerned with the purposes of marking, revealed that new developments in educational philosophy had changed the purposes of marking from the recording of results to increasing the effectiveness of student learning.

The above approach to the purposes of marking did not eliminate but rather tended to increase the motivational, guidance, informational and administrative purposes of marking. Ahmann and Glock² noted that the purpose of marks can best be defined in terms of those who use them - pupils, parents, teachers, school administrators and employers. In the final analysis Ahmann and Glock³ state that marks should serve one purpose; "to facilitate the educational

¹Lucile Morris, "Evaluating and Reporting Pupil Progress," Elementary School Journal, LIII (September, 1958), p. 532.

²J. Stanley Ahmann and Marvin D. Glock, Evaluating Pupil Growth, (Boston: Allyn and Bacon, 1958), p. 532.

³Ibid., p. 532.

development of each pupil in relation to his or her ability."

Basis of Marks

During this period, the age old question of the basis for marks was still being asked. Liggitt¹ declared that a basic unanswered question was whether the mark was based on absolute achievement or achievement in light of ability. The National Commission of Life Adjustment Education for Youth² recommended multiple appraisal with a mark related to the performance of the group. Traxler³, in a study that contrasted parental attitudes in 1957 with attitudes in 1947, reported some dissatisfaction with marking that related to ability alone and gave no comparison with the peer group. Concerned with the primary basis for marks, Ahmann and Glock had this to say:

Education based on final achievement alone does not tell the whole story of the pupil's development.

¹William L. Liggitt, "Are There Better Ways of Evaluating, Researching, and Reporting Pupil Progress in the Junior and Senior High Schools?" National Association of Secondary School Principals Bulletin, XXXIV (March, 1959), pp. 79-89.

²National Commission on Life Adjustment Education for Youth, Pupil Appraisal Practices in Secondary Schools, U. S. Office of Education No. 363 GPO (Washington, D. C.: 1952), p. 111.

³Traxler, op. cit., p. 394.

Achievement scores, for example, do not indicate whether he is at a particular level because of lack of effort or lack of talent. In the secondary school, when the pupil is looking forward to the roles he will eventually play in higher education and in professional life the teacher must be concerned with standards of achievement as well as growth in evaluating his competency.¹

Considerable criticism, during this period was directed against the competitive basis for marking. Fine² reported his observations of a definite trend away from competitive marking and toward marking related to the student's ability. However, Vredevoe³ cited numerous studies at the high school level in which guidance counselors preferred to retain the traditional marking system because of the weight that marks have in determining college admission.

Factors Involved in Assigning Marks

An important facet of marking practices, as indicated by the literature, was the delineation of the factors to be considered in the computation of a final grade or mark.

¹Ahmann and Glock, op. cit., p. 568.

²Benjamin Fine, "A, B, C of Grading Puzzles Parents," New York Times Magazine, (November 18, 1957), p. 33.

³Lawrence E. Vredevoe, "How May We Make the Recording and Reporting of Pupil Achievement More Meaningful?" National Association of Secondary School Principals Bulletin, XXXVII, (April, 1953), pp. 179-82.

Wallace¹ in answer to the question, "What are the most critical areas in the assignment of marks," cited the following: (1) Use of a marking plan, (2) Consideration of the role of individual differences, (3) Influences of non-scholastic factors including the effect of behavior upon marks, (4) Support given teachers in handling critical situations, and (5) Establishment of lines of communications with parents.

Alpern² pointed out the problems and the dilemma in trying to report reasonable academic progress and also include such items as effort and attitude. He also maintained that a single grade results in vague reports to parents and students. Alpern suggested as a solution to the problem, the adoption of two ratings for each mark, with one rating to show achievement based on accepted standards and the other to emphasize achievement based upon effort and ability to perform.

¹Joseph Wallace, "Critical Incidents in the Assignments of Marks to High School Students," (Unpublished Doctoral Dissertation, Department of Education, University of Southern California, Los Angeles, 1960), p. 48.

²Morton Alpern, "A Fair Grading System," Clearing House, (October, 1960), pp. 113-114.

Quinn and Szuberla¹, in an article titled "Relative Grading, A Consistent and Equitable Approach to the Evaluation of Student Achievement," declared that the first step in the effort to solve the grading problem is a review of the basic philosophy of grading for the school system. The statement of philosophy that should emerge, according to Quinn and Szuberla, is one that advocates that all students be evaluated on the basis of achievement alone, and that the students' achievement should be rated in relation to that of all the students in a grade or course. Grading the student on the basis of his ability or on the basis of some arbitrary standard should be rejected as being unrealistic.

While no agreement was found in the literature regarding the effect of achievement and ability on marking and grading practices, it was conclusive that defensible criteria or standards are needed if a high degree of consistency is to be reached. La Franchi, in an article titled, "High School Marks: Comparative or Individual," submitted the thesis that different types of courses should have different factors contributing to the final mark when he stated:

¹George R. Quinn and Charles A. Szuberla, "A Consistent and Equitable Approach to the Evaluation of Student Achievement" Clearing House, (April, 1963), pp. 490-94.

In elective college preparatory subjects, where alternative courses are offered for non-college preparatory pupils, evaluation should be made largely on an absolute scale with little consideration of varying individual capacities.

In courses required of all pupils, evaluation should consider the capabilities of the individual pupils, and a program should be undertaken to prevent misunderstanding of such marks.

In the case of electives, other than college preparatory, the plan should be adopted to the particular course. In purely skill subjects and in vocational courses, there is probably little need to consider varying individual capacities. For evaluation purposes, such courses would fall in the same classification as elective college preparatory subjects, with an absolute standard of accomplishment. But in other electives, consideration might well be given to the varying capacities of the individuals enrolled. For evaluation purposes, these courses would be classified in the same group as the required courses.

Thus high school courses in any one school could be divided into two groups. Those that will be evaluated and marked on an achievement scale, and those where varying individual capacities will be considered. Such a division, arrived at by the staff of the school, would help in the classification and further study of the problem in reaching a more reasonable solution than now prevails.¹

Variability and Reliability of Marks

During this period, the variability and reliability of marks were still being questioned as they were in 1910.

¹Edward H. La Franchi, "High School Marks: Comparative or Individual?" School Executive, LXXI (July, 1952), pp. 51-54.

Carter¹ summarized research studies by Swenson, Edminston and Laubaugh and indicated that girls generally receive higher marks than boys. He also concluded that marks reflect both intelligence and achievement, and that the sex of the student, with girls favored over boys, is a more important factor than the sex of the teacher. Hadley² found that the most liked students received higher marks than the least liked students. Travers and Gronlund³ found wide differences of the meaning of marks, as well as the weights given to the many factors making up the final mark. In a more recent study, Palmer⁴, an editor for Educational Testing Service, pointed out various teacher systems for marking pupils that were unfair.

Vredevoe⁵, after extensive research, found that most teachers differ in their interpretation of achievement and

¹Robert Carter, "How Invalid Are Marks Assigned by Teachers?" Journal of Educational Psychology, XLIII (April, 1952), pp. 218-28.

²Trevor S. Hadley, "School Mark - Fact or Fancy?" Educational Administration and Supervision, XL (May, 1954), pp. 305-12.

³Robert M. Travers and Norman Gronlund, "Meaning of Marks?" Journal of Higher Education, XXI (October, 1950), pp. 369-74.

⁴Oliver Palmer, "Seven Ways of Grading Dishonestly," English Journal, LI (October, 1962), pp. 464-67.

⁵Vredevoe, op. cit., p. 180.

that the value of the mark differed from school to school and among teachers in the same school.

The lack of clearly defined, uniform basis for marking and standards for the meanings of various marks tend to allow biases to lower the validity of marks. Ebel reported:

Often a student's mark has been influenced by the pleasantness of his manner, his willingness to participate in class discussion, his skill in expressing ideas orally or in writing, or his success in building an image of himself as an eager, capable student. Some of these things should not ordinarily be allowed to influence the mark he receives.¹

The above statement tends to support the accusation of students that accomplishment is not the pure and simple basis on which marks are assigned. As Palmer² noted, some instructors deliberately use high marks as rewards and low marks as punishments for behavior quite unrelated to the attainment of the objectives of instruction in a course.

After surveying the literature, Odell³ concluded that "the usual reliability of semester marks is indicated by a coefficient of from .70 to .80, perhaps even of from .80 to

¹Ebel, op. cit., p. 403.

²Palmer, op. cit., p. 464.

³C. W. Odell, "Marks and Marking Systems," Encyclopedia of Educational Research, ed. Walter S. Monroe, (New York: The Macmillan Company, 1950), pp. 711-17.

.90." With respect to the validity of marks, and on the basis of admittedly indirect and inadequate evidence, he suggested that "the degree of validity as a measure of mastery of subject matter is fairly high, probably on the average at least not below that represented by a coefficient of correlation of .70 and in many cases much higher."

In assessing estimates of reliability and validity, Ebel felt that it is helpful to keep the following two things in mind:

One is that semester marks are based on much more extensive and comprehensive observations of pupil attainments, perhaps as much as or more than eighty hours of observation. One hour of intense "observations" under the controlled conditions of a well standardized test of achievement can yield measurements whose reliability may exceed .90.

The other thing to keep in mind is that a coefficient of correlation even as high as .70 does not reflect very pure and precise measurements of the things to be measured.¹

From the evidence presented by Ebel, it is evident that the summary values for the reliability and validity of semester marks reported by Odell do not suggest utter chaos; they do suggest that considerable room for improvement remains.

¹Ebel, op. cit., p. 404.

Concerned by the increasing importance of marks to students, the National Association of Secondary School Principals employed James Terwilliger to survey secondary administrators and teachers on marking practices and policies. This survey was undertaken to explore two of the major sources of variability in the employment of marks and marking systems in the public secondary schools. These two sources were: (1) School policies affecting the assignment of marks and (2) Practices used by individual teachers. Variability in marking practices and policies lead to a basic problem and this problem is reflected in a question asked by Terwilliger¹, "What generally, if any, is there in the meaning of school marks?" Commenting on his own question, Terwilliger had this to say:

Does an "A" given by Miss Jones in English Literature reflect the same kind of student performance and behavior as an "A" given by Mr. Smith who teaches science? We would, of course, expect the marks given in these two instances to be based upon quite different subject content. However, if marks given in English and Science are to have any comparability in meaning at all, the procedures used by Miss Jones and Mr. Smith in assigning marks must be based upon the same general notions as to what is important and what is not important in the evaluation of a student.²

¹Terwilliger, op. cit., p. 5.

²Ibid., p. 36.

The author does not mean to imply that all teachers should fit their evaluation techniques to a common rigid mold. There will necessarily have to be differences from one subject to the next. Yet, there should be a general set of guidelines which will assist the teacher in a given subject area to plan the evaluation of students so that skills represented are generally the same as those of students with other teachers.

Terwilliger, in his attempt to collect data which would give a clear picture of both the diversity and agreement which exists in the practices of individual teachers and the policies of the individual schools, cited some of those findings which seem to have the greatest significance for teachers and administrators.

A fundamental question in marking is, "What should be the primary basis for assignment of marks?" The administrator responses show that 22 percent of the schools in our sample have no policy governing the basis used. Twenty-seven percent state that grades are based upon absolute standard achievement, 24 percent say that grades reflect achievement with respect to ability and 16 percent state that grades represent achievement with respect to others in the class.

A second fundamental problem concerns the role, if any, such factors as classroom behavior, absence, tardiness, and "effort" should play in the assignment of subject matter grades. Forty percent of our administrators indicated that their schools have policies regulating the reduction of grades for disciplinary or other

reasons. At least 20 percent of the teachers in each subject area report giving behavior and absences for tardiness either "moderate" or "major" consideration and at least 50 percent of the teachers in each field give either "moderate" or "major" consideration to "effort." Yet, there are substantial numbers in certain subjects who give little or no consideration to such factors.

Yet a third major problem area is created by relatively widespread adoption of homogeneous grouping in "tracks." Seventy-five percent of the schools in the sample report that tracks are used in certain subjects where tracks exists. A variation of the same problem is the weighting of grades in tracked subjects for computation of class rank. Only nine percent of the schools in the study report giving special weight to honors and advance placement courses. There is an urgent need for careful study of the situation.

A final concern is the relative consideration which a classroom teacher should give to homework, quizzes, tests, classroom performance and special projects. One might legitimately ask why we should be concerned with such matters. The answer is that the type of evaluation procedure used will determine to a large extent the type of learning which occurs.¹

If the above data suggests much of the evaluation practices presently being done in secondary education, Terwilliger suggests that what is truly needed is a revolution in marking. This means calling into question such cherished notions as the "absolute standard" of achievement, grading "with respect to ability," rewarding or penalizing students for classroom behavior or absences by revising grades. The above data tends to reflect blind adherence to tradition

¹Ibid., pp. 33-34.

much more than what is presently known about sound educational measurement.

Improving Marking Practices

The past twenty years have been marked by an attempt to consolidate the gains made in the prior years of research and experimentation. Recent improvement in marking practices has been characterized by a fundamental change in approach which involves the role of evaluation in the improvement of learning.

Recent emphasis in research has been on the need to identify all important objectives of instruction, to state these objectives clearly and specifically in behavioral terms, and to select or develop the evaluation instruments - marking practices which will provide the most valid information for instructional purposes. The ten year study at the Colorado State College of Education, directed and reported by Wrinkle, is one of the most comprehensive experiments using this approach. As a result of this study, Wrinkle¹ concluded that the first step in any improvement of marking is the identification of objectives, both general and specific, in terms of student behavior.

¹Wrinkle, op. cit., p. 30.

One of the most significant attempts to describe educational objectives in order to facilitate the processes of evaluation has come from Bloom and a committee of college and university examiners who are attempting to develop a classification of educational objectives. In a recent report the committee indicated the need for, and the potential use of, a taxonomy:

It is intended to provide for classification of the goals of our educational system. It is expected to be of general help to all teachers, administrators, professional specialists, and research workers who deal with curricular and evaluation problems. It is especially intended to help them discuss these problems with greater precision. For example, some teachers believe their students should "really understand," others desire their students to "internalize knowledge," still others want their students to "grasp the core or essence" or "comprehend." Do they all mean the same thing? Specifically, what does a student do who "really understands" which he does not do when he does not understand? Through reference to the taxonomy as a set of standard classifications, teachers should be able to define such nebulous terms as those given above. This should facilitate the exchange of information about their curricular developments and evaluation devices. Such interchanges are frequently disappointing now because all too frequently what happens to be common ground between schools disappears on closer examination of the descriptive terms being used.¹

The Taxonomy of Educational Objectives offers teachers a general set of guidelines which will assist them in a given

¹Taxonomy of Educational Objectives, ed. Benjamine Bloom, (New York: Longmans, Green and Company, 1958), p. 1.

subject area to plan the evaluation of students so that skills represented are generally the same as those of students with other teachers. As Terwilliger said, "perhaps all that can be done now is to make Bloom's work the basis for a set of non-technical recommendations which would become generally available to teachers."¹

A systematic approach to improving marking practices should provide for the clarification of learning goals.

Gronlund, comments on this:

Ideally, plans for evaluation are made at the same time instructional plans are formulated. This increases the likelihood that the desired learning outcomes will be clearly defined before instruction begins. Although goals can be developed without special attention to evaluation, they are not apt to be as clear and definite. Planning for evaluation encourages us to describe in precise terms the behaviors we are willing to accept as evidence of learning.²

Despite the planning with which educational objectives are defined and shared with pupils, they can have little influence on learning unless they are developed along with the evaluation procedures used. Cronbach cited the following warning:

What the learner tried depends on his goals. The goals of learning are supposed to be established during

¹Terwilliger, op. cit., p. 36.

²Norman E. Gronlund, Measurement and Evaluation in Teaching, (New York: The Macmillan Company, 1966), p. 362.

planning, but actually the learner's goals depend on what evaluation he anticipates. Goals not reflected in evaluation procedures will be neglected. Progress toward some objectives affects marks; the pupil pays only lip service to other objectives not represented in the evaluation.¹

In summarization, defining objectives in behavioral terms contributes to better instructional planning and more effective evaluation of learning activities. In addition, pupils operate better in a situation in which the evaluation procedures provide them with an operational definition of the objectives to be achieved.

Marking Systems

While there was not complete agreement that the five point letter system of mark reporting was the best or only answer to pupil marking, the literature confirmed that this was by far the most common method of reporting marks. Traxler² noted the continuing trend away from percentage marking toward a scale with fewer points. He also points out that the use of A, B, C, D, F, or their equivalents, was found to be the practice in most secondary schools, with experimentation generally confined to the elementary schools. Roelfs³, in a 1954 study of junior high schools, found that

¹L. J. Cronbach, Educational Psychology, (New York: Harcourt, Brace & World, 1962), p. 542.

²Traxler, op. cit., p. 394.

³Roelfs, op. cit., pp. 241-249.

82 percent used letters or numbers explained by words or phrases and 17 percent used letters or numbers with percentage equivalents. Of these schools, 81 percent used a five point scale, 3 percent used fewer than four steps, and 12 percent used more than five steps. A study by Strang¹ of 126 elementary and secondary schools revealed that 41 schools used "satisfactory" and "unsatisfactory," 56 used A, B, C, D, and 24 used numerical marks.

Barnes and Barnes² were critical of the A to F system if the marks were used to compute only a grade point average. They suggested that relative class rank was a valid criteria of ability and proposed that over a four year period that class rank would do a good rating job. Ahmann and Glock, on the difficulty of assigning marks said, "two important reasons for this are the teacher's uncertainty as to whether an appropriate amount of credit is being given to effort, growth, and final achievement for each educational objective considered, and, since the mark represents pupil progress in terms of a combination of objectives, his uncertainty as to whether each objective within the combination

¹Ruth M. Strang, "Reporting Pupil Progress," School Executive, LXXII (August, 1953), pp. 47-50.

²K. F. Barnes and E. H. Barnes, "A Realistic Approach to Grading," Clearing House, XXXVI (April, 1962), pp. 476-78.

is being appropriately weighted."¹

A variation of the traditional letter-number system is the use of not one but two marks for each subject matter area. The first mark represents the pupils final achievement in relation to some standard and the second represents the amount of effort put forth by the pupils. Commenting on the dual marking system, Ahmann and Glock said, "the dual marking system can be expected to provide a clearer picture of the pupil's progress than does the single mark so often used, but they are still an oversimplified picture of the teacher's evaluation."²

Many educators agree there is considerable merit in the statement by Ahmann and Glock, and under favorable conditions they can improve marking considerably. But they do involve problems. Ebel, cites some of the problems:

For one thing, they multiply considerably the already irksome chores of marking. For another, they create additional problems of defining precisely what is to be marked and of distinguishing clearly among the different aspects of achievement, on which to base a reliable mark. Finally, and largely as a result of the preceding difficulties, the multiple marks exhibit considerable "halo effect."³

¹Ahmann and Glock, op. cit., p. 543.

²Ahmann and Glock, op. cit., p. 545.

³Ebel, op. cit., p. 421.

From the evidence presented, it seemed that multiple marking is not the only road to improvement in marking and probably not the best road currently available.

Another system which has proven to be more useful and at the present is slowly replacing the traditional system is the checklist of behavioral description to be checked or rated. The checklist, which is most common at the elementary school level, typically includes ratings of progress toward the major goals in each subject area. Gronlund had this to say:

The checklist form of reporting has the obvious advantage of providing a detailed analysis of the pupil's strengths and weaknesses, so that constructive action can be taken to help him improve his learning. It also provides pupils, parents and others with a frequent reminder of the goals of the school. The main difficulties encountered with such reports are in keeping the list of behavioral statements down to a workable number and in stating them in such simple and concise terms that they are readily understood by the users of the reports.¹

The checklist system is considered by many to be a fundamental departure from the traditional marking system. However, Wrinkle² felt that unless a school identifies the objectives of its educational program in terms of behavior,

¹Gronlund, op. cit., p. 376.

²Wrinkle, op. cit., p. 88.

no form or practice used in reporting can be adequate.

Summary of Review of Literature

An abundance of literature and research related to marking and grading procedures was available. Certain factors or trends were evident in the literature when viewed chronologically as follows:

1. Throughout the entire span of literature many studies appeared showing the diversity of pupil marking practices and results. Variability to a large extent was shown among teachers, departments and schools.
2. Individual teachers setting their own marking standards without the benefit of marking policy led to variability, unreliability, and inconsistency of marks.
3. A constant shift was evident from a percentage method of reporting marks to a three to seven symbol system. Most frequently used was a five symbol system with one mark failing and four marks passing.
4. The scientific measurement movement and the Army Alpha and Army Beta tests had a profound effect on pupil marking in that they led to acceptance of the fact that students learn at different rates and with different abilities.

5. There was a re-examination of marking systems and a recognition that the extensive use of standardized achievement test often overlooked the importance of personal and social development.

6. The deeper perceptions and understandings of the complex phenomena of human growth and development caused less concern in research with the mechanical aspects of marking and created more concern with the purposes of marking and their relation to educational objectives.

7. Goals or objectives which have been explicitly defined in behavioral terms are of obvious value in improving the evaluation of learning activities. The precise description of behavior make signs of learning progress, or lack of progress more readily apparent during evaluation. Defining goals in behavioral terms contributes to better instructional planning and it provides pupils with an operational definition of the goals to be achieved.

8. Marking systems are expected to serve diverse functions and as a result the literature fails to report a universally satisfactory reporting method. Some of the methods that have been tried include (1) the traditional marking system (e.g. A, B, C, D, F), (2) the dual marking system, (3) informal letters, (4) parent-teacher conferences,

and (5) the checklist of objectives. Each method has rather severe limitations when used alone. Probably the best reporting system is the checklist of objectives when it combines a compact mark for administrative functions with a more detailed report for teaching and guidance purposes.

9. Whatever marking system is used, it should be meaningful both to the learner and his parents. Research indicates that this usually restricts the use of written explanations and interpretations to assure a degree of uniformity in the meaning of the symbols used.

10. All marking systems should be standardized for all teachers in the system.

11. Marking systems should be subjected to periodic review, preferably in cooperation with parents.

CHAPTER III

DESIGN AND PROCEDURE

Design of the Study

This study was designed to investigate the marking practices of teachers in secondary schools. It was believed that the results of this study would contribute, at least in a minor degree, to the research needed to help provide a consistent and equitable approach to the evaluation of student achievement.

As revealed by the review of professional literature regarding practices, there has been a very pronounced interest in the problem of pupil evaluation. Recently the problem of establishing a consistent and uniform grading procedure consistent with the nature of the changing educational program has assumed a high degree of importance. Dramatic changes in the objectives, content and structure of the curriculum have caused other modifications.

The emphasis upon coordination and integration of instruction gives impetus to the need for coordination of

evaluation procedures. To evaluate the results of a different quality of education by using traditional methods appears to be inconsistent and restrictive.

A major consideration in the design of the study was that of determining the population to be sampled. It was believed that for a study regarding marking practices to be contributive, the population should be limited to a school district. That is, the study should endeavor to sample teachers from a specific school system, rather than be concerned with secondary schools over the state or the nation.

A second consideration regarding the design of the study involved a decision affecting the delimitation of the study and the nature of the group of teachers from whom data would be collected. From the beginning it was quite obvious that some limitations must be placed upon the population to be included in the study. Therefore, the population was limited to teachers of secondary schools in the Oklahoma City Public School System. The study was further limited to teachers of the following subjects: English language arts, social studies, mathematics and science. It was believed that to investigate the marking practices of teachers, the subjects normally viewed as the heart of the academic program

of the school would yield more reliable data than could be obtained through a broad general survey of all subject matter areas.

The selection of a sample group as opposed to utilizing the total population was also considered advisable and necessary because of the number of teachers comprising the total population. The decision was made to utilize a random stratification technique. Random stratification was achieved by selecting a sample with complete randomness and then adjusting the classes of the sample to conform with a stratified distribution.

The sample was drawn from a total population of 635 teachers utilizing accepted methods for selecting random samples as described by Van Dalen.¹ From a total population of 635 teachers, a sample size of 160 was chosen according to procedures outlined by Cella.² This technique gave a sample small enough to study in depth and large enough to be representative of the total population with a sampling error of not more than 10 percent and a 95 percent confidence interval.

¹Van Dalen, op. cit., pp. 249-54.

²Cella, op. cit., pp. 160-63.

The sample used in this study was designed to be a miniature of the population from whence it was selected and presumed to embody the characteristics of the total population defined in the problem.

No attempt was made to select teachers from specific schools within the school district. The only criteria for the selection of teachers in the study was that they teach in the secondary schools and teach in the academic areas of English language arts, social studies, mathematics and science. This arrangement was in keeping with the opinion that the collection of data from the above mentioned areas would yield more reliable data than could be obtained through a broad general survey of all subject matter areas.

Procedure of the Study

This investigation was concerned with only teachers in the public secondary schools of Oklahoma City who taught in the academic areas of English language arts, social studies, mathematics and science. Permission to conduct the study was granted by the Director of Secondary Education and the Director of Research and Statistics for the Oklahoma City Public Schools. The study was supported by all administrative personnel both at the central office and individual school level.

Development and Validation of the Instrument

Since the success of this investigation rested on the development of an acceptable questionnaire, a survey of the literature and consultation with individuals experienced in the construction and use of questionnaires was made before the construction of a questionnaire.

The instrument was developed according to the following procedures as outlined by Rummel in his book, An Introduction to Research Procedures in Education.

1. Express the item as clearly as possible.
2. Choose words that have precise meanings wherever possible.
3. Avoid complex or awkward word arrangements.
4. Include all qualifications needed to provide a reasonable basis for response selection.
5. Avoid the inclusion of non-functional words in the item.
6. Avoid unessential specificity in the questions or in the responses.
7. Avoid the inclusion of trivial questions.
8. Make the suggested answers as simple as possible.
9. Be sure the items will seem to the respondent to apply to the situation.
10. Refrain from asking questions of opinion unless opinion is what is specifically required.

11. Avoid items that are not too suggestive or too unstimulating.
12. Phrase questions to avoid the academically or socially accepted responses.
13. Avoid questions that may be checked with several responses when only one response is desired.
14. Whenever possible, questions should be worded in such a way that they can be answered simply by a check-mark.
15. Ask questions in such a way that they will relieve the respondent of as much complex thinking as possible.
16. Avoid the use of words which are susceptible to different interpretations.¹

After the instrument had been developed, an objective evaluation and comparison of the questionnaire was made using the following criteria as set down by Mouly in his book, The Science of Educational Research.

1. It deals with a significant topic, it makes an important contribution, and is worthy of professional participation.
2. The importance of the problem is clearly stated in the statement of the problem and in the cover letter.
3. It seeks only information not available elsewhere.
4. It is as brief as the study of the problem will allow.
5. The directions are clear, complete, and acceptable.

¹Rummel, op. cit., pp. 126-27.

6. The questions are objective and relatively free from ambiguity and other invalidating features.
7. Questions that may embarrass the respondent or place him on the defensive are avoided.
8. The questions are in good psychological order.
9. The questions are so arranged that they can be tabulated and interpreted readily.¹

Mouly² pointed out that the questionnaire can give reliable and valid information if the above requirements are met in its formulation.

The length and the type of items on the questionnaire have a definite effect on the outcome of a study which uses the questionnaire as a method of obtaining data. A survey of the literature led to the conviction that questionnaires of from three to seven pages would be the most desirable length. The instrument used in this study was five pages in length and it included a cover letter which explained the purpose of the study.

The structured type of item was chosen over the unstructured type. One reason for this choice was expressed by Rummel in his statement that, "the unstructured item requires a respondent to do some hard, reflective thinking

¹Mouly, op. cit., p. 263.

²Mouly, op. cit., p. 263.

and necessitates a lengthy discussion on his part."¹ It was believed that more respondents would be more likely to complete more items, and return the questionnaire if the items required a simple check-mark rather than a more time consuming and demanding type of response.

The questionnaire was substantially revised many times before the first pre-test study. The pre-test questionnaire was submitted to twenty teachers of the same subject matter areas as those who were eventually to receive the final draft. After an analysis of the questionnaire used in the pre-test was made and certain revisions effected, the instrument was printed and mailed to a stratified random sample of the population included in this study.

Although an objective evaluation and comparison to which the questionnaires as an instrument of science must subscribe had been made, there remained the task of identifying the specific ways in which the validity would be established.

The actual validation of the questionnaire came in the second phase of development and validation of the instrument. A more adequate validation required checking the

¹Rummel, op. cit., p. 94.

responses which the questionnaire elicited against an external criterion. This was accomplished by following the questionnaire with a personal interview of a 20 percent sample of the respondents to see whether their responses to the questionnaire actually represented their views on the subject discussed.

A further effort was made to validate the instrument by checking questionnaire responses against actual behavior. This was accomplished through personal observation and interviews in selected classrooms of the 20 percent of teachers selected for the second phase of validation. In general, the observations made in the classrooms showed that the marking practices being used followed the same marking practices as given by the teachers in the interview and on the answered questionnaire.

Percent of Returns in this Study and Related Literature

Many studies in the literature report returns as low as 20 to 40 percent. Shannon¹ reported an average of 65 percent return for "reputable" questionnaire studies reported

¹John R. Shannon, "Percentage of Returns of Questionnaires in Reputable Educational Research," Journal of Educational Research, XLII (October, 1948), pp. 138-41.

in a sample of theses, dissertations, and professional articles.

Clark, Bradsley and Haslacher¹, a research consultant firm, stated that a normal return for a mail questionnaire is from 10 to 20 percent of the questionnaires sent, provided that an appropriate sampling technique is used. The same research firm considered a return of 50 percent or better from a homogeneous group as ample for an indicative sample.

Since 100 percent of the questionnaires distributed in this study were returned in usable form and since the respondents could be considered homogeneous, the returns in this study were regarded as indicative of the population studied.

Statistical Methods Used

In order to determine the meaning of the data derived from the questionnaires, selected statistical techniques were employed. The results of the application of percentages and the use of analysis of variance - multiple classification are presented in Chapter IV. Along with the tabular

¹Clark, Bradsley, and Haslacher, Utah Educational Association: Poll of Member Opinion, (Salt Lake City: Utah Education Association, January, 1960), pp. 42-67.

presentation is a discussion of the data as analyzed according to the selected statistical methods employed.

Percentages were used so that responses of each variable to each item in a specific question could be compared.

The analysis of variance - multiple classification was used to test significance of the difference between two or more groups simultaneously. All items in the instrument were expressed in descriptive units on a continuum. The descriptive units in each item were assigned arbitrary code numbers. Each item in the instrument was then treated by the use of analysis of variance - multiple classification.

Wert, Neidt, and Ahmann pointed out the appropriateness of this statistical treatment for data such as those to be obtained in this study.

When designing experiments in education and psychological research, the possibility of classifying the data in more than one manner is invariably considered. Indeed, it is often not only possible but also advisable to design studies so that results permit the testing of hypotheses concerning separate subdivisions of the data. It is in keeping with efficient experimental methods to incorporate logical multiple classification in such research problems.

In addition, the subdivision of the data into two or more classifications allows the investigator to control certain characteristics known to influence the result of the experiment, or perhaps only suspected of such influences. In the manner possible of bias can be

controlled and the demands of sound experimental design for experiments free of bias can be met, at least in part, by meaningful classification of the data. Such classification also sensitizes the test of significance by enabling the investigator to identify more of the sources of variation in his investigation.¹

After tabulating the data and assigning weighted code numbers, the data were punched on IBM cards. A mathematics teacher and programmer of the Oklahoma City Public Secondary Schools programmed the data for computation of the analysis of variance - multiple classification test of significant difference. The program was then run on the IBM 1401 Computer. Without the use of the computer, a considerable amount of time would have been necessary to compute the analysis of variance - multiple classification on the data analyzed.

¹Wert, Neidt, and Ahmann, op. cit., p. 188.

CHAPTER IV

PRESENTATION AND ANALYSIS OF DATA

The problem of this study was to determine the marking practices of teachers in selected academic areas in the public secondary schools of Oklahoma City and to analyze the sources of variability in those marking practices. In agreement with the design and procedures presented in Chapter III, the following hypotheses were tested:

1. There is no statistically significant difference, other than what might occur by chance, among teachers of the four academic areas of English language arts, mathematics, science, and social studies with regard to marking practices.
2. There is no statistically significant difference, other than what might occur by chance, between probationary and nonprobationary teachers with regard to marking practices.
3. There is no statistically significant difference, other than what might occur by chance, in the marking practices of probationary and nonprobationary teachers among

the four academic areas of English language arts, mathematics, science, and social studies.

An Analysis of the Teacher Sample

A breakdown of the teacher sample according to secondary school teaching level is indicated in Table 1. An inspection of the table revealed that 45 percent of the teachers taught in junior high school while 39.4 percent taught in senior high school. Only 15.6 percent taught in both junior-senior high school.

Table 2 shows the number and percentage of teacher respondents according to sex. The sample was almost evenly divided according to sex with 48.1 percent being male and 44.4 percent female. It is interesting to note that the sex category was the only item on the questionnaire to which teachers did not respond 100 percent.

Table 3 revealed the number and percentage of college or university level courses taken by teachers in testing, measurement and evaluation. An inspection of the grand total column showed that 62.5 percent of the sample completed one or no college level courses in testing, measurement and evaluation. A breakdown of the combined total column of academic areas revealed that 55 percent of the social studies teachers have completed two or more courses

TABLE 1

SECONDARY SCHOOL TEACHING LEVEL

Questionnaire Items		PROBATIONARY				NONPROBATIONARY				COMBINED TOTAL				Grand Total
		Enql	SS	Math	Sci	Enql	SS	Math	Sci	Enql	SS	Math	Sci	
Senior High	F	9	6	5	12	9	6	8	8	18	12	13	20	63
	%	45	30	25	60	45	30	40	40	45	30	32.5	50	39.4
Junior High	F	8	8	12	6	10	12	9	7	18	20	21	13	72
	%	40	40	60	30	50	60	45	35	45	50	52.5	32.5	45
Both	F	3	6	3	2	1	2	3	5	4	8	6	7	25
	%	15	30	15	10	5	10	15	25	10	20	15	17.5	15.6

TABLE 2
SEX OF RESPONDENTS

Questionnaire Items		PROBATIONARY				NONPROBATIONARY				COMBINED TOTAL				Grand Total
		Enql	SS	Math	Sci	Enql	SS	Math	Sci	Enql	SS	Math	Sci	
Male	F	3	15	12	13	2	10	7	15	5	25	19	28	77
	%	15	75	60	65	10	50	35	75	12.5	62.5	47.5	70	48.1
Female	F	15	4	7	6	16	10	10	3	31	14	17	9	71
	%	75	20	35	30	80	50	50	15	77.5	35	42.5	22.5	44.4
Unmarked	F	2	1	1	1	2		3	2	4	1	4	3	12
	%	10	5	5	5	10		15	10	10	2.5	10	7.5	7.5

TABLE 3

NUMBER OF COLLEGE LEVEL COURSES TAKEN BY TEACHERS
IN TESTING, MEASUREMENT AND EVALUATION

Questionnaire Items		PROBATIONARY				NONPROBATIONARY				COMBINED TOTAL				Grand Total
		Engl	SS	Math	Sci	Engl	SS	Math	Sci	Engl	SS	Math	Sci	
None	F	2	3	6	2	2	1	1	3	4	4	7	5	20
	%	10	15	30	10	10	5	5	15	10	10	17.5	12.5	12.5
One	F	11	7	10	10	9	7	3	7	20	14	13	17	64
	%	55	35	50	50	45	35	15	35	50	35	32.5	42.5	40
Two	F	5	8	2	7	4	9	8	4	9	17	10	11	47
	%	25	40	10	35	20	45	40	20	22.5	42.5	25	27.5	29.4
Three	F	2	1	1	1	2	2	4	5	4	3	5	6	18
	%	10	5	5	5	10	10	20	25	10	7.5	12.5	15	11.2
Four	F					1		3		1		3		4
	%					5		15		2.5		7.5		2.5
More	F		1	1		2	1	1	1	2	2	2	1	7
	%		5	5		10	5	5	5	5	5	5	2.5	4.4

in testing, measurement and evaluation, followed by mathematics teachers with 50 percent, science teachers with 45 percent and English teachers with 40 percent.

Table 4 is a breakdown of the teacher sample according to highest degree held by teachers. A high percentage of the teachers, 66.9 percent, was at the bachelor's degree level. Only 1.2 percent of the teacher sample held a special certificate, however, 31.9 percent of all teachers surveyed held a masters degree.

Table 5 indicates the number and percentage of respondents' years of teaching experience. This table shows that social studies teachers, science teachers, English teachers and mathematics teachers, in descending order of percent, had the fewest number of years of teaching experience. Approximately 70 percent of the respondents in this study had 10 years of teaching experience or less.

Table 6 indicates the number and percentage of teachers in each age category. Forty-five percent of the teachers in the survey were 30 years or younger. Twenty percent of the teachers in the sample were 51 years or older. Fifty-five percent of the probationary English and mathematics teachers were 25 years old or younger. The converse was true in the nonprobationary teacher category, with 45

TABLE 4

HIGHEST DEGREE HELD BY TEACHERS

Questionnaire Items		PROBATIONARY				NONPROBATIONARY				COMBINED TOTAL				Grand Total
		Enql	SS	Math	Sci	Enql	SS	Math	Sci	Enql	SS	Math	Sci	
Doctor's Degree	F %													
Special Certificate	F %						1 5		1 5		1 2.5		1 2.5	2 1.20
Master's Degree	F %	2 10		3 15		11 55	14 70	9 45	12 60	13 32.5	14 35	12 30	12 30	51 31.9
Bachelor's Degree	F %	18 90	20 100	17 85	20 100	9 45	5 25	11 55	7 35	27 67.5	25 62.5	28 70	27 67.5	10.7 66.9

TABLE 5

RESPONDENTS' YEARS OF TEACHING EXPERIENCE

Questionnaire Items		PROBATIONARY				NONPROBATIONARY				COMBINED TOTAL				Grand Total
		Engl	SS	Math	Sci	Engl	SS	Math	Sci	Engl	SS	Math	Sci	
0 - 4	F	16	19	17	16	3	3	1	3	19	22	18	19	78
	%	80	95	85	80	15	15	5	15	47.5	55	45	47.5	48.7
5 - 10	F	2	1	3	1	7	3	5	11	9	5	8	12	33
	%	10	5	15	5	35	15	25	55	22.5	10	20	30	20.6
11 - 15	F	1			3	3	2	2	2	4	2	2	5	13
	%	5			15	15	10	10	10	10	5	5	12.5	8.1
16 - 20	F					3	4	5		3	4	5		12
	%					15	20	25		7.5	10	12.5		7.5
21 - 25	F						1		1		1		1	2
	%						5		5		2.5		2.5	1.3
26 - 30	F					3	2	3		3	2	3		8
	%					15	10	15		7.5	5	7.5		5
30 +	F	1				1	5	4	3	2	5	4	3	14
	%	5				5	25	20	15	5	12.5	10	7.5	8.8

TABLE 6

AGE CATEGORY OF RESPONDENTS

Questionnaire Items		PROBATIONARY				NONPROBATIONARY				COMBINED TOTAL				Grand Total
		Enql	SS	Math	Sci	Enql	SS	Math	Sci	Enql	SS	Math	Sci	
20 - 25	F	11	7	11	9	1	2	1	1	12	9	12	10	43
	%	55	35	59	45	5	10	5	5	30	22.5	30	25	26.9
26 - 30	F	4	9	6	1	3		1	5	7	9	7	6	29
	%	20	45	30	5	15		5	25	17.5	22.5	17.5	15	18.1
31 - 35	F	1	2	1	3	2	4	5	6	3	6	6	9	24
	%	5	10	5	15	10	20	25	30	7.5	15	15	22.5	15
36 - 40	F	2		1	4	2	1	1	2	4	1	2	6	13
	%	10		5	20	10	5	5	10	10	2.5	5	15	8.1
41 - 45	F		1	1	2	2	4	1	1	2	5	2	3	12
	%		5	5	10	10	20	5	5	5	12.5	5	7.5	7.5
46 - 50	F				1	1	3	1		1	3	1	1	6
	%				5	5	15	5		2.5	7.5	2.5	2.5	3.8
51 - 65	F	2	1			9	6	10	5	11	7	10	5	33
	%	10	5			45	30	50	25	27.5	17.5	25	12.5	20.6

percent of the English teachers and 50 percent of the mathematics teachers reporting 51 years or older. The remaining probationary and nonprobationary teachers were distributed evenly among the age groups.

An Analysis of the Questionnaire Items

The teacher questionnaire contained two sections. Section I asked for general information about the respondent; section II included questions about practices directly related to teacher evaluation procedures.

Table 7 shows the criteria used by teachers in evaluating pupils. The responses were analyzed by probationary-nonprobationary teachers in academic areas. In keeping with the design and procedure presented in Chapter III, the four academic areas were subdivided into four columns under the headings of probationary and nonprobationary teachers. As can be seen from Table 7, there were 20 teachers in each academic area column under the probationary and nonprobationary captions. The four academic area columns under the probationary and nonprobationary captions were added together to form the combined total column. A total of 40 teachers were contained in each of the four columns under the combined total caption. Thus, the sum of the frequency responses in the grand total

TABLE 7

PRIMARY BASIS USED IN EVALUATING ACHIEVEMENT

Questionnaire Items	PROBATIONARY				NONPROBATIONARY				COMBINED TOTAL				Grand Total
	Engl	SS	Math	Sci	Engl	SS	Math	Sci	Engl	SS	Math	Sci	
Self improve- F ment or growth%	3 15	1 5	1 5	3 15	2 10	2 10	4 20	3 15	5 12.5	3 7.5	5 12.5	6 15	19 11.9
Achievement F with regard % to class	9 45	7 35	8 40	7 35	3 15	6 30	3 15	4 20	12 30	13 32.5	11 27.5	11 27.5	47 29.4
Achievement F with regard % to ability	6 30	8 40	7 35	4 20	14 70	9 45	9 45	11 55	20 50	17 42.5	16 40	15 37.5	68 42.5
Absolute F achievement %	2 10	4 20	4 20	6 30	1 5	3 15	4 20	2 10	3 7.5	7 17.5	8 20	8 20	26 16.2
Other F %													
<u>ANALYSIS OF VARIANCE</u>													
<u>Source of</u> <u>Variation</u>					<u>Degrees of</u> <u>Freedom</u>				<u>F-Value</u>				
Probationary and					F 1,152				0.0683				
Nonprobationary					F 3,152				0.2910				
Academic Area					F 3,152				0.5339				
Interaction													

column was 160 or the exact sample size.

The analysis of variance table directly beneath the raw data table shows the statistical difference between the probationary and nonprobationary teachers and also among teachers of the four academic areas. The table further shows the statistical differences among the various combinations of academic area teachers in both probationary and nonprobationary groups.

Table 7 identifies the criteria teachers use in evaluating achievement for grading purposes. An inspection of the raw data table revealed that in each of the four academic areas, achievement with regard to ability was checked more often than any other category. Over 42 percent of the teachers in the total sample selected this response. Achievement with respect to class was used by an appreciable 29.4 percent of teachers in the sample. Of the remaining teachers in the sample, 16.2 percent checked the "absolute achievement" category, while 11.9 percent checked the "self improvement or growth" category.

All of the F values were nonsignificant. Therefore, in the case of the two main effects, probationary and nonprobationary teachers did not differ in their evaluation procedures ($F = .06$), nor did the teachers in the four

academic areas differ in their evaluation procedures ($F = .29$). The nonsignificant F value for interaction ($F = .53$) was interpreted to mean that there was uniformity of marking practices among the four academic areas of probationary and nonprobationary teachers.

Since no significant F values were found among the sources of variation, the three null hypotheses of no statistically significant difference were accepted.

Table 8 presents the teachers' responses to the question, "What consideration do you give to such factors as judgment of character and citizenship in determining a pupil's final grade?" Forty percent of the teachers in the sample indicated that they gave "moderate" consideration to such factors as judgment of character and citizenship when determining a pupil's grade, and over 36 percent in the four academic areas indicated that they gave "minor" consideration. Of the remaining teachers in the sample, 15.6 percent checked the "none" category, while 7.5 percent checked the "major" category.

None of the F values were significant. Therefore, in the case of the two main effects, probationary and nonprobationary teachers did not differ in their marking practices ($F = .42$), nor did the teachers in the four

TABLE 8

CONSIDERATION GIVEN TO JUDGMENT OF CHARACTER AND CITIZENSHIP

Questionnaire Items		PROBATIONARY				NONPROBATIONARY				COMBINED TOTAL				Grand Total
		Enql	SS	Math	Sci	Enql	SS	Math	Sci	Enql	SS	Math	Sci	
Sole	F													
	%													
Major	F	2	1	2	1	2		3	1	4	1	5	2	12
	%	10	5	10	5	10		15	5	10	2.5	12.5	5	7.5
Moderate	F	9	10	3	7	7	12	6	10	16	22	9	17	64
	%	45	50	15	35	35	60	30	50	40	55	22.5	42.5	40
Minor	F	7	6	9	10	9	6	6	6	16	12	15	16	59
	%	35	30	45	50	45	30	30	30	40	30	37.5	40	36.9
None	F	2	3	6	2	2	2	5	3	4	5	11	5	25
	%	10	15	30	10	10	10	25	15	10	12.5	27.5	12.5	15.6
<u>ANALYSIS OF VARIANCE</u>														
<u>Source of Variation</u>						<u>Degrees of Freedom</u>				<u>F-Value</u>				
Probationary and Nonprobationary						F 1,152				0.4277				
Academic Area						F 3,152				1.0328				
Interaction						F 3,152				0.3811				

academic areas differ in their evaluation procedures ($F = 1.03$). The nonsignificant F value for interaction ($F = .38$) was interpreted to mean that there was uniformity of marking practices found among the four academic areas of probationary and nonprobationary teachers.

Since no significant F values were found among the sources of variation, the three null hypotheses of no statistically significant difference were accepted.

Table 9 shows the opinions of teachers concerning the consideration given to the amount of improvement the pupil has made rather than on the level of achievement the pupil has reached when determining final grades. The possible responses ranged from "none" to "sole" consideration. This same set of response alternatives was used in many subsequent questions. An inspection of the raw data table indicated that 47.5 percent of all teachers surveyed gave "moderate" consideration to the amount of improvement a pupil makes. A further inspection by item of the various positions of teacher groups revealed a general pattern of response. This pattern showed that over 40 percent of the probationary and nonprobationary teachers across academic areas checked "moderate" consideration. Thirty-five percent of the probationary and nonprobationary teachers across

TABLE 9

CONSIDERATION GIVEN TO IMPROVEMENT

Questionnaire Items		PROBATIONARY				NONPROBATIONARY				COMBINED TOTAL				Grand Total
		Enql	SS	Math	Sci	Enql	SS	Math	Sci	Enql	SS	Math	Sci	
Sole	F													
	%													
Major	F	9	8	6	5	12	4	5	7	21	12	11	12	56
	%	45	40	30	25	60	20	25	35	52.5	30	27.5	30	35
Moderate	F	10	11	10	9	8	11	9	8	18	22	19	17	76
	%	50	55	50	45	40	55	45	40	45	55	47.5	42.5	47.5
Minor	F	1	1	4	4		3	5	5	1	4	9	9	23
	%	5	5	20	20		15	25	25	2.5	10	22.5	22.5	14.4
None	F				2		2	1			2	1	2	5
	%				10		10	5			5	2.5	5	3.1
<u>ANALYSIS OF VARIANCE</u>														
<u>Source of</u>						<u>Degrees of</u>				<u>F-Value</u>				
<u>Variation</u>						<u>Freedom</u>								
Probationary and						F 1,152				0.2805				
Nonprobationary						F 3,152				4.2608				
Academic Area						F 3,152				2.2558				
Interaction														

the four academic areas indicated "major" consideration given to the amount of improvement a pupil makes. Of the remaining teachers in the sample, 3.1 percent checked the "none" category, while 14.4 percent responded to the "minor" category.

In the overall two way analysis of variance, no significant difference was found between probationary and nonprobationary teachers across academic areas ($F = .28$). In addition, no significant interaction was found among academic areas across teacher status. However, a significant difference was found among academic areas across teacher status ($F = 4.26$). In order to determine where the significant differences were located, the data were further analyzed using simple effects analysis of variance and contrast comparison analysis of variance.

A simple effects analysis of variance for probationary teachers yielded a significant F value ($F = 5.90$) indicating significant difference among academic areas in regard to the consideration given to the amount of improvement the pupil had made rather than the level of achievement reached when determining a final grade. Individual contrast comparisons indicated there were no significant differences between English and social studies ($F = .01$);

between social studies and mathematics ($F = 2.56$); and between mathematics and science ($F = .01$). However, significant differences were found between English and science teachers ($F = 3.40$); between English and mathematics teachers ($F = 2.96$); and between social studies and science teachers ($F = 2.96$).

A simple effects analysis revealed a significant difference among academic areas of nonprobationary teachers ($F = 4.25$). Individual contrast comparisons indicated significant differences between English and social studies teachers ($F = 3.87$); and between English and mathematics teachers ($F = 2.96$). No significant differences were found between social studies and mathematics teachers ($F = 2.18$); between social studies and science teachers ($F = .38$); and between mathematics and science teachers ($F = .23$).

There were no statistically significant differences found between probationary and nonprobationary teachers; therefore, the null hypothesis of no statistically significant difference between probationary and nonprobationary teachers with regard to marking practices was accepted.

In view of the significant differences among teachers of the four academic areas, the null hypothesis of no

statistically significant difference among teachers of the four academic areas with regard to marking practices was rejected.

The nonsignificant F value for interaction indicated uniformity of marking practices found among the four academic areas of probationary and nonprobationary teachers; therefore, the null hypothesis of no statistically significant difference in the marking practices of probationary and nonprobationary teachers among the four academic areas was accepted.

In Table 10, the teachers were asked what consideration they gave to the level of achievement reached by a pupil rather than the amount of improvement the pupil had made when determining a final grade. As expected, none of the teachers gave "sole" consideration to this aspect of student performance. In general, the pattern of responses in this question was similar to that in Table 6. The combined percentages for the two responses "moderate" or "major" indicated that 77.5 percent of all the teachers surveyed reported in these two groups. Of the remaining teachers in the sample, 18.1 percent checked the "minor" category, while 3.8 percent responded to the "none" category.

TABLE 10

CONSIDERATION GIVEN TO ACHIEVEMENT

Questionnaire Items		PROBATIONARY				NONPROBATIONARY				COMBINED TOTAL				Grand Total
		Enql	SS	Math	Sci	Enql	SS	Math	Sci	Enql	SS	Math	Sci	
Sole	F				1								1	1
	%				5								2.5	.6
Major	F	12	8	12	8	2	5	11	10	14	13	23	18	68
	%	60	40	60	40	10	25	55	50	35	32.5	57.5	45	42.5
Moderate	F	5	7	7	8	12	7	3	7	17	14	10	15	56
	%	25	35	35	40	60	35	15	35	42.5	35	25	37.5	35
Minor	F	3	5	1	2	5	5	5	3	8	10	6	5	29
	%	15	25	5	10	25	25	25	15	20	25	15	12.5	18.1
None	F				1	1	3	1		1	3	1	1	6
	%				5	5	15	5		2.5	7.5	2.5	2.5	3.8
<u>ANALYSIS OF VARIANCE</u>														
Source of Variation						Degrees of Freedom				F-Value				
Probationary and						F 1,152				7.5124				
Academic Area						F 3,152				2.4863				
Interaction						F 3,152				1.3906				

Tables 9 and 10 revealed that teachers who gave "moderate" consideration to improvement also tended to give "major" consideration to achievement. Conversely, those who gave "major" consideration to achievement also tended to give "moderate" consideration to improvement.

The overall analysis of variance indicated a significant difference between probationary and nonprobationary teachers ($F = 7.51$). Further overall analysis demonstrated no significant difference among academic areas across teacher status ($F = 2.48$), nor was there demonstrated a significant interaction of academic areas ($F = 1.39$).

A simple effects analysis of variance revealed a significant difference between probationary and nonprobationary teachers with probationary teachers having a higher score than nonprobationary teachers in the area of English ($F = 7.10$). No significant differences were found between probationary and nonprobationary teachers in the area of social studies ($F = 2.94$); mathematics ($F = 1.78$); and science ($F = .04$).

In view of the statistically significant difference found between probationary and nonprobationary teachers, the null hypothesis of no statistically significant difference between probationary and nonprobationary teachers with

regard to marking practices was rejected.

There was not a significant difference among the four academic areas; therefore, the null hypothesis of no statistically significant difference among teachers of the four academic areas with regard to marking practices was accepted.

The nonsignificant F value for interaction was interpreted to mean that there was uniformity of marking practices among teachers of the four academic areas; therefore, the null hypothesis of no statistically significant difference among probationary and nonprobationary teachers of the four academic areas with regard to marking practices was accepted.

The next question was concerned with the consideration given to the process of total educational evaluation when determining a pupil's final grade. The responses are presented in Table 11. It is obvious that the total process of educational evaluation was only "moderately" considered by about one-half of the teachers in the sample. Twenty-seven percent of the teachers in the four academic areas reported that they gave "major" consideration to the total process of evaluation. Of the remaining teachers in the sample, 18 percent indicated "minor," while 7 percent

TABLE 11

CONSIDERATION GIVEN TO EDUCATIONAL EVALUATION

Questionnaire Items		PROBATIONARY				NONPROBATIONARY				COMBINED TOTAL				Grand Total
		Enql	SS	Math	Sci	Enql	SS	Math	Sci	Enql	SS	Math	Sci	
Sole	F		1					1	1		1	1	1	3
	%		5					5	5		2.5	2.5	2.5	2
Major	F	9	7	7	5	6	3	1	6	15	10	8	11	44
	%	45	35	35	25	30	15	5	30	37.5	25	20	27.5	27.5
Moderate	F	7	10	5	9	12	10	12	8	19	20	17	17	73
	%	35	50	25	45	60	50	60	40	47.5	50	42.5	42.5	45.5
Minor	F	4	1	6	4	2	4	5	3	6	5	11	7	29
	%	20	5	30	20	10	20	25	15	15	12.5	27.5	17.5	18
None	F		1	2	2		3	1	2		4	3	4	11
	%		5	10	10		15	5	10		10	7.5	10	7
<u>ANALYSIS OF VARIANCE</u>														
<u>Source of Variation</u>						<u>Degrees of Freedom</u>				<u>F-Value</u>				
Probationary and Nonprobationary						F 1,152				0.9503				
Academic Area						F 3,152				1.4112				
Interaction						F 3,152				1.6416				

responded to the category "none."

All of the F values were nonsignificant. Therefore, in the case of the two main effects, evidence was found that proves that the probationary and nonprobationary teachers did not differ in regard to marking practices ($F = .95$), nor did the teachers of the four academic areas differ ($F = 1.41$) in their evaluation procedures. The nonsignificant F value for interaction ($F = 1.64$) was interpreted to mean that there was uniformity of marking practices found among probationary and nonprobationary teachers of the various academic areas.

Since no significant F values were found among the sources of variation, the three null hypotheses of no significant difference were accepted for question 8.

Responses to a question concerning teacher consideration given to absolute measurement that concentrates on a specific, well defined trait are presented in Table 12. As expected relatively few teachers gave "sole" or "none" consideration to this aspect of evaluation. The most popular response for all groups was the "moderate" category. This item exhibited consistency across the four academic areas. Percentage differences among the various academic areas was determined by comparing the distribution between

TABLE 12

CONSIDERATION GIVEN TO ABSOLUTE MEASUREMENT

Questionnaire Items		PROBATIONARY				NONPROBATIONARY				COMBINED TOTAL				Grand Total
		Engl	SS	Math	Sci	Engl	SS	Math	Sci	Engl	SS	Math	Sci	
Sole	F													
	%													
Major	F	4	5	14	10	6	4	8	7	10	9	22	17	58
	%	20	25	70	50	30	20	40	35	25	22.5	55	42.5	36.3
Moderate	F	8	8	5	8	10	9	6	10	18	17	11	18	64
	%	40	40	25	40	50	45	30	50	45	42.5	27.5	45	40
Minor	F	4	7		1	3	5	3	3	7	12	3	4	26
	%	20	35		5	15	25	15	15	17.5	30	7.5	10	16.2
None	F	4		1	1	1	2	3		5	2	4	1	12
	%	20		5	5	5	10	15		12.5	5	10	2.5	7.5
<u>ANALYSIS OF VARIANCE</u>														
<u>Source of</u>						<u>Degrees of</u>				<u>F-Value</u>				
<u>Variation</u>						<u>Freedom</u>								
Probationary and						F 1,152				0.8134				
Nonprobationary						F 3,152				3.5137				
Academic Area						F 3,152				2.6353				
Interaction														

the "minor" and "major" response alternatives. Between 22.5 and 25 percent of the English and social studies teachers checked the "major" category, while between 42.5 and 55 percent of the mathematics and science teachers responded to this category. The percentages for the "minor" response were somewhat smaller. Between 7.5 and 17.5 percent of the English, mathematics, and science teachers checked this response, while 30 percent of the social studies teachers responded to this category. Other interesting comparisons can be made by looking at the rows of Table 12. This permits the checking of agreement among teachers within a subject field and across subject matter areas. It can readily be seen that there were marked differences from field to field.

The overall analysis of variance indicated no significant difference between probationary and nonprobationary teachers ($F = .81$). A significant F value was obtained among academic areas across teacher status ($F = 3.51$). Further overall analysis demonstrated no significant interaction of academic areas as a function of teacher status ($F = 2.60$).

The simple effects analysis of variance revealed no significant differences among academic areas of nonprobationary

teachers ($F = .93$). Simple effects analysis of probationary teachers among academic areas revealed a significant F value of ($F = 5.27$). Individual comparisons showed a significant difference among probationary teachers in the areas of English and mathematics ($F = 4.38$), and mathematics and science ($F = 2.90$).

No significant differences appeared between English and social studies ($F = .39$); between English and science ($F = 2.46$); between social studies and mathematics ($F = 2.4$); and between social studies and science ($F = .88$).

In the case of the first main effect, evidence was found proving that the probationary and nonprobationary teachers did not differ in regard to marking practices; therefore, the null hypothesis of no significant difference between probationary and nonprobationary teachers with regard to marking practices was accepted.

In view of the statistically significant differences found among the academic areas, the null hypothesis of significant differences among teachers of the four academic areas with regard to marking practices was rejected.

The nonsignificant F value for interaction was interpreted to mean that there was uniform marking practices found between probationary and nonprobationary teachers of

the various academic areas; therefore, the null hypothesis of no statistically significant difference in the marking practices of probationary and nonprobationary teachers among the four academic areas was accepted.

The data in Table 13 refers to the consideration given to the course objectives stated in terms of desired behavior outcomes. The results generally indicated that 66.9 percent of the teachers gave "major" consideration to course objectives as stated in terms of desired behavior outcomes. Differences among the various academic areas were determined by combining the "moderate" and "minor" categories under the combined total column. In the combined total column, 17 percent of the mathematics and science teachers checked the "moderate" or "minor" category, while over 28 percent of the English and social studies teachers checked the same levels of response. The extent of this practice reflects the degree of importance the four academic areas attached to the stating of course objectives in terms of desired behavior outcomes.

The overall analysis of variance indicated no significant difference between probationary and nonprobationary teachers ($F = 1.16$). Further overall analysis demonstrated no significant differences of marking practices among

TABLE 13

CONSIDERATION GIVEN TO OBJECTIVES THAT ARE STATED
IN TERMS OF DESIRED BEHAVIOR OUTCOMES

Questionnaire Items		PROBATIONARY				NONPROBATIONARY				COMBINED TOTAL				Grand Total
		Enql	SS	Math	Sci	Enql	SS	Math	Sci	Enql	SS	Math	Sci	
Sole	F	1	1	4						1	1	4		6
	%	5	5	20						2.5	2.5	10		3.8
Major	F	13	11	12	15	15	12	12	17	28	23	24	32	107
	%	65	55	60	75	75	60	60	85	70	57.5	60	80	66.9
Moderate	F	4	8	3	3	5	7	4	3	9	15	7	6	37
	%	20	40	15	15	25	35	20	15	22.5	37.5	17.5	15	23.1
Minor	F	2		1	1			1		2		2	1	5
	%	10		5	5			5		5		5	2.5	3.1
None	F				1		1	3			1	3	1	5
	%				5		5	15			2.5	7.5	2.5	3.1
<u>ANALYSIS OF VARIANCE</u>														
<u>Source of Variation</u>						<u>Degrees of Freedom</u>				<u>F-Value</u>				
Probationary and						F 1,152				1.1600				
Nonprobationary						F 3,152				0.4021				
Academic Area						F 3,152				3.2324				
Interaction														

academic areas across teacher status ($F = .40$). However, a significant interaction of academic areas as a function of teacher status was obtained ($F = 3.23$).

In order to determine the location of variability a simple effects analysis was completed. This analysis revealed no significant difference among academic areas for probationary teachers. However, a simple effects analysis of nonprobationary teachers indicated significant differences among academic areas ($F = 2.71$).

Contrast analysis of variance among nonprobationary teachers revealed no significant difference between English and social studies teachers ($F = .37$); between English and mathematics teachers ($F = 1.56$); between science and English teachers ($F = .05$); between social studies and mathematics teachers ($F = .39$); between social studies and science teachers ($F = .77$); and between science and mathematics teachers ($F = 2.26$).

Simple effects analysis comparing probationary and nonprobationary teachers within a given subject indicated a highly significant F value between probationary mathematics teachers and nonprobationary mathematics teachers ($F = 9.24$). No significant differences between probationary and nonprobationary teachers in the areas of English ($F = .18$),

social studies ($F = .43$), and science ($F = 1.98$) were found.

In view of the no statistically significant difference found between probationary and nonprobationary teachers, the null hypothesis of no significant difference between probationary and nonprobationary teachers with regard to marking practices was accepted.

There was not a significant difference among the teachers in the four academic areas; therefore, the null hypothesis of no statistically significant difference among teachers of the four academic areas with regard to marking practices was accepted.

The significant F value for interaction was interpreted to mean that there were different marking practices found between probationary and nonprobationary teachers of the various academic areas; therefore, the null hypothesis of no statistically significant difference in the marking practices of probationary and nonprobationary teachers among the four academic areas was rejected.

The next question considered the achievement of course objectives. There was an attempt to determine the extent to which the achievement of course objectives determined a pupil's grade. The responses are presented in Table 14. The results generally indicated that 67.5 percent of the

TABLE 14

CONSIDERATION GIVEN TO ACHIEVEMENT OF COURSE OBJECTIVES

Questionnaire Items		PROBATIONARY				NONPROBATIONARY				COMBINED TOTAL				Grand Total
		Enql	SS	Math	Sci	Enql	SS	Math	Sci	Enql	SS	Math	Sci	
Sole	F			1				1				2		2
	%			5				5				5		1.3
Major	F	13	14	15	16	12	7	17	14	25	21	32	30	108
	%	65	70	75	80	60	35	85	70	62.5	52.5	80	75	67.5
Moderate	F	6	6	4	3	8	11	1	6	14	17	5	9	45
	%	30	30	20	15	40	55	5	30	35	42.5	12.5	22.5	28.1
Minor	F	1			1		2	1		1	2	1	1	5
	%	5			5		10	5		2.5	5	2.5	2.5	3.1
None	F													
	%													
<u>ANALYSIS OF VARIANCE</u>														
<u>Source of Variation</u>						<u>Degrees of Freedom</u>				<u>F-Value</u>				
Probationary and Nonprobationary						F 1,152				1.7547				
Academic Area						F 3,152				4.0657				
Interaction						F 3,152				1.8127				

teachers in the study gave "major" consideration to the achievement of course objectives. Again there were percentage differences from academic area to academic area. Differences among the various groups were determined by combining the "moderate" and "minor" categories under the combined total column. Over 21 percent of the mathematics and science teachers checked the "moderate" or "minor" category, while 42 percent of the English and social studies teachers checked the same level of response.

It should be noted that certain parallels appear in responses presented here and in Table 13. In both Tables 13 and 14, English and social studies teachers considered course objectives and the achievement of course objectives less important than did mathematics and science teachers.

In the overall two way analysis of variance, a significant F value ($F = 4.06$) was obtained among academic areas across teacher status. No overall significant differences were found between probationary and nonprobationary teachers across academic areas ($F = 1.75$). The F value of ($F = 1.81$) was not significant for interaction.

Simple effects analysis of probationary teachers indicated no significant difference in marking practices among teachers of the four academic areas ($F = .75$). A

significant F value ($F = 5.28$) was found among academic areas for nonprobationary teachers. Individual contrast comparisons revealed a significant difference among nonprobationary teachers between social studies and mathematics teachers ($F = 5.03$). However, no significant differences were found between English and social studies teachers ($F = 1.39$); between English and mathematics teachers ($F = 1.07$); between English and science teachers ($F = .10$); between science and social studies teachers ($F = 2.39$); between science and mathematics teachers ($F = .46$).

There were no overall significant differences found between probationary and nonprobationary teachers; therefore, the null hypothesis of no statistically significant difference between probationary and nonprobationary teachers with regard to marking practices was accepted.

In view of the statistically significant difference found among the academic areas, the null hypothesis of no statistically significant difference among teachers of the four academic areas with regard to marking practices was rejected.

The nonsignificant F value for interaction was interpreted to mean that there was uniformity found among teachers

of the various academic areas; therefore, the null hypothesis of no statistically significant difference between probationary and nonprobationary teachers of the various academic areas with regard to marking practices was accepted.

The data in Table 15 revealed the consideration teachers gave to the fact that pupils vary in their ability to achieve the objectives of a specific subject matter. As indicated in the grand total column, 49.4 percent of the teachers gave "major" consideration to the fact that pupils vary in their ability to achieve the course objectives. However, there were differences from one subject to the next with respect to the frequency with which the response "major" was checked. For example, 70 percent of the English teachers checked "major" while only 30 percent of the mathematics teachers indicated a consideration for that response. Therefore, this response category provided a rough distinction of percentage differences among the academic areas. Across teacher status, a similar picture can be obtained by combining the "none," "minor" and "moderate" response categories under the combined total column. In doing this, the percentages for social studies, mathematics, and science teachers varied between 50 and 62.5, whereas the percent for English teachers was exactly 30.

TABLE 15

CONSIDERATION GIVEN TO THE VARIATION OF
INDIVIDUAL PUPIL ABILITY

Questionnaire Items		PROBATIONARY				NONPROBATIONARY				COMBINED TOTAL				Grand Total
		Enql	SS	Math	Sci	Enql	SS	Math	Sci	Enql	SS	Math	Sci	
Sole	F			2			1	1			1	3		4
	%			10			5	5			2.5	7.5		2.5
Major	F	12	10	6	10	16	9	6	10	28	19	12	20	79
	%	60	50	30	50	80	45	30	50	70	47.5	30	50	49.4
Moderate	F	8	5	11	5	3	7	7	6	11	12	18	11	52
	%	40	25	55	25	15	35	35	30	27.5	30	45	27.5	32.5
Minor	F		5	1	4		2	5	4		7	6	8	21
	%		25	5	20		10	25	20		17.5	15	20	13.1
None	F				1	1	1	1		1	1	1	1	4
	%				5	5	5	5		2.5	2.5	2.5	2.5	2.5
<u>ANALYSIS OF VARIANCE</u>														
Source of Variation Probationary and Nonprobationary Academic Area Interaction						Degrees of Freedom				F-Value				
						F 1,152				0.0363				
						F 3,152				2.1676				
						F 3,152				0.9082				

In the overall two way analysis of variance, all of the F values were nonsignificant. Therefore, in the case of the two main effects, it was found that the probationary and nonprobationary teachers did not differ in regard to marking practices ($F = .03$), nor did the teachers of the four academic areas differ ($F = 2.16$). The nonsignificant F value for interaction ($F = .90$) was interpreted to mean that there was uniformity of marking practices found among probationary and nonprobationary teachers of the various academic areas.

Since no significant F values were found among the sources of variation, the three null hypotheses of no statistically significant difference were accepted for question twelve.

Responses to the question concerning the consideration given to classroom behavior in determining course grades are presented in Table 16. As expected, no teacher gave "sole" consideration to this aspect of student performance. Similarities among the various academic areas were determined by checking the percentage distribution in the "minor" response alternative. Between 40 and 47 percent of the teachers checked this category. A similar picture was obtained by checking the "moderate" response category under

TABLE 16

CONSIDERATION GIVEN TO CLASSROOM BEHAVIOR

Questionnaire Items		PROBATIONARY				NONPROBATIONARY				COMBINED TOTAL				Grand Total
		Encl	SS	Math	Sci	Encl	SS	Math	Sci	Encl	SS	Math	Sci	
Sole	F													
	%													
Major	F	4	5	1		1	2	4	3	5	7	5	3	20
	%	20	25	5		5	10	20	15	12.5	17.5	12.5	7.5	12.5
Moderate	F	8	7	5	10	5	7	3	6	13	14	8	16	51
	%	40	35	25	50	25	35	15	30	32.5	35	20	40	31.9
Minor	F	6	7	9	9	13	9	9	9	19	16	18	18	71
	%	30	35	45	45	65	45	45	45	47.5	40	45	45	44.4
None	F	2	1	5	1	1	2	4	2	3	3	9	3	18
	%	10	5	25	5	5	10	20	10	7.5	7.5	22.5	7.5	11.2
ANALYSIS OF VARIANCE														
<u>Source of Variation</u>						<u>Degrees of Freedom</u>				<u>F-Value</u>				
Probationary and Nonprobationary						F 1,152				0.7069				
Academic Areas						F 3,152				1.5681				
Interaction						F 3,152				1.3819				

the combined total column. The percentages for the "none" category are smaller with 7.5 percent of the English, social studies and science teachers checking this alternative, whereas, 22.5 percent of the mathematics teachers checked the "none" response category. It should be noted that 12.5 percent of the teacher sample responded to the "major" category.

The overall analysis of variance revealed no significant F values. Therefore, in the case of the two main effects, evidence again indicated that the probationary and nonprobationary teachers did not differ in regard to marking practices ($F = .70$), nor did the teachers of the four academic areas differ ($F = 1.56$) in their consideration given to classroom behavior. The nonsignificant F value for interaction ($F = 1.38$) was interpreted to mean that there was uniformity of marking procedures found between probationary and nonprobationary teachers of the various academic areas.

Since no significant F values were found among the sources of variation, the three null hypotheses of no significant difference were accepted for question 16.

The data in Table 17 gives the responses to the question "What consideration do you give to 'effort' in

TABLE 17

CONSIDERATION GIVEN TO PUPIL EFFORT

Questionnaire Items		PROBATIONARY				NONPROBATIONARY				COMBINED TOTAL				Grand Total
		Enql	SS	Math	Sci	Enql	SS	Math	Sci	Enql	SS	Math	Sci	
Sole	F			1			1		1		1	1	1	3
	%			5			5		5		2.5	2.5	2.5	1.9
Major	F	13	12	10	8	12	8	6	7	25	20	16	15	76
	%	65	60	50	40	60	40	30	35	62.5	50	40	37.5	47.5
Moderate	F	5	6	6	10	7	8	11	7	12	14	17	17	60
	%	25	30	30	50	35	40	55	35	30	35	42.5	42.5	37.5
Minor	F	2	1	3	2	1	3	3	4	3	4	6	6	19
	%	10	5	15	10	5	15	15	20	7.5	10	15	15	11.9
None	F		1						1		1		1	2
	%		5						5		2.5		2.5	1.2
<u>ANALYSIS OF VARIANCE</u>														
<u>Source of</u>						<u>Degrees of</u>				<u>F-Value</u>				
<u>Variation</u>						<u>Freedom</u>								
Probationary and														
Nonprobationary						F 1,152				1.2821				
Academic Area						F 3,152				1.3387				
Interaction						F 3,152				0.2649				

determining a pupil's grade?" Relatively few teachers (1.9 percent) in the sample responded at the "sole" category. Among teachers of the four academic areas, the "major" category received 47.5 percent of the responses. However, there are some percentage differences from one area to the next with respect to the frequency with which "major" was checked. An inspection of the combined total column revealed that between 50 and 62 percent of the English and social studies teachers responded to the "major" item, while between 37 and 40 percent of the mathematics and science teachers checked this response. Of the remaining teachers in the sample, 11.9 percent responded to the "minor" category, while 1.2 percent checked the "none" category.

All of the F values in Table 17 were nonsignificant. Therefore, in the case of the two main effects, evidence demonstrated that probationary and nonprobationary teachers did not differ in their marking practices ($F = 1.28$), nor did teachers in the four academic areas differ in the evaluation procedures ($F = 1.03$). The nonsignificant F value for interaction ($F = 2.26$) was interpreted to mean that there was uniformity of marking practices found among the four academic areas of probationary and nonprobationary teachers.

Since no significant F values were found among the sources of variation, the three null hypotheses of no statistically significant difference were accepted.

Teacher consideration given to unexcused absences and tardiness in determining a pupil's grade is revealed in the data of Table 18. As expected, no teacher gave this question "sole" consideration. Among teachers of the four academic areas, the "moderate" response received 34.4 percent of the responses followed closely by the "minor" category. The combined percentages for these two responses under the combined total column demonstrated the consideration given to unexcused absences and tardiness by individual academic areas. English teachers responded to the two categories with 57.5 percent followed by mathematics teachers with 59 percent, social studies teachers with 53 percent and science teachers with 45 percent. Over 16 percent of the total teacher sample reported that they gave no consideration to unexcused absences and tardiness when determining a pupil's grade. It should be noted that 18.7 percent of the total teacher sample gave "major" consideration to unexcused absences and tardiness.

The overall analysis of variance revealed no significant F values. Therefore, in the case of the two main

TABLE 18

CONSIDERATION GIVEN TO UNEXCUSED ABSENCES AND TARDINESS

Questionnaire Items		PROBATIONARY				NONPROBATIONARY				COMBINED TOTAL				Grand Total
		Enql	SS	Math	Sci	Enql	SS	Math	Sci	Enql	SS	Math	Sci	
Sole	F													
	%													
Major	F	4	3	4	3	8	4	2	2	12	7	6	5	30
	%	20	15	20	15	40	20	10	10	30	17.5	15	12.5	18.7
Moderate	F	4	9	9	3	7	6	7	10	11	15	16	13	55
	%	20	45	45	15	35	30	35	50	27.5	37.5	40	32.5	34.4
Minor	F	8	6	1	10	4	6	7	7	12	12	8	17	49
	%	40	30	5	50	20	30	35	35	30	30	20	42.5	30.6
None	F	4	2	6	4	1	4	4	1	5	6	10	5	26
	%	20	10	30	20	5	20	20	5	12.5	15	25	12.5	16.3
ANALYSIS OF VARIANCE														
Source of						Degrees of				F-Value				
<u>Variation</u>						<u>Freedom</u>								
Probationary and														
Nonprobationary						F 1,152				1.5058				
Academic Area						F 3,152				0.8633				
Interaction						F 3,152				2.0412				

effects, evidence demonstrated that probationary and non-probationary teachers did not differ in their consideration given to unexcused absences and tardiness ($F = 1.50$), nor did the teachers in the four academic areas differ ($F = .86$). The nonsignificant F value for interaction ($F = 2.04$) indicated that there was uniformity of consideration found among the four academic areas of probationary and nonprobationary teachers.

Since no significant F values were found among the sources of variation, the three null hypotheses of no statistically significant difference were accepted.

Consideration given to quiz scores in determining grades at the end of the grading period is indicated by the data in Table 19. (A quiz was defined as a test of less than 15 minutes duration). As expected, no teacher gave this question "sole" consideration. Among teachers of the four academic areas, the "moderate" category received 44.4 percent of the responses of the teacher sample. The "minor" category was checked by 29.4 percent of the teachers in the sample. Combining the percentages for these two categories under the combined total column revealed the consideration given to quiz scores in determining grades at the end of each grading period. English teachers

TABLE 19

CONSIDERATION GIVEN TO QUIZ SCORES

Questionnaire Items		PROBATIONARY				NONPROBATIONARY				COMBINED TOTAL				Grand Total
		Enql	SS	Math	Sci	Enql	SS	Math	Sci	Enql	SS	Math	Sci	
Sole	F													
	%													
Major	F	3	4	4	5	2	3	5	4	5	7	9	9	30
	%	15	20	20	25	10	15	25	20	12.5	17.5	22.5	22.5	18.7
Moderate	F	11	5	9	7	14	8	9	8	25	13	18	15	71
	%	55	25	45	35	70	40	45	40	62.5	32.5	45	37.5	44.4
Minor	F	5	9	5	6	3	7	5	7	8	16	10	13	47
	%	25	45	25	30	15	35	25	35	20	40	25	32.5	29.4
None	F	1	2	2	2	1	2	1	1	2	4	3	3	12
	%	5	10	10	10	5	10	5	5	5	10	7.5	7.5	7.5
ANALYSIS OF VARIANCE														
Source of Variation					Degrees of Freedom					F-Value				
Probationary and Nonprobationary					F 1,152					0.2109				
Academic Area					F 3,152					0.7512				
Interaction					F 3,152					0.0535				

responded to the two with 82.5 percent followed by mathematics teachers with 80 percent, social studies teachers with 72.5 percent and science teachers with 70 percent. Of the remaining teachers in the sample, 18.7 percent responded to the "major" category, while 7.5 percent checked the "none" category.

The overall analysis of variance revealed no significant F values. Therefore, in the case of the two main effects, the probationary and nonprobationary teachers did not differ in their consideration for quiz results ($F = .21$), nor did the teachers in the four academic areas differ ($F = .75$). The nonsignificant F value for interaction ($F = .05$) indicated there was uniformity of marking practices found among the four academic areas of probationary and nonprobationary teachers.

Since no significant F values were found among the sources of variation, the three null hypotheses of no statistically significant difference were accepted.

The data in Table 20 presents the teachers' responses to the question, "What consideration do you give to homework in determining grades at the end of each grading period?" Relatively few teachers responded at either extreme on the continuum. Further inspection of the raw data table revealed

TABLE 20

CONSIDERATION GIVEN TO THE PLACE OF HOMEWORK
IN DETERMINING GRADES

Questionnaire Items		PROBATIONARY				NONPROBATIONARY				COMBINED TOTAL				Grand Total
		Engl	SS	Math	Sci	Engl	SS	Math	Sci	Engl	SS	Math	Sci	
Sole	F													
	%													
Major	F	3	2	3	3	6	3		3	9	5	3	6	23
	%	15	10	15	15	30	15		15	22.5	12.5	7.5	15	14.4
Moderate	F	10	11	9	15	9	6	12	11	19	17	21	26	83
	%	50	55	45	75	45	30	60	55	47.5	42.5	52.5	65	51.9
Minor	F	5	6	8	2	4	7	6	6	9	13	14	8	44
	%	25	30	40	10	20	35	30	30	22.5	32.5	35	20	27.5
None	F	2	1			1	4	2		3	5	2		10
	%	10	5			5	20	10		7.5	12.5	5		6.2
<u>ANALYSIS OF VARIANCE</u>														
Source of Variation Probationary and Nonprobationary Academic Area Interaction						Degrees of Freedom				F-Value				
						F 1,152				0.8602				
						F 3,152				2.3896				
						F 3,152				1.3134				

that a substantial number of teachers checked either the "moderate" or "minor" category. The combined percentages for these two response categories under the combined total column were between 70 and 87.5 percent. The "none" category under the combined total column for the four academic areas resulted in percentages ranging between 0 to 12.5, while the "major" category varied between 7.5 to 22.5 percent.

The overall analysis of variance revealed no significant F values. Therefore, in the case of the two main effects, evidence indicated that probationary and nonprobationary teachers did not differ in their consideration of homework ($F = .21$), nor did the teachers in the four academic areas differ ($F = .75$). The nonsignificant F value for interaction ($F = .05$) meant that there was uniformity of consideration for homework among the four academic areas of probationary and nonprobationary teachers.

Since no significant F values were found among the sources of variation, the three null hypotheses of no statistically significant difference were accepted.

The data in Table 21 refers to the use of test scores in determining grades at the end of each grading period. (A test was defined as of more than 15 minutes duration, but

TABLE 21

CONSIDERATION GIVEN TO TEST SCORES

Questionnaire Items		PROBATIONARY				NONPROBATIONARY				COMBINED TOTAL				Grand Total
		Enql	SS	Math	Sci	Enql	SS	Math	Sci	Enql	SS	Math	Sci	
Sole	F													
	%													
Major	F	11	16	16	11	9	10	13	12	20	26	29	23	98
	%	55	80	80	55	45	50	65	60	50	65	72.5	57.5	61.3
Moderate	F	8	4	3	9	8	8	5	8	16	12	8	17	53
	%	40	20	15	45	40	40	25	40	40	30	20	42.5	33.1
Minor	F	1		1		3	2	1		4	2	2		8
	%	5		5		15	10	5		10	5	5		5
None	F							1				1		1
	%							5				2.5		.6
<u>ANALYSIS OF VARIANCE</u>														
Source of Variation						Degrees of Freedom				F-Value				
Probationary and Nonprobationary						F 1,152				4.2149				
Academic Area						F 3,152				1.0976				
Interaction						F 3,152				0.9220				

not longer than one class period). As expected, relatively few teachers gave no consideration to this aspect of student performance. No teacher in any of the four academic areas used test results as the only criterion for assigning grades. A very substantial number of the teacher sample checked either the "major" or "moderate" category. The combined percentage for these two responses under the grand total column was 94.4 percent. We thus find that the total teacher sample gave very little consideration to the two responses "minor" or "none." The only percentage differences in Table 21 were in the responses to the category "major." Probationary teachers all appeared to use test results extensively as indicated by the fact that between 55 and 80 percent of this status group checked "major." Between 45 and 60 percent of the nonprobationary teachers checked the "major" category.

The overall analysis of variance revealed a significant difference between probationary and nonprobationary teachers across academic areas ($F = 4.21$). A significant difference was not found among academic areas ($F = 1.09$), nor was there a significant interaction indicated ($F = .92$).

The contrast analysis of variance yielded a significant difference between probationary social studies teachers and nonprobationary social studies teachers.

($F = 6.75$). The contrast analysis did not reveal a significant difference between probationary and nonprobationary teachers in the areas of English ($F = 1.08$), mathematics ($F = 1.70$), and science ($F = .08$).

In view of the statistically significant difference found between probationary and nonprobationary teachers, the null hypothesis of no statistically significant difference between probationary and nonprobationary teachers with regard to marking practices was rejected.

There were no significant differences among teachers of the four academic areas; therefore, the null hypothesis of no statistically significant difference among teachers of the four academic areas with regard to marking practices was accepted.

The nonsignificant F value for interaction demonstrated the uniformity of marking practices found among the four academic areas of probationary and nonprobationary teachers; therefore, the null hypothesis of no statistically significant difference in the marking practices of probationary and nonprobationary teachers among the four academic areas was accepted.

The next question was concerned with the use of marks as persuasive devices to induce an increase of student effort. The responses are presented in Table 22. A substantial

TABLE 22

CONSIDERATION GIVEN TO THE USE OF MARKS
AS PERSUASIVE DEVICES

Questionnaire Items		PROBATIONARY				NONPROBATIONARY				COMBINED TOTAL				Grand Total
		Engl	SS	Math	Sci	Engl	SS	Math	Sci	Engl	SS	Math	Sci	
Sole	F													
	%													
Major	F	4	8	8	5	4	4	2	7	8	12	10	12	42
	%	20	40	40	25	20	20	10	35	20	30	25	30	26.2
Moderate	F	11	10	8	11	6	9	8	9	17	19	16	20	72
	%	55	50	40	55	30	45	40	45	42.5	47.5	40	50	45
Minor	F	5	1	2	4	9	5	5	4	14	6	7	8	35
	%	25	5	10	20	45	25	25	20	35	15	17.5	20	21.9
None	F		1	2		1	2	5		1	3	7		11
	%		5	10		5	10	25		2.5	7.5	17.5		6.9
<u>ANALYSIS OF VARIANCE</u>														
Source of Variation Probationary and Nonprobationary Academic Area Interaction						Degrees of Freedom				F-Value				
						F 1,152				7.4772				
						F 3,152				1.7159				
						F 3,152				1.8345				

number of the total teacher sample checked either the "major" or "moderate" category. The combined percentage for these two response categories was 71.2 for each of the four academic areas. Combining the "none" and "minor" categories for the academic areas resulted in 28.8 percent of the teacher sample. The only major percentage difference in Table 22 was in the responses to the two categories "major" and "moderate." Probationary teachers all appeared to use marks as persuasive devices extensively as indicated by the fact that between 75 and 90 percent of this status group checked either "major" or "moderate." Between 50 and 80 percent of the nonprobationary teachers checked the "major" or "moderate" category.

The overall analysis of variance revealed no significant difference among academic areas across teacher status ($F = 1.71$). No significant interaction of probationary and nonprobationary teachers was found among the four academic areas ($F = 1.74$). However, an F value of ($F = 7.47$) indicated a significant difference between probationary and nonprobationary teachers.

Further analysis using simple effects analysis of variance revealed a significant difference between probationary and nonprobationary mathematics teachers in their

consideration to the use of marks as persuasive devices ($F = 8.04$). No significant differences were found between probationary and nonprobationary teachers in the areas of English ($F = 1.28$), social studies ($F = 3.57$), and science ($F = .14$), with regard to this criterion.

In view of no significant difference among academic areas, the null hypothesis of no statistically significant difference among teachers of the four academic areas with regard to marking practices was accepted.

The nonsignificant F value for interaction meant that there was uniformity of marking practices found among the four academic areas of probationary and nonprobationary teachers; therefore, the null hypothesis of no statistically significant difference in the marking practices of probationary and nonprobationary teachers among the four academic areas was accepted.

In view of the statistically significant difference found between probationary and nonprobationary teachers; the null hypothesis of no statistically significant difference between probationary and nonprobationary teachers with regard to marking practices was rejected.

Consideration given to classroom performance is indicated by the data in Table 23. As expected, very few of the

TABLE 23

CONSIDERATION GIVEN TO CLASSROOM PERFORMANCE

Questionnaire Items		PROBATIONARY				NONPROBATIONARY				COMBINED TOTAL				Grand Total
		Enql	SS	Math	Sci	Enql	SS	Math	Sci	Enql	SS	Math	Sci	
Sole	F													
	%													
Major	F	4	2	3	4	12	8	7	10	16	10	10	14	50
	%	20	10	15	20	60	40	35	50	40	25	25	35	31.3
Moderate	F	11	13	9	9	6	8	5	4	17	21	14	13	65
	%	55	65	45	45	30	40	25	20	42.5	52.5	35	32.5	40.6
Minor	F	5	5	7	6	2	4	8	3	7	9	15	9	40
	%	25	25	35	30	10	20	40	15	17.5	22.5	37.5	22.5	25
None	F			1	1				3			1	4	5
	%			5	5				15			2.5	10	3.1
ANALYSIS OF VARIANCE														
Source of Variation						Degrees of Freedom				F-Value				
Probationary and Nonprobationary						F 1,152				7.3743				
Academic Area						F 3,152				1.7558				
Interaction						F 3,152				0.3010				

teachers responded at either extreme on the continuum. The most popular response was the "moderate" category with 40.6 percent of the teachers responding. A clearer picture of the consideration the academic areas give to classroom performance was determined by comparing the distribution between the "major" and "moderate" response alternatives. The combined percentages for these two response categories was between 60 and 82.5 for each of the four academic areas. Combining the "none" and "minor" categories for the academic areas resulted in percentages ranging between 17.5 and 40.

The only percentage differences in Table 23 was in the responses to the two alternatives "major" and "moderate." Nonprobationary teachers all appeared to give extensive consideration to classroom performance as indicated by the fact that between 60 and 90 percent of this group checked either the "major" or "moderate" category; whereas, between 60 and 75 percent of the probationary teachers checked either "major" or "moderate."

The overall analysis of variance indicated no significant difference among academic areas across teacher status ($F = 1.71$). No significant interaction of probationary and nonprobationary teachers was found among the four academic areas ($F = .30$). However, a significant F value of ($F = 7.37$)

demonstrated a significant difference between probationary and nonprobationary teachers. Further analysis using simple effects analysis of variance indicated a significant difference between probationary and nonprobationary English teachers in the consideration they gave to classroom performance ($F = 4.59$). No significant differences were found between probationary and nonprobationary teachers in the areas of social studies ($F = 1.86$), mathematics ($F = .95$), and science ($F = .95$).

There were no significant differences among the academic areas; therefore, the null hypothesis of no statistically significant difference among teachers of the four academic areas with regard to marking practices was accepted.

The nonsignificant F value for interaction indicated that there was uniformity of marking practices among the four academic areas of probationary and nonprobationary teachers; therefore, the null hypothesis of no statistically significant difference in the marking practices of probationary and nonprobationary teachers among the four academic areas was accepted.

In view of the statistically significant difference found between probationary and nonprobationary teachers, the

null hypothesis of no statistically significant difference between probationary and nonprobationary teachers with regard to marking practices was rejected.

Table 24 shows the consideration given to work done in a previous grading period. This is an attempt to determine the extent to which grades reflect an assessment of cumulative performance over an extended period. The results generally indicated that 61.9 percent of the teachers in the study gave "none" or "minor" consideration to work done in previous grading periods. The extent of this practice reflected the degree of importance attributable to the acquisition of basic skills or improvement. Of the remaining teachers in the sample, 28.7 percent checked the "moderate" category, while 9.4 percent checked the "major" response.

All of the F values in Table 24 were nonsignificant. Therefore, in the case of the two main effects, evidence demonstrated that probationary and nonprobationary teachers did not differ in their marking practices ($F = .73$), nor did teachers in the four academic areas differ in the consideration given to work done in a previous grading period ($F = .16$). The nonsignificant F value for interaction ($F = 2.07$) was interpreted to mean that there was uniformity of

TABLE 24

CONSIDERATION GIVEN TO WORK DONE IN A
PREVIOUS GRADING PERIOD

Questionnaire Items		PROBATIONARY				NONPROBATIONARY				COMBINED TOTAL				Grand Total
		Enql	SS	Math	Sci	Enql	SS	Math	Sci	Enql	SS	Math	Sci	
Sole	F													
	%													
Major	F	1	1	3	1	3	3	1	2	4	4	4	3	15
	%	5	5	15	5	15	15	5	10	10	10	10	7.5	9.4
Moderate	F	4	5	5	9	6	6	5	6	10	11	10	15	46
	%	20	25	25	45	30	30	25	30	25	27.5	25	37.5	28.7
Minor	F	3	6	7	3	7	3	5	6	10	9	12	9	40
	%	15	30	35	15	35	15	25	30	25	22.5	30	22.5	25
None	F	12	8	5	7	4	8	9	6	16	16	14	13	59
	%	60	40	25	35	20	40	45	30	40	40	35	32.5	36.9
<u>ANALYSIS OF VARIANCE</u>														
Source of <u>Variation</u>						Degrees of <u>Freedom</u>				<u>F-Value</u>				
Probationary and														
Nonprobationary						F 1,152				0.7390				
Academic Area						F 3,152				0.1690				
Interaction						F 3,152				2.0745				

marking practices found among the four areas of probationary and nonprobationary teachers.

Since no significant F values were found among the sources of variation, the three null hypotheses of no significant difference were accepted.

Table 25 presents the teachers' responses to the question, "With what frequency do you typically record a grade for a pupil in your class record book?" The results generally indicated that 41 percent of the teacher sample recorded a grade twice a week for their pupils. Recording one grade a week tended to be done by an appreciable proportion (27.5%) of teachers in the sample. One of the percentage differences in Table 25 was in the responses to the "three times a week" category. Five to 10 percent of the nonprobationary social studies and science teachers responded to this category, while 20 and 35 percent of the nonprobationary English and mathematics teachers checked this response. A similar situation exists in the category "once a week." While only 10 percent of the nonprobationary English teachers responded to this category, between 30 and 50 percent of the nonprobationary social studies, mathematics and science teachers checked this response.

TABLE 25

FREQUENCY WITH WHICH A GRADE IS RECORDED IN THE
CLASS RECORD BOOK

Questionnaire Items		PROBATIONARY				NONPROBATIONARY				COMBINED TOTAL				Grand Total
		Engl	SS	Math	Sci	Engl	SS	Math	Sci	Engl	SS	Math	Sci	
Daily	F	1	1	2		2	1	2		3	2	4		9
	%	5	5	10		10	5	10		7.5	5	10		5.6
Four times a week	F	2		1	1	2		2	2	4		3	3	10
	%	10		5	5	10		10	10	10		7.5	7.5	6.2
Three times a week	F	4	3	5	5	7	1	4	2	11	4	9	7	31
	%	20	15	25	25	35	5	20	10	27.5	10	22.5	17.5	19.4
Twice a week	F	9	9	9	9	7	8	5	10	16	17	14	19	66
	%	45	45	45	45	35	40	25	50	40	42.5	35	47.5	41.3
Once a week	F	4	7	3	5	2	10	7	6	6	17	10	11	44
	%	20	35	15	25	10	50	35	30	15	42.5	25	27.5	27.5
<u>ANALYSIS OF VARIANCE</u>														
Source of Variation Probationary and Nonprobationary Academic Area Interaction						Degrees of Freedom				F-Value				
						F 1,152				0.0220				
						F 3,152				3.9363				
						F 3,152				0.7404				

The overall analysis of variance revealed a significant difference among teachers in the academic areas ($F = 3.93$). However, no significant difference between probationary and nonprobationary teachers was revealed ($F = .02$), nor was there a significant interaction ($F = .74$).

In order to determine the location of variability among academic areas, a simple effects analysis of variance was done. No significant differences were found among academic areas of probationary teachers again. However, a significant difference among academic areas for nonprobationary teachers was revealed ($F = 3.62$).

The contrast analysis of variance taking individual comparison among academic areas for nonprobationary teachers revealed a significant difference only between English and social studies teachers ($F = 3.24$). No significant differences were indicated between English and mathematics ($F = 1.24$); between social studies and science ($F = .26$); and between mathematics and science ($F = .36$).

In view of the significant difference among the teachers of the four academic areas, the null hypothesis of no statistically significant difference among teachers of the four academic areas with regard to marking practices was rejected.

There were no significant differences found between probationary and nonprobationary teachers; therefore, the null hypothesis of no statistically significant difference between probationary and nonprobationary teachers among the four academic areas with regard to marking practices was accepted.

The nonsignificant F value for interaction means that there was uniformity of marking practices found among the four academic areas of probationary and nonprobationary teachers; therefore, the null hypothesis of no statistically significant difference in the marking practices of probationary and nonprobationary teachers among the four academic areas was accepted.

The next question was concerned with the consideration given to final examination results in determining final grades. The appropriate data are presented in Table 26. The most popular response for all groups was the "moderate" category with 40 percent of the teacher sample checking this response. A clearer picture of the consideration the teachers in the different academic areas gave to final examinations was determined by comparing the combined percentage distributions of the two response alternatives "moderate" and "minor." Seventy-five percent of the English,

TABLE 26

CONSIDERATION GIVEN TO FINAL EXAMINATIONS

Questionnaire Items		PROBATIONARY				NONPROBATIONARY				COMBINED TOTAL				Grand Total
		Enql	SS	Math	Sci	Enql	SS	Math	Sci	Enql	SS	Math	Sci	
Sole	F													
	%													
Major	F	1	2	1	1	2	1	2	1	3	3	3	2	11
	%	5	10	5	5	10	5	10	5	7.5	7.5	7.5	5	6.9
Moderate	F	9	8	6	11	8	4	7	11	17	12	13	22	64
	%	45	40	30	55	40	20	35	55	42.5	30	32.5	55	40
Minor	F	6	8	7	6	7	10	10	4	13	18	17	10	58
	%	30	40	35	30	35	50	50	20	32.5	45	42.5	25	36.2
None	F	4	2	6	2	3	5	1	4	7	7	7	6	27
	%	20	10	30	10	15	25	5	20	17.5	17.5	17.5	15	16.9
ANALYSIS OF VARIANCE														
Source of Variation						Degrees of Freedom				F-Value				
Probationary and Nonprobationary						F 1,152				0.0088				
Academic Area						F 3,152				0.5944				
Interaction						F 3,152				1.7890				

social studies and mathematics teachers checked the "moderate" or "minor" response alternatives, while 80 percent of the science teachers responded to these two categories. Of the remaining teachers in the sample, 6.9 percent responded to the "major" category, while 16.9 percent checked the "none" response.

All of the F values in Table 26 were nonsignificant. Therefore, in the case of the two main effects, the probationary and nonprobationary teachers did not differ in their marking practices ($F = .008$), nor did teachers in the four academic areas differ in the consideration given to final examinations ($F = .59$). The nonsignificant F value for interaction ($F = 1.78$) was interpreted to mean that there was uniformity of marking practices found among the four academic areas of probationary and nonprobationary teachers.

Since no significant F values were found among the sources of variation, the three null hypotheses of no statistically significant difference were accepted.

Table 27 presents the responses to the question, "With what frequency do you grade homework for quality?" Fifteen percent of the teachers in the sample did not grade homework at all. At the other end of the continuum, over

TABLE 27

FREQUENCY WITH WHICH HOMEWORK IS GRADED FOR QUALITY

Questionnaire Items		PROBATIONARY				NONPROBATIONARY				COMBINED TOTAL				Grand Total
		Engl	SS	Math	Sci	Engl	SS	Math	Sci	Engl	SS	Math	Sci	
Not at all	F	4	3	3	2	2	8	1	1	6	11	4	3	24
	%	20	15	15	10	10	40	5	5	15	27.5	10	7.5	15
Once a month	F	3	3	1			1	1	1	3	4	2	1	10
	%	15	15	5			5	5	5	7.5	10	5	2.5	6.3
Once every two weeks	F	1	6	2	4	6	3		6	7	9	2	10	28
	%	5	30	10	20	30	15		30	17.5	22.5	5	25	17.5
Once or twice a week	F	11	8	9	13	8	7	10	11	19	15	19	24	77
	%	55	40	45	65	40	35	50	55	47.5	37.5	47.5	60	48.1
Daily	F	1		5	1	4	1	8	1	5	1	13	2	21
	%	5		25	5	20	5	40	5	12.5	2.5	32.5	5	13.1
<u>ANALYSIS OF VARIANCE</u>														
<u>Source of</u> <u>Variation</u>						<u>Degrees of</u> <u>Freedom</u>				<u>F-Value</u>				
Probationary and						F 1,152				0.7585				
Nonprobationary						F 3,152				6.0610				
Academic Area						F 3,152				1.3690				
Interaction														

13 percent of the teacher sample graded homework on a "daily" basis. The most popular item was the "once or twice a week" category, with 48.1 percent of the teacher sample checking this response. One of the percentage differences in Table 27 was in the responses to the "not at all" category. Five to 10 percent of the nonprobationary English, mathematics and science teachers responded to this category, while 40 percent of the nonprobationary social studies teachers checked this response. A similar situation existed in the "daily" category. While 5 percent of the nonprobationary social studies and science teachers responded to this category, between 20 and 40 percent of the nonprobationary English and mathematics teachers checked this response.

The overall two way analysis of variance revealed no significant difference between probationary and nonprobationary teachers ($F = .75$). The analysis of variance did not reveal a significant interaction ($F = 1.36$). However, a significant difference was found among the academic areas ($F = 6.06$). Further analysis using simple effects analysis revealed no significant differences among academic areas for probationary teachers. The simple effects analysis did reveal, however, a significant F value

($F = 5.92$) across academic areas for nonprobationary teachers.

Individual comparisons using contrast analysis of variance indicated a significant difference for nonprobationary teachers of social studies in comparison with mathematics teachers ($F = 5.76$). The contrast analysis revealed no significant difference between English and social studies teachers ($F = 2.39$); between English and mathematics teachers ($F = .72$); between English and science teachers ($F = .23$); between social studies and science teachers ($F = 1.94$); and between mathematics and science teachers ($F = 1.01$).

There were no significant differences found between probationary and nonprobationary teachers; therefore, the null hypothesis of no statistically significant difference between probationary and nonprobationary teachers with regard to the frequency which homework is graded was accepted.

In view of the significant difference among teachers of the four academic areas, the null hypothesis of no statistically significant difference among teachers of the four academic areas with regard to the frequency with which homework is graded was rejected.

The nonsignificant F value for interaction indicated uniformity of marking practices found among the four

academic areas of probationary and nonprobationary teachers; therefore, the null hypothesis of no statistically significant difference in the marking practices of probationary and nonprobationary teachers among the four academic areas was accepted.

The data in Table 28 describes the frequency with which quizzes were given. (A quiz is a test of less than 15 minutes duration). Over 15 percent of the teachers in the sample reported that they did not give quizzes at all. At the other end of the continuum, less than one percent reported that they gave quizzes on a "daily" basis. The most popular item was the "once a week" category, with 42.5 percent of the teacher sample checking this response. Of the remaining teachers in the sample, 24.4 percent checked the "once every two weeks" category, while 16.9 percent checked the "once a month" response.

The overall analysis of variance revealed a significant difference among teachers in the academic areas across teacher status ($F = 2.72$). The two way analysis revealed an F value of ($F = .45$) which was not significant for interaction at the .05 level of confidence. Further analysis demonstrated no significant difference between probationary and nonprobationary teachers ($F = 2.27$).

TABLE 28

FREQUENCY WITH WHICH SHORT QUIZZES ARE ADMINISTERED

Questionnaire Items		PROBATIONARY				NONPROBATIONARY				COMBINED TOTAL				Grand Total
		Enql	SS	Math	Sci	Enql	SS	Math	Sci	Enql	SS	Math	Sci	
Not at all	F		2	2	4	2	8	3	4	2	10	5	8	25
	%		10	10	20	10	40	15	20	5	25	12.5	20	15.6
Once a month	F	3	7	4	2	3	3	3	2	6	10	7	5	27
	%	15	35	20	10	15	15	15	10	15	25	17.5	10	16.9
Once every two weeks	F	7	3	6	2	6	5	6	4	13	8	12	6	39
	%	35	15	30	10	30	25	30	20	32.5	20	30	15	24.4
Once a week	F	10	8	7	12	9	4	8	10	19	12	15	22	68
	%	50	40	35	60	45	20	40	50	47.5	30	37.5	55	42.5
Daily	F			1								1		1
	%			5								2.5		.6
<u>ANALYSIS OF VARIANCE</u>														
<u>Source of</u> <u>Variation</u>						<u>Degrees of</u> <u>Freedom</u>				<u>F-Value</u>				
Probationary and						F 1,152				2.2725				
Nonprobationary						F 3,152				2.7261				
Academic Area						F 3,152				0.4587				
Interaction														

Simple effects analysis of variance revealed no significant difference for probationary teachers among academic areas ($F = .69$). However, further single effects analysis of nonprobationary teachers across academic areas revealed an F value of 2.48 which approached significance. (F required for significance at the .05 level of confidence was 2.67).

In order to determine where the significant differences were located, a contrast analysis of variance was conducted on the two academic areas having the greatest numerical difference. However, this analysis revealed no significant difference between English and social studies teachers.

Further contrast analysis of variance was conducted on column totals of academic areas across the probationary - nonprobationary variable in order to further determine where the differences were located. As a result of this analysis, an F ratio of 2.50 was obtained which approached significance. (F required for significance at the .05 level of confidence was 2.67). Since the contrast analysis of variance did not reveal a significant F value, further analysis using t tests was conducted. The t test revealed a significant t value of 2.24 (with 18 degrees of freedom

a t value of 2.10 was required at the .05 level of confidence).

There were no significant differences found between probationary and nonprobationary teachers; therefore, the null hypothesis of no statistically significant difference between probationary and nonprobationary teachers with regard to marking practices was accepted.

In view of the significant difference among teachers of the four academic areas, the null hypothesis of no statistically significant difference among teachers of the four academic areas with regard to marking practices was rejected.

The nonsignificant F value for interaction indicated the uniformity of marking practices found among the four academic areas of probationary and nonprobationary teachers; therefore, the null hypothesis of no statistically difference in the marking practices of probationary and nonprobationary teachers among the four academic areas was accepted.

Table 29 presented the responses to the question, "With what frequency do you typically give longer tests to your students?" (A test was defined as of more than 15 minutes duration, but not longer than one class period). Four percent of the teachers in the sample did not give tests

TABLE 29

FREQUENCY WITH WHICH LONGER TESTS ARE ADMINISTERED

Questionnaire Items		PROBATIONARY				NONPROBATIONARY				COMBINED TOTAL				Grand Total
		Enql	SS	Math	Sci	Enql	SS	Math	Sci	Enql	SS	Math	Sci	
Not at all	F		1		2		2		2		3		4	7
	%		5		10		10		10		7.5		10	4.4
1 or 2 per semester	F	5	1	3		1	1	1	2	6	2	4	2	14
	%	25	5	15		5	5	5	10	15	5	10	5	8.8
3 or 4 per semester	F	4	2	1	3	11	4	1	4	15	6	2	7	30
	%	20	10	5	15	55	20	5	20	37.5	15	5	17.5	18.7
5 or 6 per semester	F	8	9	8	4	2	4	6	3	10	13	14	7	44
	%	40	45	40	20	10	20	30	15	25	32.5	35	17.5	27.5
More than 6 per semester	F	3	7	8	11	6	9	12	9	9	16	20	20	65
	%	15	35	40	55	30	45	60	45	22.5	40	50	50	40.6
<u>ANALYSIS OF VARIANCE</u>														
Source of Variation						Degrees of Freedom				F-Value				
Probationary and Nonprobationary						F 1,152				0.0190				
Academic Area						F 3,152				2.4952				
Interaction						F 3,152				0.8698				

at all. At the other end of the continuum, over 40 percent of the teacher sample gave more than six tests per semester. A further inspection of the raw data table revealed that teachers of social studies, mathematics and science reported more frequent use of tests than did teachers of English. Teachers of English were also much more variable in the frequency of their testing.

All of the F values in Table 29 were nonsignificant. Therefore, in the case of the two main effects, probationary and nonprobationary teachers did not differ in their marking practices ($F = 2.44$), nor did teachers in the four academic areas differ in their marking practices ($F = 1.39$). The nonsignificant F value for interaction ($F = .86$) indicated the uniformity of marking practices found among the four academic areas of probationary and nonprobationary teachers.

Since no significant F values were found among the sources of variation, the three null hypotheses of no statistically significant difference were accepted.

Data on the percent of students the teacher individually discussed grades with during the school term are presented in Table 30. Over 6 percent of the teachers in the sample reported that they discussed grades with "less

TABLE 30

PERCENT OF STUDENTS WITH WHICH GRADES WERE
DISCUSSED ON AN INDIVIDUAL BASIS

Questionnaire Items		PROBATIONARY				NONPROBATIONARY				COMBINED TOTAL				Grand Total
		Enql	SS	Math	Sci	Enql	SS	Math	Sci	Enql	SS	Math	Sci	
More than 50	F	4	4	2	8	3	3	6	4	7	7	8	12	34
	%	20	20	10	40	15	15	30	20	17.5	17.5	20	30	21.3
Between 20 and 50	F	8	6	9	4	6	5	6	7	14	11	15	11	51
	%	40	30	45	20	30	25	30	35	35	27.5	37.5	27.5	31.9
Between 10 and 20	F	4	4	6	4	7	3	4	6	11	7	10	10	38
	%	20	20	30	20	35	15	20	30	27.5	17.5	25	25	23.7
Between 5 and 10	F	3	5	3	3	2	6	3	2	5	11	6	5	27
	%	15	25	15	15	10	30	15	10	12.5	27.5	15	12.5	16.9
Less than 5	F	1	1		1	2	3	1	1	3	4	1	2	10
	%	5	5		5	10	15	5	5	7.5	10	2.5	5	6.2
<u>ANALYSIS OF VARIANCE</u>														
<u>Source of</u> <u>Variation</u>					<u>Degrees of</u> <u>Freedom</u>					<u>F-Value</u>				
Probationary and					F 1,152					0.8746				
Nonprobationary					F 3,152					1.3922				
Academic Area					F 3,152					0.3867				
Interaction														

than 5 percent" of their students. At the other end of the continuum, 21.3 percent reported that they discussed grades with "more than 50 percent" of their students. The most popular item was the "between 20 and 50" category, with 31.9 percent of the teacher sample checking this response. Of the remaining teachers in the sample, 23.7 percent checked the "between 10 and 20" category, while 16.9 percent checked the "between 5 and 10" response.

All of the F values in Table 30 were nonsignificant. Therefore, in the case of the two main effects, probationary and nonprobationary teachers did not differ in their marking practices ($F = .87$), nor did the teachers in the four academic areas differ in their marking practices ($F = 1.39$). The nonsignificant F value for interaction ($F = .38$) was interpreted to mean that there was uniformity of marking practices among the four academic areas of probationary and nonprobationary teachers.

Since no significant F values were found among the sources of variation, the three null hypotheses of no statistically significant difference were accepted.

Table 31 described responses to the question, "When computing pupil's grades, what consideration do you give to special distinction between classes organized on the

TABLE 31

CONSIDERATION GIVEN TO HOMOGENEOUS GROUPING

Questionnaire Items		PROBATIONARY				NONPROBATIONARY				COMBINED TOTAL				Grand Total
		Enql	SS	Math	Sci	Enql	SS	Math	Sci	Enql	SS	Math	Sci	
Sole	F		1	1	1	1			1	1	1	1	2	5
	%		5	5	5	5			5	2.5	2.5	2.5	5	3.1
Major	F	13	5	7	7	12	6	5	5	25	11	12	12	60
	%	65	25	35	35	60	30	25	25	62.5	27.5	30	30	37.5
Moderate	F	4	8	8	5	5	8	8	11	9	16	16	16	57
	%	20	40	40	25	25	40	40	55	22.5	40	40	40	35.6
Minor	F		3	3	3	1	3	4	2	1	6	7	5	19
	%		15	15	15	5	15	20	10	2.5	15	17.5	12.5	11.9
None	F	3	3	1	4	1	3	3	1	4	6	4	5	19
	%	15	15	5	20	5	15	15	5	10	15	10	12.5	11.9
<u>ANALYSIS OF VARIANCE</u>														
<u>Source of Variation</u>						<u>Degrees of Freedom</u>				<u>F-Value</u>				
Probationary and						F 1,152				0.0058				
Nonprobationary						F 3,152				2.3971				
Academic Area						F 3,152				0.9530				
Interaction														

basis of ability or achievement levels?" The results indicated that 73.1 percent of the teacher sample gave either "moderate" or "major" consideration to special distinction between classes organized on the basis of ability or achievement levels. There were percentage differences, however, when comparisons were made among the four academic areas. Twenty-seven to 30 percent of the social studies, mathematics and science teachers responded to the "major" category, while over 60 percent of the English teachers checked this response. A similar situation existed in the "moderate" category. Forty percent of the social studies, mathematics and science teachers checked the "moderate" response, while 22.5 percent of the English teachers responded to this category. It should be noted that 23.8 percent of the teacher sample indicated that they gave "minor" or "no" consideration to special distinction between classes when computing grades.

None of the F values in Table 31 were significant. Therefore, in the case of the two main effects, probationary and nonprobationary teachers did not differ in their marking practices ($F = .008$), nor did teachers in the four academic areas differ in their marking practices ($F = 2.39$). The nonsignificant F value for interaction ($F = .95$) indicated

the uniformity of marking practices found among the four academic areas of probationary and nonprobationary teachers.

Since no significant F values were found among the sources of variation, the three null hypotheses of no statistically significant difference were accepted.

Responses to the question concerning the consideration given to special weighting factors related to computing pupil grades in courses where ability or achievement level grouping exists are presented in Table 32. Less than 1 percent of the teachers in the sample reported that they gave "sole" consideration to special weighting factors when computing grades. At the other end of the continuum, 11.3 percent of the teacher sample checked the "none" response. The most popular item was the "moderate" category, with 40.6 percent of the teachers checking this response.

It should be noted that some percentage differences were present when comparisons were made among the teachers in the four academic areas. Twenty-two to 25 percent of the social studies, mathematics and science teachers responded to the "major" category, while over 40 percent of the English teachers checked this category. It should be noted that there were certain parallels in the responses

TABLE 32

CONSIDERATION GIVEN TO SPECIAL WEIGHTING PROCEDURES

Questionnaire Items		PROBATIONARY				NONPROBATIONARY				COMBINED TOTAL				Grand Total
		Enql	SS	Math	Sci	Enql	SS	Math	Sci	Enql	SS	Math	Sci	
Sole	F					1				1				1
	%					5				2.5				.6
Major	F	6	6	5	6	11	3	4	4	17	9	9	10	45
	%	30	30	25	30	55	15	20	20	42.5	22.5	22.5	25	28.1
Moderate	F	9	10	9	5	5	11	7	9	14	21	16	14	65
	%	45	50	45	25	25	55	35	45	35	52.5	40	35	40.6
Minor	F	1	3	5	7	1	2	6	6	2	5	11	13	31
	%	5	15	25	35	5	10	30	30	5	12.5	27.5	32.5	19.4
None	F	4	1	1	2	2	4	3	1	6	5	4	3	18
	%	20	5	5	10	10	20	15	5	15	12.5	10	7.5	11.3
ANALYSIS OF VARIANCE														
Source of Variation						Degrees of Freedom				F-Value				
Probationary and Nonprobationary						F 1,152				0.0271				
Academic Area						F 3,152				1.2811				
Interaction						F 3,152				1.9938				

to the "minor" and "none" categories presented here and in Table 31. Twenty-three percent of the teacher sample responded to these two categories in Table 28, while 30.7 percent of the teacher sample responded to these two categories in Table 32.

None of the F values in Table 32 were significant. Therefore, in the case of the two main effects, evidence indicated that probationary and nonprobationary teachers did not differ in their marking practices ($F = .02$), nor did teachers in the four academic areas differ in their marking practices ($F = 1.28$). The nonsignificant F value for interaction ($F = 1.99$) indicated the uniformity of marking practices found among the four academic areas of probationary and nonprobationary teachers.

Since no significant F values were found among the sources of variation, the three null hypotheses of no statistically significant difference were accepted.

Table 33 was concerned with the methods used in determining the distribution of grades. The results showed that 33.1 percent of the teacher sample used a "compromise procedure" when determining the distribution of grades in their classes. The use of an "absolute standard" was reported by 23.1 percent of the teachers in the sample. Of

TABLE 33

METHODS USED WHEN DETERMINING THE DISTRIBUTION OF
STUDENT GRADES

Questionnaire Items		PROBATIONARY				NONPROBATIONARY				COMBINED TOTAL				Grand Total
		Enql	SS	Math	Sci	Enql	SS	Math	Sci	Enql	SS	Math	Sci	
Other	F	1	5	3	2	3	3	3	3	4	8	6	5	23
	%	5	25	15	10	15	15	15	15	10	20	15	12.5	14.4
Compromise procedure	F	10	4	5	6	10	4	7	7	20	8	12	13	53
	%	50	20	25	30	50	20	35	35	50	20	30	32.5	33.1
Absolute standard	F	6	5	6	5	3	6	4	2	9	11	10	7	37
	%	30	25	30	25	15	30	20	10	22.5	27.5	25	17.5	23.1
Relative marking	F	2	2	5	5	2	3	5	4	4	5	10	9	28
	%	10	10	25	25	10	15	25	20	10	12.5	25	22.5	17.5
Grading on the curve	F	1	4	1	2	2	4	1	4	3	8	2	6	19
	%	5	20	5	10	10	20	5	20	7.5	20	5	15	11.9
<u>ANALYSIS OF VARIANCE</u>														
Source of Variation						Degrees of Freedom				F-Value				
Probationary and Nonprobationary						F 1,152				0.0040				
Academic Area						F 3,152				0.8796				
Interaction						F 3,152				0.1749				

the remaining teachers in the sample, 14.4 percent reported that they used some "other" method, while 17.5 percent checked the "relative marking" response. It should be noted that 11.9 percent of the teachers in the sample responded to the "grading on the curve" category.

There were some percentage differences when comparisons were made among teachers in specific subject areas. Teachers of mathematics, social studies and science reported less frequent use of the "compromise procedure" than did English teachers. English and mathematics teachers indicated less response to "grading on the curve" than did social studies and science teachers.

None of the F values in Table 33 were significant. Therefore, in the case of the two main effects, evidence indicated that probationary and nonprobationary teachers did not differ in their marking practices ($F = .004$), nor did teachers in the four academic areas differ in their marking practices ($F = .87$). The nonsignificant F value for interaction ($F = .17$) indicated the uniformity of marking practices found among the four academic areas of probationary and nonprobationary teachers.

Since no significant F values were found among the sources of variation, the three null hypotheses of no

statistically significant difference were accepted.

Responses to a question concerning the use of weighting formulas for homework, tests, projects, etc., are presented in Table 34. The most popular response for the teacher sample was the "formulas which weigh all factors approximately equal" category, with 37.5 percent of the teachers checking this response. Over 14 percent of the teachers indicated that they used "no formula" when determining grades. Of the remaining teachers in the sample, 36.9 percent indicated they used "formulas which give most emphasis to test scores," while 8.7 percent checked the "formulas which give most emphasis to daily work" response.

There were some percentage differences when comparisons were made among specific subject areas. The raw data indicated that between 32 and 37 percent of the science and social studies teachers used "formulas which give most emphasis to test scores," while 62.5 percent of the mathematics teachers and 15 percent of the English teachers checked this response. A substantial 45 percent of the English and science teachers reported that they used "formulas which weigh all factors approximately equal," while between 27 and 32 percent of the mathematics and social studies teachers indicated they used this formula. The only four

TABLE 34

FORMULAS USED WHEN WEIGHTING SUCH FACTORS
AS HOMEWORK, TESTS, PROJECTS, ETC.

Questionnaire Items	PROBATIONARY				NONPROBATIONARY				COMBINED TOTAL				Grand Total
	Enql	SS	Math	Sci	Enql	SS	Math	Sci	Enql	SS	Math	Sci	
No formula used	F 2 % 10	5 25		3 15	6 30	4 20	2 10	1 5	8 20	9 22.5	2 5	4 10	23 14.4
Formulas which give most to other factors	F %				1 5		1 5	2 10	1 2.5		1 2.5	2 5	4 2.5
Formulas which give most to test scores	F 4 % 20	9 45	16 80	8 40	2 10	6 30	9 45	5 25	6 15	15 37.5	25 62.5	13 32.5	59 36.9
Formulas which give most to daily work	F 5 % 25	1 5		2 10	2 10	2 10	1 5	1 5	7 17.5	3 7.5	1 2.5	3 7.5	14 8.7
Formulas which weigh all factors approx. equal	F 9 % 45	5 25	4 20	7 35	9 45	8 40	7 35	11 55	18 45	13 32.5	11 27.5	18 45	60 37.5
<u>ANALYSIS OF VARIANCE</u>													
<u>Source of Variation</u>					<u>Degrees of Freedom</u>				<u>F-Value</u>				
Probationary and Nonprobationary					F 1,152				0.2074				
Academic Area					F 3,152				0.8554				
Interaction					F 3,152				1.2701				

responses to the category "formulas which give most emphasis to other factors" occurred in the areas of English, mathematics and science.

All of the F values in Table 34 were nonsignificant. Therefore, in the case of the two main effects, probationary and nonprobationary teachers did not differ in their marking practices ($F = .20$), nor did teachers of the four academic areas differ in their marking practices ($F = .85$). The nonsignificant F value for interaction ($F = 1.27$) indicated the uniformity of marking practices found among the four academic areas of probationary and nonprobationary teachers.

Since no significant F values were found among the sources of variation, the three null hypotheses of no statistically significant difference were accepted.

The data in Table 35 reflected the general attitudes of teachers toward grades. Between 32.5 and 57.5 percent of the teachers in the four academic areas indicated that "grades are of some value to the teacher and student." Over 38 percent of the teachers in the sample expressed a negative attitude ("abolish" or "necessary evil") about grades. Responses to the "great value" category included 16.2 percent of the teachers.

TABLE 35

TEACHER ATTITUDES TOWARD SCHOOL GRADES

Questionnaire Items		PROBATIONARY				NONPROBATIONARY				COMBINED TOTAL				Grand Total
		Enql	SS	Math	Sci	Enql	SS	Math	Sci	Enql	SS	Math	Sci	
Necessary	F	9	5	3	5	6	3	9	11	15	8	12	16	51
evil	%	45	25	15	25	30	15	45	55	37.5	20	30	40	31.9
Great value	F	2	3	1	4	3	4	4	5	5	7	5	9	26
	%	10	15	5	20	15	20	20	25	12.5	17.5	12.5	22.5	16.2
Should	F	3	2		2	3	1			6	3		2	11
abolish	%	15	10		10	15	5			15	7.5		5	6.9
Some value	F	6	10	16	9	8	12	7	4	14	22	23	13	72
	%	30	50	80	45	40	60	35	20	35	55	57.5	32.5	45
Unconcerned	F													
	%													
<div> <div>Source of Variation</div> <div>Probationary and Nonprobationary Academic Area Interaction</div> </div> <div> <div>ANALYSIS OF VARIANCE</div> <div>Degrees of Freedom</div> </div> <div> <div>F-Value</div> <div>F 1,152</div> <div>F 3,152</div> <div>F 3,152</div> </div> <div> <div>3.4742</div> <div>2.4602</div> <div>3.8551</div> </div>														

The overall analysis of variance revealed a significant difference between probationary and nonprobationary teachers across academic areas ($F = 3.47$) and a significant interaction with an F value of ($F = 3.85$). However, no overall significant differences were found among the academic areas.

A simple effects analysis of variance revealed a significant difference between probationary and nonprobationary mathematics teachers ($F = 8.90$) and between probationary and nonprobationary science teachers ($F = 5.03$). However, no significant differences were found between probationary and nonprobationary English teachers ($F = .76$) and probationary and nonprobationary social studies teachers ($F = .39$).

There were significant differences found between probationary and nonprobationary teachers; therefore, the null hypothesis of no statistically significant difference between probationary and nonprobationary teachers with regard to the teacher attitudes concerning school grades was rejected.

In view of the no significant difference among teachers of the four academic areas, the null hypothesis of no statistically significant difference among teachers

of the four academic areas with regard to the attitude of teachers concerning school grades was accepted.

The significant F value for interaction indicated variability of marking practices found among the four academic areas of probationary and nonprobationary teachers; therefore, the null hypothesis of no statistically significant difference in the marking practices of probationary and nonprobationary teachers among the four academic areas was rejected.

Table 36 identifies what the teacher feels is the "best" student attitude toward grades. A large majority of the teachers in the sample indicated that "high grades should be one of several major goals." Over 73 percent of the total teacher sample in the four academic areas selected this response. Of the remaining teachers in the sample, more would prefer (13.7 percent) to have grades be "the primary goal" than to be "a minor goal" or "no goal."

All of the F values in Table 36 were nonsignificant. Therefore, in the case of the two main effects, probationary and nonprobationary teachers did not differ in their marking practices ($F = .37$), nor did teachers in the four academic areas differ in their marking practices ($F = .27$). The nonsignificant F value for interaction ($F = .70$) was

TABLE 36

THE ATTITUDE HELD BY STUDENTS TOWARD GRADES WHICH IS MOST
CONDUCTIVE TO DERIVING THE MAXIMUM BENEFIT
THE TEACHERS' SUBJECT

Questionnaire Items		PROBATIONARY				NONPROBATIONARY				COMBINED TOTAL				Grand Total
		Enql	SS	Math	Sci	Enql	SS	Math	Sci	Enql	SS	Math	Sci	
No goal	F	2	1	1			1	3		2	2	4		8
	%	10	5	5			5	15		5	5	10		5
One of sev- eral major goals	F	14	15	13	16	15	17	13	15	29	32	26	31	118
	%	70	75	65	80	75	85	65	75	72.5	80	65	77.5	73.8
Minor goal	F	2		1	2	3	1		3	5	1	1	5	12
	%	10		5	10	15	5		15	12.5	2.5	2.5	12.5	7.5
Primary goal	F	2	4	5	2	2	1	4	2	4	5	9	4	22
	%	10	20	25	10	10	5	20	10	10	12.5	22.5	10	13.7
No opinion	F													
	%													
ANALYSIS OF VARIANCE														
Source of Variation						Degrees of Freedom				F-Value				
Probationary and Nonprobationary						F 1,152				0.3734				
Academic Area						F 3,152				0.2766				
Interaction						F 3,152				0.7052				

interpreted to mean that there was uniformity of marking practices found among the four academic areas of probationary and nonprobationary teachers.

Since no significant F values were found among the sources of variation, the three null hypotheses of no statistically significant difference were accepted.

Teachers were asked to rate the importance of grading as a part of the teacher's job. The results are tabulated in Table 37. Forty-three percent of the teachers in the sample indicated that the consideration they attached to evaluation as a part of the job of the teachers was of "major importance," while 35.6 percent of the total sample checked the "moderate importance" category. Of the remaining teachers in the sample, 18.1 percent indicated "some importance," while 2.5 percent responded to the category "minor importance."

All of the F values in Table 37 were nonsignificant. In the case of the two main effects, probationary and nonprobationary teachers did not differ in their attitudes about evaluation as a part of the job of the teacher ($F = 1.59$), nor did teachers in the four academic areas differ in their attitudes about evaluation as a part of the job of the teacher ($F = 1.65$). The nonsignificant F value for

TABLE 37

IMPORTANCE OF EVALUATION AS A PART OF THE JOB OF THE TEACHER

Questionnaire Items		PROBATIONARY				NONPROBATIONARY				COMBINED TOTAL				Grand Total
		Engl	SS	Math	Sci	Engl	SS	Math	Sci	Engl	SS	Math	Sci	
Minor importance	F	1		2			1			1	1	2		4
	%	5		10			5			2.5	2.5	5		2.5
Major importance	F	12	6	8	12	8	5	10	9	20	11	18	21	70
	%	60	30	40	60	40	25	50	45	50	27.5	45	52.5	43.8
Some importance	F	2	5	3	3	5	4	2	5	7	9	5	8	29
	%	10	25	15	15	25	20	10	25	17.5	22.5	12.5	20	18.1
Moderate importance	F	5	9	7	5	7	10	8	6	12	19	15	11	57
	%	25	45	35	25	35	50	40	30	30	47.5	37.5	27.5	35.6
None	F													
	%													

<u>ANALYSIS OF VARIANCE</u>		
Source of Variation	Degrees of Freedom	F-Value
Probationary and Nonprobationary	F 1,152	1.5993
Academic Area	F 3,152	1.6562
Interaction	F 3,152	0.3104

interaction ($F = .31$) indicated the uniformity of teacher attitudes about evaluation as a part of the job of the teacher among the four academic areas of probationary and nonprobationary teachers.

Since no significant F values were found among the sources of variation, the three null hypotheses of no statistically significant difference were accepted.

CHAPTER V

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

The purpose of this study was to determine the marking practices of teachers in the public secondary schools of Oklahoma City and to analyze the sources of variability in those marking practices. The investigator attempted to present the findings in the light of both descriptive and statistical analysis. The descriptive findings were utilized in determining the marking practices of secondary teachers. The statistical analysis was utilized to test the following null hypotheses: (1) There is no statistically significant difference among teachers of the four academic areas of English language arts, mathematics, science and social studies with regard to marking practices; (2) There is no statistically significant difference between probationary and nonprobationary teachers with regard to marking practices; and (3) There is no statistically significant difference in the marking practices of probationary and nonprobationary teachers among the four academic

areas of English language arts, mathematics, science, and social studies.

Summary of Findings

Classification Data

A summary of responses to the questionnaire revealed that 45 percent of the teachers taught in junior high school and 39.4 percent taught in senior high school. Of the remaining teachers in the sample, 15.6 percent taught in both junior-senior high school.

The teacher sample was almost evenly divided according to sex with 48.1 percent being male and 44.4 percent reporting female. The sex category was the only item on the questionnaire to which teachers did not respond 100 percent. Included in this "failure to respond" category were 7.5 percent of the teachers.

This study also found that 66.9 percent of the teachers in the sample were at the bachelor's degree level. Only 1.2 percent of the teacher sample held a special certificate, while 31.9 percent of all teachers surveyed had obtained a master's degree.

The summary of responses further revealed that 45 percent of the teachers in the survey were 30 years of age

or younger. The age category of 31 to 45 years of age contained 30.6 percent of the teachers in the sample, while 24.4 percent were between 46 and 65 years of age.

This study also found that 48.7 percent of the teachers had four years or less of teaching experience. The data further revealed that 20.6 percent of the teachers in the survey had 5 to 10 years teaching experience. Of the remaining teachers in the sample, 16.9 percent had between 11 and 25 years of experience, while 13.5 percent of the sample had 26 years or more of teaching experience.

Questionnaire Data

From the outset, the investigator attempted to present the findings in light of both descriptive and statistical analysis. The descriptive findings were utilized to determine teacher marking practices and the statistical analysis was utilized to make an analysis of the sources of variability in those marking practices.

1. Approximately one-half of the teachers in the sample selected the "achievement with respect to ability" category as the primary basis used in evaluating achievement for grading purposes.

2. Teachers who gave "moderate" consideration to improvement gave "major" consideration to achievement when

determining a pupil's grade. Conversely, those who gave "major" consideration to achievement also gave "moderate" consideration to improvement.

3. Probationary teachers differed significantly with regard to the amount of consideration given to improvement when determining a pupil's grade.

4. A significant difference was found between probationary and nonprobationary teachers with regard to the amount of consideration given to achievement when determining a pupil's grade.

5. Approximately one-half of the teachers in the sample indicated that they gave "moderate" consideration to such factors as judgment of character and citizenship when determining a pupil's grade.

6. When determining a pupil's final grade, more teachers responded to "major" consideration given to absolute measurement than gave "major" consideration to the total process of educational evaluation.

7. Consideration given to absolute measurement revealed a significant difference among probationary teachers in the academic areas of English language arts, mathematics and science.

8. Approximately one-third of the respondents indicated that they gave very little consideration to educational objectives stated in terms of desired behavioral outcomes.

9. A significant difference was found between probationary and nonprobationary teachers regarding the consideration given to course objectives stated in terms of desired behavioral outcomes.

10. Approximately one-third of the respondents reported that they gave very little consideration to course objectives in the evaluation of pupils.

11. English language arts and social studies teachers considered course objectives stated in terms of desired behavioral outcomes and the evaluation of pupils in terms of the achievement of course objectives less important than did mathematics and science teachers.

12. Over 40 percent of the teachers surveyed reported that they gave classroom behavior either "moderate" or "major" consideration when determining a pupil's grade.

13. Approximately one-half of the teachers in the sample indicated that they gave "major" consideration to "effort" when determining a pupil's grade.

14. Approximately 45 percent of the teachers in each academic area reported that they gave unexcused absences and/or tardiness either "moderate" or "major" consideration when determining a pupil's grade.

15. English language arts and mathematics teachers gave more consideration to quiz scores when determining grades at the end of the grading period than did social studies and science teachers.

16. A definite majority of the teachers in each academic field gave either "minor" or "moderate" consideration to homework when determining grades.

17. Probationary teachers all appeared to use tests more extensively than did the nonprobationary teachers.

18. A definite majority of the classroom teachers in the sample reported that they gave "moderate" or "major" consideration to the use of marks as persuasive devices.

19. Probationary teachers tended to use marks as persuasive devices more extensively than did the nonprobationary teachers.

20. At least 60 percent of the teachers in each of the four academic areas reported that they gave "moderate" or "major" consideration to classroom performance when determining pupil's grades.

21. An attempt to determine the extent to which grades reflect an assessment of cumulative performance over an extended period revealed that 61.9 percent of the teachers in the study gave "none" or "minor" consideration to work done in a previous grading period.

22. A significant difference was found among non-probationary teachers regarding the frequency a grade should be recorded for an individual pupil in the class record book.

23. Seventy-five percent of the classroom teachers in each of the four academic areas reported giving "minor" or "moderate" consideration to final examination results in determining a pupil's grade.

24. Approximately one-half of the teachers in the four academic areas reported that they graded homework for quality "once or twice a week."

25. Social studies, mathematics and science teachers reported more frequent use of tests than did teachers of English language arts.

26. Most of the teachers in the study reported that they discussed grades with less than 50 percent of their assigned students.

27. Approximately one-fourth of the respondents indicated that when computing pupil's grades, they gave "minor" or "no" consideration to special distinction between classes organized on the basis of ability or achievement levels.

28. Twenty-eight percent of the teacher sample reported that they gave "major" consideration to special weighting factors related to grading where ability or achievement level grouping exists. At the other end of the continuum, 30.7 percent reported that they gave "minor" or "no" consideration to special weighting factors.

29. When weighting formulas for homework, tests, projects, etc., were used in determining a pupil's grade, mathematics teachers gave more emphasis to test scores than did social studies, English language arts or science teachers.

30. Nonprobationary teachers expressed more of a negative attitude toward grades than did probationary teachers.

31. Forty-three percent of the teacher sample reported giving "major" consideration to the importance they attach to evaluation as a part of the job of the teacher.

Conclusions

It was concluded that variability in the marking practices of teachers in the four academic areas resulted in the rejection of the following null hypotheses:

1. There is no statistically significant difference between probationary and nonprobationary teachers with regard to consideration given to achievement.
2. There is no statistically significant difference among teachers of the four academic areas with regard to the consideration given to improvement.
3. There is no statistically significant difference among teachers of the four academic areas with regard to consideration given to absolute measurement.
4. There is no statistically significant difference of probationary and nonprobationary teachers among the four academic areas with regard to consideration given to objectives stated in terms of desired behavior outcomes.
5. There is no statistically significant difference among teachers of the four academic areas with regard to consideration given to course objectives in the evaluation of pupils.
6. There is no statistically significant difference between probationary and nonprobationary teachers with

regard to consideration given to test scores.

7. There is no statistically significant difference between probationary and nonprobationary teachers with regard to the consideration given to the use of marks as persuasive devices.

8. There is no statistically significant difference between probationary and nonprobationary teachers with regard to the consideration given to classroom performance.

9. There is no statistically significant difference among teachers of the four academic areas with regard to the frequency a grade is recorded in the teacher's class record book.

10. There is no statistically significant difference among teachers of the four academic areas with regard to the frequency that homework is graded for quality.

11. There is no statistically significant difference among teachers of the four academic areas with regard to the frequency that short quizzes are given to students.

12. There is no statistically significant difference of probationary and nonprobationary teachers among the four academic areas with regard to teachers' attitudes concerning school grades.

Other conclusions reached through descriptive findings and statistical analysis of the data are:

13. The grade a pupil receives appeared to depend to some degree on classroom behavior, unexcused absences and/or tardiness, judgment of character and citizenship, and the amount of effort put forth.

14. A high percentage of teachers did not evaluate pupils in terms of course objectives achieved.

15. Evaluation of pupil progress appeared to some degree to be based on a combination of amount of growth and/or level of achievement.

16. Differentiating between marks in classes having different levels of academic ability appeared to depend to some degree on the academic area of the teacher.

17. Secondary teachers to some degree gave very little consideration to the weighting of grades where ability grouping exists and the fact that pupils vary in individual ability to achieve course objectives.

18. The interrelationship between educational objectives, evaluation procedures and pupil learning theory did not appear to be clearly understood by the teachers.

19. The grades pupils receive appeared to lose their meaning because of a lack of clearly defined marking practices and procedures.

Recommendations

1. In order to establish greater uniformity among teachers in their marking practices and hence in the meaning of marks, it is recommended that a set of guidelines be developed which can be used by school systems to assist teachers in planning for the evaluation of pupils.

2. The wealth of existing literature on the subject of evaluation, learning objectives, purposes, and procedures should be utilized when developing a set of guidelines for pupil evaluation.

3. It is recommended that many instructional objectives be stated in terms of desired pupil behavior.

4. The educational objectives are the basis on which the evaluation procedures should be planned, therefore, it is recommended that pupil evaluation should be based chiefly on the degree to which pupil behavior has changed as defined in the objectives.

5. It is recommended that a method of converting numerical measures of achievement into marks be adopted so as to establish greater uniformity among teachers when determining the distribution of grades in class, and hence make the marks issued much more consistently meaningful.

6. It is recommended that when teachers convert numerical measures of achievement into marks, provisions should be made for classes having different levels of academic ability.

7. Evaluation of pupil achievement appears to have in it many elements of subjectivity, therefore, it is recommended that pupil marks be based on achievement and that they should not reflect deportment, unexcused absences and/or tardiness, effort or judgment of character.

8. It is recommended that continuing efforts should be directed toward a study of the development of a sound theoretical or philosophical basis for the evaluation of pupil achievement.

BIBLIOGRAPHY

Books and Pamphlets

- Ahmann, J. Stanley and Glock, Marvin D. Evaluating Pupil Growth. Boston: Allyn and Bacon, 1963, pp. 566-567.
- Bloom, Benjamin S. et al. Taxonomy of Educational Objectives, Handbook I: Cognitive Domain. David McKay, Inc., 1956.
- Bloom, Benjamin S., Drathwohl, David and Masia, Bertram. Taxonomy of Educational Objectives, Handbook II: Affective Domain. David McKay, Inc., 1964.
- Burton, William H. The Guidance of Learning Activities. New York: Appleton-Century Company, 1944.
- Cella, Francis R. Sampling Statistics in Business and Economics. Norman, Oklahoma: Bureau of Business Research, University of Oklahoma, 1950, pp. 160-163.
- Clark, Bradsley and Haslacher. Utah Educational Association: Poll of Member Opinion. Salt Lake City: Utah Educational Association, January, 1960, pp. 42-67.
- Cronbach, L. J. Educational Psychology. New York: Harcourt, Brace & World, 1962.
- Dearborn, Walter. "Marks and Marking Systems," in Encyclopedia of Educational Research. Ed. Chester W. Harris. New York: Macmillan Company, 1960, p. 784.
- Ebel, Robert L. Measuring Educational Achievement. New Jersey: Prentice-Hall, Inc., 1965, p. 397.

- Fisher, Ronald A., and Yates, Frank. Statistical Tables for Biological, Agricultural and Medical Research. Hafner Publishing Company, Inc., 1953.
- Goode, William J. and Hatt, Paul K. Methods in Special Research. New York: McGraw-Hill Book Company, Inc., 1952, p. 326.
- Greer, Edith S. and Harbeck, Richard M. What High School Pupils Study. U. S. Office of Education, OE-33025 (Bulletin 1962, No. 10). Washington: U. S. Government Printing Office, 1962, pp. 25-36.
- Gronlund, Norman E. Measurement and Evaluation in Teaching. New York: The Macmillan Company, 1966.
- Kelly, Fredrick. "Marks and Marking Systems," in Encyclopedia of Educational Research. Ed. Chester W. Harris. New York: Macmillan Company, 1960, p. 784.
- Kerlinger, Fred N. Foundations of Behavioral Research. New York: Holt, Rhinehart and Winston, Inc., 1964.
- "Marks and Marking Systems," in Encyclopedia of Educational Research. Ed. Chester W. Harris. New York: Macmillan Company, 1960, p. 784.
- Meyer, Max. "Mark and Marking Systems," in Encyclopedia of Educational Research. Ed. Chester W. Harris. New York: Macmillan Company, 1960, p. 784.
- Mouly, George J. The Science of Educational Research. New York: American Book Company, 1963, p. 263.
- National Commission on Life Adjustment Education for Youth. Pupil Appraisal Practices in Secondary Schools. U. S. Office of Education No. 363 GPO: Washington: 1952.
- Odell, Charles. Educational Measurements in High Schools. New York: Century Company, 1930.
- Odell, C. W. "Marks and Marking Systems," in Encyclopedia of Educational Research. Ed. Walter S. Monroe. New York: Macmillan Company, 1959, pp. 711-717.

- Rothney, John W. M. Evaluating and Reporting Pupil Progress: What Research Says to the Teacher. National Education Association Bulletin No. 7, Washington, D. C., 1955, p. 4.
- Ruch, Giles M. Objectives or New Type Examinations; An Introduction to Educational Measurement. New York: Scott, Foresman and Company, 1929, p. 478.
- Rummel, J. Francis. An Introduction to Research Procedures in Education. Harper and Row, Publishers, 1958, pp. 126-27.
- Smith, Eugene R. Appraising and Recording Student Progress. New York: Harper and Brothers, 1942.
- Thomas, Robert Murray. Judging Student Progress. New York: Longmans, Green and Company, 1954, pp. 283-329.
- Thorndike, E. L. An Introduction to the Theory of Mental and Social Measurement, 2nd ed. New York: Teachers College, Columbia University, 1913, Chapter 2.
- Traxler, Arthur. Techniques of Guidance. New York: Harper and Brothers, 1947.
- Van Dalen, Deobold B. Understanding Educational Research. New York: McGraw-Hill Book Company, Inc., 1962, pp. 249-54.
- Wert, James E., Neidt, Charles O. and Ahmann, J. Stanley. Statistical Methods in Educational and Psychological Research. New York: Appleton-Century-Crafts, Inc., 1954, p. 188.
- Wright, Grace S. and Greer, Edith S. The Junior High School. U. S. Office of Education, OE-20046 (Bulletin 1963, No. 32). Washington: U. S. Government Printing Office, 1963, pp. 41-48.
- Wrinkle, William. Improving Marking and Reporting Practices in Elementary and Secondary Schools. New York: Rinehart & Company, Inc., 1947, p. 30.

Periodicals and Articles

- Alphren, Morton. "A Fair Grading System," Clearing House, XXXV (October, 1960), pp. 113-114.
- Austin, David B. "The Need for Marking Systems at the Secondary School Level," The High School Journal, XXXVI (April, 1953), p. 194.
- Barnes, K. F. and Barnes, E. H. "A Realistic Approach to Grading," Clearing House, XXXVI (April, 1962), pp. 476-78.
- Bohlmeier, Edward C. "Principals Pertaining to Marking and Reporting Procedures," School Review, LIX (January, 1951), pp. 15-24.
- Bostrom, John, Ulandix, W., and Rosenbaum. "Grades as Reinforcing Contingencies and Attitude Change," Journal of Educational Psychology, LII (May, 1961), pp. 112-115.
- Cagle, Dan F. "How May We Make The Evaluation and Reporting of Student Achievement More Meaningful?" Bulletin of the National Association of Secondary School Principals, XXXIX (April, 1955), pp. 24-27.
- Campbell, Laurence. "So Pupils May Know," School and Society, XXXII (December, 1930), pp. 726-63.
- Carter, Robert. "How Invalid Are Marks Assigned by Teachers?" Journal of Educational Psychology, XLIII (April, 1952), pp. 218-28.
- Crooks, A. B. "Marks and Marking System; A Digest:" Journal of Educational Research, XXVII (December, 1933), pp. 259-72.
- Cummings, Robert E. "Evaluating and Grading," Education, LXXXII (March, 1962), pp. 403-405.
- Drake, C. E. "Trends in the Field of Evaluating Secondary Education," Educational Administration and Supervision, XXVI (April, 1949), pp. 241-56.

- Fine, Benjamine. "A, B, C of Grading Puzzles Parents," New York Times Magazine, (November 18, 1957), p. 33.
- Hadley, Trever S. "School Mark - Fact or Fancy?" Educational Administration and Supervision, XL (May, 1954), pp. 305-12.
- Halliwell, Joseph W. "The Relationship of Certain Factors to Marking Practices in Individualized Reporting Programs," Journal of Educational Research, LIV (October, 1960), pp. 76-78.
- Hick, Arch. "Contributions of Research to the Classification, Promotion, Marking and Certification of Pupils," Thirty Seventh Yearbook of the National Society for the Study of Education, Part II (Bloomington, Ill.: Public School Publishing Company, 1938), pp. 187-99.
- Jansen, Udo H. "Educators Look At Grading - - - A Survey of Systems," The Texas Outlook, XLV (July, 1961), p. 15.
- Johnson, L. R. "Are There Better Ways of Evaluating, Recording, and Reporting Pupil Progress in the Junior and Senior High School?" Bulletin of the National Association of Secondary School Principals, XXXIV (March, 1959), p. 75.
- Johnson, Mauritz, Jr. "Solving the Mess in Marks," Education Digest, XXVII (February, 1962), pp. 12-14.
- Keller, Irvin A. "A More Comprehensive and Significant Marking System," Bulletin of the National Association of Secondary School Principals, XXXVI (January, 1952), pp. 70-78.
- LaFranchi, Edward H. "High School Marks: Comparative or Individual?" School Executive, LXXI (July, 1952), pp. 51-54.
- Leese, Joseph. "Mess in Marks," New York State Education, XLVIII (May, 1961).

Liggitt, William L. "Are There Better Ways of Evaluating, Recording, and Reporting Pupil Progress in Junior and Senior High Schools?" Bulletin of the National Association of Secondary School Principals, XXXIV (March, 1959), pp. 79-89.

Los Angeles Committee of Secondary School Principal's Association. "Marking Slow Pupils," California Quarterly of Secondary Education, I (June, 1926), pp. 386-91.

Morris, Lucile. "Evaluating and Reporting Pupil Progress," Elementary School Journal, LIII (September, 1958), p. 532.

Morton, Alpert. "A Fair Grading System," Clearing House, (October, 1960), pp. 113-114.

Moskowits, Samuel O. "Guide Lines for Marking Procedures," High Points, XLIV (April, 1962), pp. 26-30.

Palmer, Oliver. "Seven Ways of Grading Dishonestly," English Journal, LI (October, 1962), pp. 464-67.

Peterson, Russell S. "Marking Achievement in Relation to Ability," Bulletin of the National Association of Secondary School Principals, XXXVIII (February, 1954), p. 63.

Quinn, George R. and Szuberla, Charles A. "A Consistent and Equitable Approach to the Evaluation of Student Achievement," Clearing House, (April, 1963), pp. 490-94.

Roelf, R. M. "Trends in Junior High School Reporting," Journal of Educational Research, IXL (December, 1955), pp. 241-49.

Rothney, John W. M. "Evaluating and Reporting Pupil Progress," Research Bulletin of the National Education Association, VII (1965), p. 4.

- Rugg, Harold. "Teacher Marks and Marking Systems," Educational Administration and Supervision, VII (January, 1915), pp. 117-42.
- Russell, James W. "Marks of the Best," Bulletin of the National Association of Secondary School Principals, XLV (February, 1961), pp. 112-114.
- Shannon, John R. "Percentage of Returns of Questionnaires in Reputable Educational Research," Journal of Educational Research, XLII (October, 1948), pp. 138-41.
- Starch, Daniel and Elliott, E. C. "Reliability of Grading Work in History," School Review, XXI (October, 1913), pp. 676-81.
- Starch, Daniel and Elliott, E. C. "Reliability of Grading Work in Mathematics," School Review, XXI (April, 1913), pp. 254-59.
- Strang, Ruth M. "Reporting Pupil Progress," School Executive, LXII (August, 1953), pp. 47-50.
- Terwilliger, James S. "Self Reported Marking Practices and Policies in Public Secondary Schools," Bulletin of the National Association of Secondary School Principals, I (March, 1966), pp. 3-34.
- Thorndike, E. L. and Bregman, E. O. "On The Form of the Distribution of Intellect in the Ninth Grade," Journal of Educational Research, X (November, 1924), pp. 271-278.
- Travers, Robert M. and Gronlund, Norman. "Meaning of Marks," Journal of Higher Education, XXI (October, 1950), pp. 369-74.
- Vredevoe, Lawrence E. "How May We Make the Recording and Reporting of Pupil Achievement More Meaningful?" Bulletin of the National Association of Secondary School Principals, XXXVI (April, 1953), pp. 179-180.
- Wagner, Guy W. "What Schools Are Doing - Reporting Pupils Progress," Education, LXXX (February, 1960), p. 381.

Wakehom, Glen. "Humanizing Grades," School and Society, XXXIV (October, 1931), pp. 596-98.

Weber, Alan G. "Three Track Grading," The Texas Outlook, XLV (July, 1961), p. 45.

Williamson, E. G. "The Cooperative Guidance Program," School Review, XLIII (April, 1935), pp. 273-80.

Yauch, Wilbur A. "What Research Says About School Marks and Their Reporting," National Educational Association Journal, L (May, 1961).

Unpublished Dissertations

Wallace, Joseph. "Critical Incidents in the Assignments of Marks to High School Students." Unpublished Doctoral Dissertation, Department of Education, University of Southern California, Los Angeles, 1960, p. 48.

APPENDIX

TEACHER MARKING PRACTICE QUESTIONNAIRE

SECTION I - GENERAL INFORMATION - PLEASE CIRCLE THE APPROPRIATE NUMBER.

- | | |
|--|---|
| 1. At what level do you teach? | Senior High School..... 1 |
| | Junior High School..... 2 |
| | Both..... 3 |
| 2. Sex | Male..... 1 |
| | Female..... 2 |
| 3. Teaching experience in the Oklahoma City Public Schools. | 0 - 3 Years..... 1 |
| | 4 or More..... 2 |
| 4. Number of college level courses completed in Testing, Measurement and Evaluation. | None..... 1 |
| | One..... 2 |
| | Two..... 3 |
| | Three..... 4 |
| | Four..... 5 |
| | More..... 6 |
| 5. Please indicate the academic area in which you are presently teaching. | English-Language Arts.. 1 |
| | Social Studies..... 2 |
| | Mathematics..... 3 |
| | Science..... 4 |
| 6. Please indicate the highest degree held. | Doctor's Degree..... 1 |
| | Diploma or Specialist's Certificate beyond the Master's Degree..... 2 |
| | Master's Degree..... 3 |
| | Bachelor's Degree..... 4 |

7. Please indicate your age by checking one of the age groups.
- | | |
|--------------|---|
| 20 - 25..... | 1 |
| 26 - 30..... | 2 |
| 31 - 35..... | 3 |
| 36 - 40..... | 4 |
| 41 - 45..... | 5 |
| 46 - 50..... | 6 |
| 51 - 65..... | 7 |
8. Total number years of teaching experience.
- | | |
|--------------|---|
| 0 - 4..... | 1 |
| 5 - 10..... | 2 |
| 11 - 15..... | 3 |
| 16 - 20..... | 4 |
| 21 - 25..... | 5 |
| 26 - 30..... | 6 |
| 30 +..... | 7 |

SECTION II - THE PRIMARY BASIS FOR THE ASSIGNMENT OF MARKS.

Please react to the following statements by circling the word or statement on the continuum which best represents your marking procedures or opinions.

1. Which of the following best represents the primary basis you use in evaluating pupil achievement for grading purposes?

Absolute achievement	Achievement with regard to class
Achievement with regard to ability	Self improvement or growth

2. What consideration do you give to such factors as judgment of character and citizenship in determining a pupil's final grade?

None	Minor	Moderate	Major	Sole
------	-------	----------	-------	------

3. What consideration do you give to the amount of improvement the pupil has made rather than on the level of achievement he has reached when determining the final grade?

None	Minor	Moderate	Major	Sole
------	-------	----------	-------	------

4. What consideration do you give to the level of achievement reached by the pupil rather than the amount of improvement the pupil has made?

None Minor Moderate Major Sole

5. What consideration do you give to the process of total educational evaluation when determining a pupil's final grade? (Evaluation is a process in which a teacher commonly uses information derived from many sources to arrive at a value judgment).

None Minor Moderate Major Sole

6. When determining a final grade for a pupil, what consideration do you give to absolute measurement that concentrates on a specific, well defined trait and tries to characterize a pupil in terms of that trait in as objective a way as possible? (Measurement is the process that attempts to obtain a quantitative representation of the degree to which a trait or skill is possessed by a pupil).

None Minor Moderate Major Sole

7. When determining course objectives, what consideration do you give to objectives that are stated in terms of desired behavior outcomes - that is what the learner should do? (Desired behavior outcomes - knowledge, understandings, attitudes, appreciations, abilities, skills in a specific subject matter, etc.).

None Minor Moderate Major Sole

8. When determining a pupil's final grade, what consideration do you give to the extent to which the course objectives are achieved by the pupil?

None Minor Moderate Major Sole

9. When determining a pupil's final grade, what consideration do you give to the fact that pupils vary in their ability to achieve the objectives of a specific subject matter area?

None Minor Moderate Major Sole

10. What consideration do you give to classroom behavior (e.g. causing distractions or conversely being very cooperative) in determining a pupil's grade at the end of each grading period?

None Minor Moderate Major Sole

11. What consideration do you give to "effort" in determining a pupil's grade at the end of each grading period?

None Minor Moderate Major Sole

12. What consideration do you give to unexcused absences and tardiness in determining a pupil's grade at the end of each grading period?

None Minor Moderate Major Sole

13. What consideration do you give quiz scores in determining grades at the end of each grading period? (A quiz for purposes here is a test less than 15 minutes).

None Minor Moderate Major Sole

14. What consideration do you give to homework in determining grades at the end of each grading period?

None Minor Moderate Major Sole

15. What consideration do you give to test scores in determining grades at the end of each grading period? (A test for purposes here is defined as one more than 15 minutes in duration).

None Minor Moderate Major Sole

16. What consideration do you give to the use of marks as persuasive devices to induce an increased application of student effort?

None Minor Moderate Major Sole

17. What consideration do you give to classroom performance (i.e., recitation, voluntary participation, etc.) in determining grades at the end of each grading period?

None Minor Moderate Major Sole

18. What consideration do you give to work done in previous grading periods when assigning course grades for a grading period just ended?

None Minor Moderate Major Sole

19. With what frequency do you typically record a grade for a pupil in your class record book?

Daily Three times a week

Four times a week Twice a week Once a week

20. What consideration do you give to final examinations (end of semester or year) in determining final grades?

None Minor Moderate Major Sole

21. With what frequency do you typically grade homework for quality?

Daily Once every two weeks

Once a week Once a month Not at all

22. With what frequency do you typically give short quizzes to your students?

Daily Once every two weeks

Once a week Once a month Not at all

23. With what frequency do you typically give longer tests to your students?

More than 6 per semester 3 or 4 per semester

5 or 6 per semester 1 or 2 per semester Not at all

24. With what percent of your students did you individually discuss grades during this school term?

Less than 5 Between 10 and 20

Between 5 and 10 Between 20 and 50 More than 50

25. When computing pupil grades, what consideration do you give to special distinction between classes organized on the basis of ability or achievement levels?

None Minor Moderate Major Sole

26. When computing pupil grades, what consideration do you give to special procedures (i.e., special weighting factors) related to grading in courses where ability or achievement level grouping exists?

None Minor Moderate Major Sole

27. Please indicate which of the following methods you use when determining the distribution of grades in your class.

1. Grading on the curve (i.e., 4% A's, 20% B's, 52% C's, etc.).
2. Relative marking or rank in class marking: How well each student did with respect to others in class (i.e., third from the top out of 35 will receive A's or with percentiles and/or standard scores).
3. Absolute standard (standard set up by the teacher to judge the knowledge of students without recourse to any other standards).
4. Compromise procedure (jointly compounded of rank in class or absolute standards).
5. Other.

28. Which of the following types of formula do you use when weighting the factors (homework, tests, projects, etc.) which you consider in determining a pupil's grade at the end of each grading period?

1. Formulas which weigh daily work. (Homework or classwork) test (both quizzes and major examinations) and other factors (i.e., special projects) approximately equal.
 2. Formulas which give most weight to daily work.
 3. Formulas which give most weight to test scores.
 4. Formulas which give most weight to other factors.
 5. No weighted formula used.
29. Which of the following statements best represents your general attitude concerning school grades?
- | | | |
|-------------|----------------|----------------|
| Unconcerned | Should abolish | |
| Some value | Great value | Necessary evil |
30. In your opinion which of the following student attitudes toward grades is most conducive to deriving the maximum benefits from your subject?
- | | |
|--------------|----------------------------|
| Primary goal | One of several major goals |
| Minor goal | No goal |
31. Which of the following statements best describes the importance you attach to evaluation as a part of the job of the teacher?
- | | |
|---------------------|--------------------------------------|
| None | Some importance |
| Moderate importance | Major importance Minor importance |

Dear Teacher:

I am conducting, as my dissertation study, an investigation of the marking and grading practices in the public secondary schools of Oklahoma City. The study will require certain data that only you, as a classroom teacher, can furnish. Therefore, I solicit your cooperation in completing the enclosed questionnaire and returning it to me at your earliest convenience. Although your questionnaire is identified, you may have my assurance that the information will be treated anonymously and confidentially.

I am cognizant of the fact that this request for your assistance is really an imposition. My only defense, however, is that I believe that this study will contribute, at least in a minor degree, to the research needed to help provide a consistent and equitable approach to the evaluation of student achievement.

Sincerely,

Leon V. Crowley
Doctoral Candidate
The University of Oklahoma

Enclosures:

1. Return Envelope
2. Questionnaire