A STUDY OF SOME OF THE DETERMINANTS
OF TEACHER MARKS

A Dissertation
Presented to
the Faculty of the College of Education
University of Houston

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Education

---

by
Glenn Lewis Morris
August 1965

ACKNOWLEDGMENTS

I wish to express my deepest gratitude to the dissertation committee for their patience and understanding and for their contributions of their time in this study. The members of the committee--Dr. Joseph L. Fearing, Dr. J. M. Muse, Mr. Stanley Sanders, and Dr. Marvin Sterrett--provided valuable counsel and assistance. Especially do I wish to thank Dr. Gerald T. Kowitz, Chairman of the Committee, for the assistance he provided toward developing the methodology and presentation of the data. It was a privilege to work under his guidance, and I shall never be able to properly express my gratitude to him for his selfless assistance.

Sincere appreciation is also expressed to the teachers in the schools throughout the United States who participated in the study by grading the sample papers.

Appreciation also goes to my wife and daughter who not only encouraged me throughout the study but actively participated in the typing of the manuscript.

Finally, my special thanks go to Mr. J. C. Rogers, Jr., Superintendent of Lamar Consolidated Independent School District, for his understanding of this project. Without his cooperation and assistance this undertaking would not have been possible.

A STUDY OF SOME OF THE DETERMINANTS OF TEACHER MARKS

An Abstract of a Dissertation

Presented to

the Faculty of the College of Education

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Education

by

Glenn Lewis Morris

August 1965

A B S T R A C T

Morris, Glenn Lewis. A Study of Some of the Determinants of Teacher Marks. Unpublished Doctoral Dissertation, University of Houston, Houston, Texas, August, 1965.

Statement of the problem. Most of the writings about marks in recent years have been concerned with the recording or reporting of pupil progress. Little attention has been given to how teachers decide upon marks or what the basic determinants of marks are. One of the purposes of this study was to find if the information a teacher has as to the student's age, grade, sex, ability, past achievement, or conduct, affects the grades assigned to papers. These are the factors mentioned most often in the literature as factors teachers consider in assigning marks. It was also the intent of this study to determine if the nature of the assignment of a paper, i.e., whether a paper is designated a test paper or a daily paper affects the mark assigned by the teacher. Another objective was to determine if the above-mentioned information caused a reduction of variability in the scoring of papers. A further purpose was to determine whether the variability in scoring was greater in scoring English papers or in the scoring of arithmetic papers. To determine if teachers used a fixed point system or floating point system of grading was also a goal of the study.

Fixed point and floating point systems of grading. The fixed point system of grading presumes a single standard for all; the focus is upon meeting the terms of the prescribed standard. The floating point system usually sets no definite standards that must be met by every child. In such a system the grades may be modified by such factors as information about the student or about the assignment. In this study, it was assumed in the rationale that teachers would use a floating point system. It was also assumed that teachers would grade papers in accordance with the expectations which they had for students. For example, if a student's past achievement had been consistently low, the teacher would not expect so much from that student and would tend to give an "average" paper a higher score. If the student's past achievement had been consistently above average, the teacher would expect more from that student than average work. If an average paper were presented, the paper would receive a lower score.

Procedures and sources of data. Two hundred twenty-five female, fifth grade teachers from all of the geographical areas of the United States scored the same arithmetic and English papers under different conditions of information. The teachers had been selected by their supervisors on the basis of rating and years of service. All of them were given instructions to score the papers on the basis of 0-100 points with 70 considered passing. One-third of the teachers were given no information about the student or the assignment. One-third of the teachers were given information that according to the rationale was designed to depress the scores. One-third were given information that was designed to elevate the scores if teachers graded on the basis of expectations for the students. A random half of those teachers who were given information about the student were also given information that the paper was a test paper. An equal number were given information that the paper was a daily paper.

The resulting scores were statistically treated through a multiple classification analysis of variance which provides a test of differences among the means for several groups simultaneously. A ratio of variances was used to test for significant reduction in observed variance.

Conclusions. Some of the findings of the study were the following:

1. In the scoring of arithmetic papers, the fixed point method of grading was used. However, variance was reduced when information about the student was given to the teachers, and when given the information that the assignment was a daily paper.

2. In the scoring of English papers, information that was designed to depress the grades actually elevated the grades. Teachers graded on the basis of the floating point, but the point floated with the reputation of the student rather than expectation for the student. If the student's

past achievement was good, the grade was elevated. Variance was reduced when teachers were given information that was designed to depress the score. Subsequent analysis showed that information about the conduct of the student reduced variance.

3. Variance was less in the no-information category in English than in the same category in arithmetic.

4. Although teachers grade on the basis of factors other than achievement, when scoring English papers, this study did not identify these factors.

Recommendations. Some of the recommendations derived from the study are as follows:

1. Further study should be organized in attempts to analyze the evidence used in arriving at grades.

2. The study should be replicated with a design which would generate a more heterogeneous sample.

3. The study should be replicated with a greater amount of information supplied to the teacher about the student such as the information a teacher would have available from school records.

4. Further attempts should be originated to identify some determinants of marks other than achievement.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

THE PROBLEM AND DEFINITIONS OF TERMS USED

Introduction. Probably the most frustrating aspect of a teacher's job is giving grades. The activity of marking occupies much of the conscientious teacher's out-of-class time. Marks may serve a variety of purposes. One of the administrative purposes is the reporting of pupil progress. It should be pointed out that the progress record of the child, which includes teacher marks, is a legal record and remains in the school in perpetuity. Through marks the teacher intends to convey to the school, the parents, and the child the student's success or failure. If the child has not learned much, this may be an unpleasant experience for both the parents and teachers, as well as for the child. In most cases, the message which the teacher is trying to transmit to the parents and the child is garbled because the mode of the scale is not clear. Evidence of the problem of communicating with parents at the end of the reporting period is the numerous telephone calls asking for clarification of assigned marks. The mark is usually in the form of one number or letter that is supposed to convey to the parents, the school, and the child what the child has achieved.[1] In some instances the grade may mean that, compared to the work of the rest of the class, the child has failed or succeeded. In other cases, the mark may mean that, in relation to what the student has the ability to do, he did not meet expectations or he exceeded expectations. Little is known about the underlying rationale of the marking procedure, and, consequently, most people have a feeling of dissatisfaction with the whole process of marking.[2]

I. THE PROBLEM

Statement of the problem. Although much has been written about marking, most of it has been concerned with the recording of pupil progress or reporting progress to students, which are only two of the aspects of marking. Little attention has been given to the basic determinants of marks or how the teacher decides upon them. In some cases, the final grade in a course is the accumulation or the average of many marks that the student has made on individual papers in his course. However, the averaging of grades does not reveal the basis on which the marks were assigned, nor does the process itself provide a basis.

---

[1] C. C. Ross, Measurement in Today's Schools (New York: Prentice Hall, Incorporated, 1947), p. 409.

[2] Frederick B. Davis, Educational Measurements and Their Interpretation (Belmont: Wadsworth Publishing Co., 1964), p. 287.

If the meaning of the original scores is not known, the averaging of these scores does not add meaning, nor does it create meaning. For example, whether a paper was difficult to complete or easy to write is not indicated by the score on the paper.

Carter concluded, after examining three separate studies, that marks are determined by factors other than achievement. He recommended that a study be made on the significance of non-intellectual factors involved in the assignment of marks. Some of the non-intellectual factors which he listed were age, grade, sex, ability, and the behavior of students.[3] The same factors referred to by Carter as determinants of teacher marks were considered important determinants by other writers such as Norsted[4] and Farwell.[5] Neither of the writers gives evidence to substantiate his claims concerning the determinants of marks. If their observations were correct, it should be possible to identify some of the factors and bring some understanding out of the confusion that currently exists.

---

[3]R. S. Carter, "How Invalid Are Marks Assigned by Teachers?" Journal of Educational Psychology, XLIII (May, 1952), pp. 218-228.

[4]R. A. Norsted, "To Mark or Not to Mark?" Journal of Education, CXXI (February, 1938), pp. 81-84.

[5]G. H. Farwell, "Pressures Behind the Grade," Clearinghouse, XXXIX (April, 1964), pp. 462-466.

That confusion exists is indicated by Crooks in his report of a survey of forty-three teachers and principals. He found that forty-nine different bases were used in marking.[6] Regardless of the base or purpose of marks, the explanation of a grade assigned by a teacher is extremely difficult for the teacher who does the marking and may be interpreted quite differently by the person who receives the mark.

Another factor that is a possible determinant of marks is the type of paper that is being scored. A test paper might be scored on a different basis from daily work, and the determinants of teacher marks may vary among subject areas.

The purpose of this study is to determine whether or not the non-intellectual factors such as age, grade, sex, teacher's estimate of ability, past achievement, and behavior of students are significant factors in the assignment of marks. If information is significant in the determination of marks, the study will identify some of the individual informational factors that help to determine marks given by teachers. For example, if the chronological age of the student is a determinant of assigned marks, the scores will reflect the finding. If the determinants of marks can be identified, the recognition of the determinants

---

[6]A. D. Crooks, "Marks and Marking Systems," Journal of Educational Research, XXVII (March, 1933), pp. 259-272.

would help to reduce observed variance. The reduction of variance would mean a reduction in the spread of the scores given on the same paper. This would add to the reliability of marks, since the reduction in variance is essentially a reduction in error. In other words, less error shows greater reliability; more error is evidence of greater unreliability.[7] Reliability in marks means that marks are more stable and consistent. It also means that the marks are more useful for communication purposes and for the records of the institution.[8]

Teachers' marks, as well as all sets of measurement, have a total obtained variance. The total observed variance includes systematic variance that causes the scores to vary in one direction more than another. For example, the information a teacher has about a student may systematically influence the mark assigned by the teacher to the child's paper. Some of this variance is due to random or error variance; scores tend to fluctuate in an unstable manner. They are the combination of a number of causes: the chance elements present in all measures due to unknown causes, temporary fatigue, fortuitous conditions that may affect the

---

[7]Fred N. Kerlinger, Foundations of Behavioral Research (New York: Holt, Rinehart, and Winston, Inc., 1964), pp. 430-435.

[8]J. Stanley Ahmann and Marvin D. Glock, Evaluating Pupil Growth (Boston: Allyn and Bacon, Inc., 1963), p. 565.

object measured, fluctuations of memory, and other agents that are temporary and shifting. Any obtained variance is said to be composed of two components, i.e., the true component and the error component. All systematic variance is included in the true component. In measurement, if the proportion of true variance to the total variance is high, grades are reliable. If the proportion of error variance to total observed variance is high, reliability is low.[9]

In this study an attempt was made to determine if the error variance was reduced by supplying teachers with information about the student whose paper was being graded.

Another goal was to determine if the structure of the subject affects marking of papers in that subject. Since English and arithmetic have markedly different structures, these two subject areas were chosen. In the grading of arithmetic there are relatively few permissible alternatives. Some teachers grade arithmetic papers on the basis of the answer as either correct or incorrect, and the pupil receives full credit or no credit. The other alternative is that the paper is graded on the basis of evidence of appropriate processes being applied to obtain the answer. Thus, if the pupil has selected the method which has the capacity to solve the problem but had made a

[9]Kerlinger, op. cit.

computational error, he may be given partial credit. The grading of a composition in English, in contrast, presents many permissible alternatives. In fact, there are so many that some teachers find it necessary to give as many as three grades for one paper. They may score such things as writing, content, and creativity. Within the areas which are given a mark are a number of other alternatives such as punctuation, spelling, neatness, capitalization, sentence structure, originality, and many others. The exploration of whether variance in marking procedures is a function of this aspect of subject matter structure is, therefore, another goal of the present study.

Finally, the study was designed to indicate whether variance in marking procedures was a function of the nature of the assignment, i.e., did the designation of a paper as "test" or "daily" affect the grade assigned? It was reasoned that daily papers are many times assigned as homework. The student has opportunities to receive help from parents and from other students with his homework. Therefore, the daily paper may have a different value from the test paper on which the pupil can get no help, since it was completed under the supervision of the teacher.

The effects of non-intellectual factors of information about the student and the influence of the nature of the assignment being scored were examined in terms of how they affected the mean scores assigned to papers by teachers. How the informational factors affected variability was determined by observing the reduction or increase in variance that occurred when the non-intellectual factors of information were given to teachers.

Importance of the study. The recurring tasks of evaluating the progress of students and the assignment of marks are related to many other facets of formal education. Marks have far-reaching effects because they are used as a basis for informing pupils of progress, reporting to parents, admitting students to high schools and colleges, promoting competition, inciting study, determining honors, rewarding of scholarships and grants, grouping, and predicting pupil success.

It would seem that, since marks are so important in the life of the student, they should be standardized so that a mark given by one teacher would carry a similar informational content to one given by another teacher in the same school. Unfortunately, this is not true, since teachers do not even agree on the meaning of marks.[10] Mandel had this to say about the importance of marks: "In these days, when so much is based on scholastic averages,

---

[10]Robert M. Travers and Norman E. Gronlund, "Meaning of Marks," Journal of Higher Education, XXI (May, 1950), p. 370.

the mark should be a very accurate evaluation of the student's ability and achievement."[11]

Inequities occur because of the confusion in the marking processes. Wood stated that the College Entrance Examination Board examinations deny entrance of 1,279 out of 10,000 applicants for college who, if given a different form of the same test, would pass. He also indicated that a similar number who passed the first test would fail a different form, if it were administered.[12] This represents an error of approximately 13 per cent, and it points to the problem that even carefully standardized tests contain a margin of error. Standardized tests with their margin of error should be more reliable than marks issued on the basis of individual judgments of teachers.

Errors similar to those recorded by Wood, occur daily in public schools. In grouping for instruction, about 10 per cent of the students who belong in the A section, i.e., those with highest ability, are placed in the B section, or a section of lower ability, and vice versa. If homogeneous grouping is used, the separation of the brighter child from the duller child should be done as accurately as possible,

---

[11] B. Mandel, "Appraising Ability and Achievement," High Points, XLIV (April, 1962), pp. 31-37.

[12] B. D. Wood, "Measurement of College Work," Educational Administration and Supervision, VII (September, 1921), pp. 301-304.

since their self concepts are affected by the group in which they are placed.[13] The expectations of the teacher are different, also. According to Symonds, it becomes increasingly important that individuals be treated "fairly" in the critical points of their career and that vocational counsel be based on the best evidence obtainable and not on opinion.[14] If pupils base their vocational aspirations on the subjective opinions of others, they will more than likely vacillate from one goal to another because of the variety of opinions they will encounter.

The greatest inequity resulting from imperfect measurement lies closer to the learning process. A pupil's present learning activities are affected by the evidence of his past learning, i.e., his grades. Several writers have shown experimentally that praise is a valuable incentive for school work. Hurlock states as the conclusion of her experiment that praise is the most valuable of incentives in school work regardless of age, sex, initial ability, or accuracy.[15] Mandel, in agreeing with Hurlock, said that nothing is gained by giving a low mark; on the contrary,

---

[13] P. M. Symonds, Measurements in Secondary Education (New York: The Macmillan Company, 1934), p. 19.

[14] Ibid.

[15] E. B. Hurlock, "The Value of Praise and Reproof as Incentives for Children," Archives of Pyschology, LXXI (July, 1924), p. 462.

such marks will usually discourage future efforts by the student.[16]

School marks constitute an important kind of praise or recrimination.[17] A pupil who receives a mark of "A" or "B" may consider himself a good student; whereas, one who receives a below-passing mark may consider himself a failure, or conclude that the teacher is unfair. A grade can mean different things to different people. A student who considers himself an "A" student may view a "B" as a failure; while a child who considers himself a "D" student may view a "B" as an astounding success. It is important that the student be given an accurate evaluation of his achievement so that he may be encouraged to set realistic goals. It would be unwise for a person who was deficient in communication skills to aspire to the legal profession. It would be equally unwise for a person who could not achieve in quantitative skills to pursue a career in engineering.

Realizing the problems that exist in the area of grading, school districts have expended much time and effort in attempts to develop new marking systems. The different marking systems that have been developed have

---

[16]Mandel, op. cit.

[17]Symonds, op. cit., p. 20.

not provided a permanent, workable solution; none of them has been more than one effort in a long series of abortive attempts to satisfy the various expectations of parents and teachers.[18]

Further evidence of the attention given to the problem was noted in an examination of the *Educational Index* for the period of July, 1961, to June, 1963. Considering only those journals which were included in the *Index*, one finds that during this period there were seventy-three entries dealing with marks, marking systems, and report cards, for an average of 3.04 articles per month or almost one each week! Many of these articles were not pertinent to this study, since they did not deal with studies in reliability; nor did they show research in finding determinants of teacher marks. Yauch stated that the emphasis in research in grading has shifted away from earlier attempts to establish statistical reliability and writers have chosen instead to be satisfied with surveys of current practice. A record of what schools are currently doing in no way validates the practice.[19] Hence, it is difficult to see how this kind of research contributed to deeper understandings of the problems involved.

---

[18]J. Stanley Ahmann, Marvin D. Glock, and Helen Wardeberg, *Evaluating Elementary School Pupils* (Boston: Allyn and Bacon, 1960), p. 355.

[19]W. A. Yauch, "School Marks and Their Reporting," *National Education Association Journal*, L (May, 1961), p. 50.

II. KINDS OF MARKING SYSTEMS

A variety of marking methods has been devised over a period of years; some of them have had extensive use in the schools. The most common marking systems are the percentage system, the symbolic system, the dichotomous system, and a ratio system which shows a relationship between a child's ability and his achievement.[20]

Percentage system. The percentage system has been the target of much criticism. In the percentage system marks are given from zero to one hundred with a mark of seventy generally considered passing. The objection is raised at this point that the three reference points --zero, seventy, and one hundred--are not stable; neither are they operationally defined. A mark of zero to one person who is scoring an arithmetic paper may mean that the problems were worked by the correct process, but the student did not get the right answers. To another teacher the mark of zero may mean that the student did not turn in a paper. If he had turned in a paper, the teacher would have given him fifty or thirty-five or some other figure for "minimum effort" in meeting the requirement.

---

[20] A. E. Traxler, Techniques of Guidance (New York: Harper and Row, Publishers, 1957), pp. 235-242.

The same problem occurs in defining one hundred. A score of one hundred is generally considered to be 100 per cent. The problem then becomes one of defining what 100 per cent means. Some teachers may consider it 100 per cent of all knowledge of a particular subject. Since testing cannot test all of the knowledge about a subject, it would be difficult for a pupil to score 100 per cent of such a standard. Another teacher may expect the student to give a complete return of all knowledge the instructor has presented. Still another teacher may consider 100 per cent to be the amount of information the "best" student has acquired. To one teacher who grades an arithmetic paper, the process is of the essence; a student who uses correct processes is given 100, even though he erred in computation.

Seventy is even more difficult to define, since it is supposedly not on either end of the scale. No one has adequately explained why a mark of seventy is considered passing. Many questions arise as to how the figure was ascertained. The answer may be that someone decided that a child should know 70 per cent of what was taught him and that, subsequently, the figure became accepted through use. In our current technological world, a knowledge of somewhat less than three-fourths of the essentials does not seem to be a reasonable standard. Another criticism is that the percentage method provides for finer distinctions than human judgment is able to make. Symonds reasoned that, for rating human traits, seven is the optimal number. More than seven steps yields an increase in reliability that is slight. In fact, too refined a scale makes it difficult for an observer to discriminate between steps. Less than seven steps leads to appreciable loss in reliability.[21]

Symbolic system. The symbolic system is one in which letters or symbols are used to indicate the achievement of pupils. It is usually a five-point system, although it may include only three, or as many as nine, symbols. The main argument for this system is that it is easier to classify students into five general groups with symbols such as "A," "B," "C," "D," and "E" than it is to classify them into 101 groups as is done in the percentage system. It must be pointed out that, in many cases, the percentage system is used in scoring papers and later, the results are translated into a symbolic grade often through rather exotic mathematical manipulations. The manipulations increase neither the validity nor the reliability of the marks.

A symbolic system may be related to the normal curve and is so related by some teachers. A common plan is that of assigning a mark of "C" to all students falling

---

[21]Symonds, op. cit., p. 511.

between plus 0.5 standard deviation and minus 0.5 standard deviation; "B" to all those between plus 0.5 standard deviation and plus 1.5 standard deviation; "D" to those between minus 0.5 standard deviation and minus 1.5 standard deviation; "A" to those above plus 1.5 standard deviation; and "F" to those below minus 1.5 standard deviation. In a normal distribution the following percentages of the different marks would result: "A," 7 per cent; "B," 24 per cent; "C," 38 per cent; "D," 24 per cent; and "F," 7 per cent.[22] By chance alone, a student's grade could vary as much as from an "A" to a "C," or from an "F" to a "B."

That there are no absolute fixed percentages observed by persons using the normal curve is indicated by the fact that ten or more different distributions, each providing for a five-point system, have been defended by educators of importance.[23]

If a basis exists for transmuting point scores into letter grades as shown above, the justification is when a large enough sample is used, the distribution of scores will follow the normal probability curve with 68 per cent of the scores falling between plus 1.0 standard deviation and minus 1.0 standard deviation. Between plus 2.0 standard deviation and

---

[22] Ibid.

[23] C. C. Ross, Measurement in Today's Schools (New York: Prentice Hall, Incorporated, 1942), p. 436.

minus 2.0 standard deviation will lie 95 per cent of the scores. Between plus 3.0 standard deviation and minus 3.0 standard deviation lie 98.7 per cent of the scores. With a class of under forty pupils it is only by chance that a pupil could vary 2.5 standard deviation from the mean of the group in either direction. An "A" or an "F" could occur by chance alone in such a system.

All such schemes are based on the central tendency and the variability of the group, with the mean used as the measure of central gendency and the sigma or standard deviation used as a measure of variability.

Dichotomous system. In a dichotomous system the only marks that are assigned are pass or fail, or satisfactory or unsatisfactory. The former system is curriculum-centered, while the latter system is child-centered. The satisfactory or unsatisfactory marking is more in keeping with the spirit of the modern school.[24] The dichotomous system intends to do away with harmful competition, since there is no degree of passing or failing. The system has much to commend it, especially for reporting to parents.

---

[24] Ibid.

Like other marking systems, this system has not been wholly satisfactory, in that some of the purposes for which marks are issued are not realized. As a result, many schools have gradually abandoned it. The first step away from this dichotomy is generally the establishment of a middle group. An honors group may be added later; and soon the school is back to the five-point or seven-point system. The gradual breakdown of the dichotomous system is characterized by the giving of "S+" and "S-." Although it would seem that teachers would favor a system in which they only had to evaluate a student's work as satisfactory or unsatisfactory, many of them do not. At the initiation of the dichotomous system, they do approve of it; but soon they find that they no longer have marks to use as incentives for students. Two marks do not separate students into enough groups for most teachers, pupils, and parents. Also, they provide even less information to parents about the pupil's learning progress. No information is given as to the degree of proficiency attained in a subject area.

Ratio system. Rather than using group achievement as a basis for individual marks, some reports use the child's ability as a basis for marking. If a pupil is achieving up to capacity, his mark is high, even if his achievement is less than that of a more able pupil. The latter might well be given a low mark because he could have done much better.[25] The ratio system which is exemplified below is one example of a floating point system.

An example of a system that indicates relationship of pupils' achievement to ability is the accomplishment quotient system that was popularized by Franzen. The I.Q., or accomplishment quotient, equals the educational age divided by the mental age.[26] For example, if a pupil whose mental age is 10-0 has an educational age of 9-2, his accomplishment quotient is 110 divided by 120, or 92. A pupil might have an educational age of 9-2, but a mental age of only 8-3. His accomplishment quotient would be 111. The interpretation of the first case is that the pupil is not achieving to the limits of his capacity. The interpretation of the second case appears to imply that this pupil has exceeded his capabilities by 11 per cent. Another explanation may be that both of the measurements on which the A.Q. is based are composites of a number of measures. Both the mental age and the educational age are measures of varying reliability. To complicate matters further, they are used

---

[25]Ahmann, Glock, and Wardeburg, op. cit., p. 365.

[26]R. Franzen, "The Accomplishment Quotient," Teachers College Record, XXI (November, 1920), pp. 432-440.

conjunctively. It would be difficult to determine the meaning of the resulting index.[27]

Supplementary marking systems. To overcome the limited information provided by report cards and to improve cooperation between teachers and parents, some schools have used informal letters or parent-teacher conferences. The major advantage of both methods is the greater flexibility in reporting. Particular points that might cause misunderstanding on an abbreviated form can be described and explained in detail. The completeness of the report is limited only by the data available about the pupil and the ingenuity of the teacher in organizing it. The parent-teacher conference has an additional advantage of providing parents with an opportunity to ask questions and discuss plans for the pupil's further development.

The informal letter and the parent-teacher conference are seemingly limited to a supplementary role because of their shortcomings: (1) Comprehensive reports require an excessive amount of time to prepare. (2) Such informal reports do not provide a systematic report of pupil progress toward goals of the school. (3) Different aspects of the pupil's development are apt to be stressed from one report to another. Letters and conference reports are difficult

[27]Kerlinger, *op. cit.*, p. 618.

to summarize for administrative records. (4) It is seldom possible to obtain the cooperation of all parents.[28] Some of the criticisms may reflect the misuse of the system rather than an inherent weakness in the plan.

III. BASES FOR MARKS

Teachers and supervisors are often unsure about what should be the basis for marks. Some teachers have said that grades should represent as accurately as possible only the level of achievement of the pupils. Other teachers suggest that marks should represent the gain of the student; while still others suggest that the effort, potential, ability, and study habits of the student should be reflected in marks that are given.

The issue as to the basis of marks resolves itself into the question of whether marks should be based on some absolute standard of achievement or on some comparison of achievement with the student's ability or other factors. To state the issue mathematically, should marks be given by a fixed point method or a floating point method? A fixed point method presumes a single standard for all; the focus is upon meeting the terms of the prescribed standard. The floating point system considers other characteristics of the child than his achievement in assigning marks.

---

[28] Norman E. Gronlund, _Measurement and Evaluation in Teaching_ (New York: The MacMillan Company, 1965), pp. 377-378.

The fixed point system uses a clearly defined operational standard, which may include sub-answers and processes for the product of the student's work. The standardized achievement test is an example of this system. The test is scored objectively. The score given to the student is determined only by his performance on the test; personal characteristics, situational factors, or other items of information about him are not considered. Another example of the fixed point system of grading may be found in the scoring of arithmetic papers by a teacher. The teacher sets a standard that no errors in method or computation will be tolerated. For example, if a pupil works a problem and uses the correct method but makes an error in a partial product, by the fixed point standard the whole problem is wrong and is so graded. Similar standards may be operationally defined by teachers in other subject areas. Under such a system no attempt is made to distinguish between the bright but indolent pupils and the dull but industrious ones. The relative amount of effort put forth by the student is not a factor, since every student is graded by a single standard.

The floating point method usually sets no definite standards that must be met by every child. A child's work is marked according to an estimate of his ability derived either from a test score or an opinion of the teacher, or in relation to past achievement, age, or other factors. Even though the teacher judges that the child has not achieved the required grade level of work, the student may still receive a favorable mark because the factors that modified his achievement are given due consideration.

IV. LIMITATIONS OF THE STUDY

Several elements created limits on the scope and findings of the study. They were these: (1) limitations arising from the characteristics of the sample, (2) the attributes of instruments to be scored, (3) the structure of academic disciplines or subject content, and (4) the value placed on factors of information transmitted.

Limitation of the sample. The study was restricted to 225 female fifth grade teachers with from two to fifteen years experience, who were rated good or excellent by their supervisors. The teachers were from schools in metropolitan areas and from cities of 100,000 to 400,000 population. The cities represented all the major geographical areas of the United States. The results may or may not apply to teachers with less than two years experience or more than fifteen. The results may or may not apply to scoring done by men, since it was advisable to control the sex factor so as not to introduce another possible variable. Douglas had stated that marks are determined by factors other than achievement, especially marks assigned by women teachers.[29]

Limitation of the instruments to be scored. Both of the papers were done by one fifth grade student and were selected as average work for a particular school. The work may or may not be considered average by other schools, but it is of no consequence in this study. Any stimulus within the average range, even broadly conceived, would have served equally well. Since no remuneration could be offered the cooperating teachers, the work samples were kept short so that the demands on the teachers' time would be minimal.

Limitation of academic subjects. The study was limited to the academic subjects of arithmetic and English. Statistical inference concerning marking in other subject areas may or may not be valid.

Limitation of factors of information. The study included only information about the student as to age, grade, sex, past achievement, ability, and past behavior. These are the only non-intellectual factors of information that were tested to discover whether information is a factor in assigning marks. Other factors of information

---

[29]H. R. Douglas and N. E. Olson, "Relation of High School Marks to Sex in Four Minnesota Senior High Schools," School Review, XLV (October, 1937), pp. 282-288.

may or may not have been operating in the assignment of marks which were unknown.

V. DEFINITION OF TERMS

The following definitions are presented to clarify the use of terms in this study.

Mark. The teacher's numerical or letter evaluation of pupil achievement in a course or area of performance.[30]

Grade. The meaning is essentially the same as "mark" and is frequently used instead of it.

Score. The number of points representing the actual performance on a test or examination. Grade refers to the interpretation of the score.[31]

Achievement. The knowledge, understandings, and skills of students at a specific point in time in a specific subject area.[32]

---

[30]Harry A. Greene, *Measurement and Evaluation in the Elementary School* (New York: Longmans, Green, and Co., 1953), p. 591.

[31]L. W. Webb and A. M. Shotwell, *Testing in the Elementary School* (New York: Farrar and Rinehart, Incorporated, 1939), p. 11.

[32]Davis, *op. cit.*, p. 114.

Variance. The reflection of the range of difference among a group of scores. In statistical analysis, the variance is defined as the mean of the sum of squares of deviation of scores from the grand mean, i.e., $V = \frac{Ex^2}{n}$ where V is the variance; Ex^2 is the sum of the squared deviation scores; and n is the number of cases. The squared deviation scores are found by: $x^2 = (X-\overline{X})^2$ where X is any individual score and $\overline{X}$ is the mean of the distribution.[33]

Error variance. The variance left over in a set of measures after all known sources of systematic variance have been removed from the measures.[34]

Systematic variance. The variation in grades due to known or unknown influences that cause grades to vary in one direction more than another.[35]

Variability in marks. The extent to which scores spread above and below the average.[36]

---

[33] J. P. Guilford, *Fundamental Statistics in Psychology and Education* (New York: McGraw-Hill Book Company, Inc., 1956), p. 85.

[34] Kerlinger, *op. cit.*, p. 101.

[35] *Ibid.*

[36] Ross, *op. cit.*, p. 81.

Validity in marking. The extent to which the results of an evaluation procedure serve the particular uses for which they are intended.[37]

Reliability in marking. The extent to which the results of an evaluation procedure are consistent.[38]

Fixed point grading. The method of grading in which a clearly defined operational standard is utilized.[39]

Floating point grading. The method of grading in which a ratio of a clearly defined operational standard to other factors, such as information, is utilized.[40]

Structure of a subject. The organization of a subject which permits greater or lesser optionality in content, emphasis, and grading.

In the next chapter, a review of the literature related to reliability of marks and possible explanations of this reliability will be given. As will be shown, the

---

[37]Gronlund, *op. cit.*, p. 59.

[38]*Ibid.*, p. 60.

[39]Interview with Dr. Gerald T. Kowitz, Director of the Bureau of Education Research and Services, University of Houston, 1965.

[40]*Ibid.*

reliability of teachers' marks is disappointingly low; the variance observed in marks assigned by teachers is largely unexplained. Some of the systematic variance may be a function of the information teachers have about their students.

VI. SUMMARY

Teachers are commonly required to evaluate the progress of students by assigning number or letter grades in a percentage, symbolic, dichotomous, or a system that indicates relationship of a pupil's achievement to ability. These grades are based on the marks that student have made in daily work and on tests.

Marks are supposed to have far-reaching effects on the lives of students because they should inform them of their levels of attainment and may motivate them to learn more effectively.

The study attempted to do the following:

1. To determine if the information that a teacher has about a student affects the marks assigned by a teacher either by elevating or depressing the score.

2. To determine if information that a teacher has about a student reduces observed variance in assigned marks.

3. To determine whether teachers grade on a fixed point or floating point basis.

4. To determine if variance in marking is greater in the more loosely structured subject areas such as English than in the more rigidly structured subject areas such as arithmetic.

5. To determine if the nature of the assignment affects variance in marking.

CHAPTER II

RELATED LITERATURE

The concern with marks by educators can be traced to ancient Greece, Egypt, Babylon, and China. Current interest parallels the testing movement, which had its inception in the first part of the twentieth century, initiated by the studies of that period, which pointed to the variability and unreliability of marks.[1] These studies showed that something was awry but failed to demonstrate the locus of the problem.

I. VARIABILITY OF TEACHER MARKS

In 1912, Starch and Elliot investigated the scoring of examination papers. A geometry paper written by a student in one of the larger high schools of Wisconsin was reproduced and sent to 180 high schools in the North Central Association of Schools. Of the one hundred sixteen replies received, two scores were above 90, and one was below 30. Twenty were 80 or over, and twenty were below 60. Forty-seven teachers thought the paper was passing or above, but sixty-nine considered it to be failing.[2]

---

[1]Daniel E. Starch, Educational Measurements (New York: The Macmillan Company, 1916), pp. 20-39.

[2]Daniel E. Starch and E. C. Elliot, "Reliability of Grading Work in Mathematics," School Review, XXI (April, 1913), pp. 254-259.

Similar results were obtained in scoring history and English papers. The history paper was awarded marks of from 43 per cent to 90 per cent by seventy teachers.[3] The marking of the English paper by 142 teachers resulted in marks ranging from 50 per cent to 98 per cent.[4] The study in which the English paper was scored has direct implications in the current study. It should be noted that the sample had obvious limitations in the geographical area sampled, the number of responses, and the size of the high schools.

Ruch demonstrated by an experiment that the scoring of examination papers was extremely subjective. To ninety-one teachers he submitted a single question in geography with three answers actually written by pupils in a geography class. The answers were selected from the best papers, median papers, and the poorest papers in the class. Teachers were asked to score the papers on a scale of 0-20. While one teacher assigned a value of 2 to one of the papers, another teacher assigned it a value of 20.[5]

---

[3]Daniel E. Starch and E. C. Elliot, "Reliability of Grading Work in History," School Review, XXI (December, 1913), pp. 676-681.

[4]Daniel E. Starch and E. C. Elliot, "Reliability of Grading Work in English," School Review, XX (September, 1912), pp. 442-457.

[5]G. M. Ruch, Improvement of the Written Examination (Chicago: Scott, Foresman Company, 1924), pp. 54-57.

The case of a pupil who had been presenting the writing of de Maupassant as his own was reported by Morlan. The instructor rated the work as good, but not exceptional. He felt that one of the themes was similar to something he had read before by de Maupassant. A comparison of the French writer's story with the theme of the young man revealed that they were the same. He checked other stories the student had written and found that they were also the stories of de Maupassant, which he had scored as "B." Many people would grade the writings of the French author higher than a "B."[6] Some teachers might possibly grade the author lower.

Another humorous incident was reported by Wood, which illustrates the subjectivity of marks:

> The facts of a subjective scale are well illustrated in the following anecdote concerning the grading of history papers by a group of college professors of history in the summer of 1920. One of the five or six expert readers assigned to a certain group of history papers, after scoring a few, wrote out for his own convenience what he considered a model paper for the given set of ten questions. By some mischance this paper fell into the hands of another reader who graded it in a perfectly bona fide fashion. The mark he assigned was below passing, and in accordance with the custom, this model was rated by a number of other expert readers in order to insure that it was properly marked. The marks assigned to it by these readers varied from 40 to 90.[7]

---

[6]G. Morlan, "He Gave de Maupassant a Grade of B," School and Society, XXXIV (September, 1931), pp. 321-322.

[7]Wood, op. cit., p. 301.

Kelly compared the final grades of a class in two successive semesters taught by different teachers. Concerning the experiment, he reported that there were four ward schools that sent their pupils to the same seventh grade. An attempt was made to determine how far the pupils maintained their relative positions in the common seventh grade classes. It was theorized that, if differences were found in the amount of increase or decrease in marks among the four school groups, it would be possible to measure the difference in the standard between a mark in one sixth grade and the same mark in a sixth grade in another school. Marks recorded on the cards were "E," "G," "F," "P," which were symbols for excellent, good, fair, and poor.

Kelly reported such differences as these:

> For work which the teacher in School C would give a mark of "G" in language, penmanship, or history, the teacher in School D would give less than a mark of "F." A pupil whose card in School C had been a "G" card, on the whole, would be dismayed in receiving an "F" when he moved to School D.[8]

Rugg said that individual teachers set their own standards for evaluation of students resulting in variability, unreliability, and inconsistency of distribution. Teachers intend to use actual achievement as the basis of

---

[8]F. J. Kelly, Teacher Marks, Their Variability and Standardization (New York: Teachers College, 1914), p. 142.

marks, but other factors enter into their evaluation.[9] It should be pointed out that other factors mentioned here were not identified. It was merely stated that other factors enter into teacher evaluation.

In a more recent report, Muessig declared that teachers have varying skills and abilities in measurement practices. He reported that one teacher may grade by homework; another may merely check in the work. One teacher permits students to do extra work to build grades; another does not. One teacher bases grades on true-false tests, another on essay tests. One teacher fails a certain percentage each year, while another fails none. One teacher hates to give "A's," yet another gives an equal number of "A's," "B's," and "C's."[10] All of these practices could contribute to variability in progress reports of students. However, they do not arrive at the specifics of the problem of marking papers.

As a result of interviews with teachers and administrators, Vredevoe stated that most teachers differ in their interpretation of achievement and that the value

---

[9] H. O. Rugg, "Teachers' Marks and Marking Systems," Educational Administration and Supervision, I (March, 1915), pp. 117-142.

[10] R. H. Muessig, "Mystery of Grading and Reporting," Education, LXXXIII (October, 1962), pp. 90-93.

of marks differ from school to school and among teachers in the same school.[11]

As recently as 1962, an arithmetic test paper was given to thirty-five mathematics teachers representing various geographic areas of the United States. These thirty-five teachers were members of a National Science Foundation Institute and were selected on the basis of training and demonstrated competence. They were asked to imagine themselves to be the teacher of the child who submitted the paper and to grade the test. The range of numerical grades of the group reported in the study was 10 to 75, with a mean of 46.85 and a standard deviation of 17.4. The members of the group were surprised at the divergence of scores.[12]

In a recent study at Michigan State University, it was found that students who received all "A's" in their freshman year tended not to receive all "A's" in subsequent years, which is indicative of the fluctuating nature of

---

[11]L. E. Vredevoe, "How May We Make the Recording and Reporting of Pupil Progress More Meaningful?" National Association of Secondary School Principal's Bulletin, XXXVII (February, 1953), pp. 179-182.

[12]Reuben R. Rusch, John A. Brown, and Arthur R. Delong, "Meaning of an Arithmetic Score," Arithmetic Teacher (March, 1962), p. 145.

grades.[13] Although it was noted here that grades do fluctuate, no attempt was made to account for this fluctuation. No one really knows why these students do not continue as "A students," nor is it known why they were "A students" their freshman year, since "a grade is a mystery to students, parents, and all others, perhaps even the teacher."[14]

Marks are not easily defined. Travers surveyed fifty college faculty members to determine if they interpreted grades in a similar manner. They did not agree as to the meaning of the final mark in a course. When asked to describe the meaning of the letter grade "C," they differed in their description in a range of six categories. Some described "C" as a failure, while others interpreted it to be a fair grade.[15]

Teachers differ widely in the factors to which they assign weight in determining the mark. One teacher said that "A" was such an outstanding grade that it was impossible to make one in his class. However, another teacher said

---

[13]N. M. Chansky, "X-Ray of the School Mark," Educational Forum, XXVI (March, 1962), pp. 347-352.

[14]Muessig, op. cit., pp. 90-93.

[15]Robert M. Travers and Norman E. Gronlund, op. cit., pp. 369-374.

that, if a teacher were really teaching, he should give 50 per cent "A's."[16]

There is reason to believe that a teacher who assigns marks if given the same evidence at a different time under different circumstances would assign different marks.[17] Although Jones alleges that there is reason to believe this, no empirical evidence is presented to support his allegation.

The bulk of the evidence presented in the literature emphasizes the notion that school marks, as assigned by teachers, are unreliable and variable. A few writers are in disagreement. Bloom found that high school teachers make relatively precise judgments about students and that these assessments are highly related to evaluations made by college teachers and to the results of achievement measures. However, he also stated that the differences in meaning of grades that reflect differences in programs and objectives can be made clear only by careful analysis

---

[16]Arch O. Heck, Thirty-seventh Yearbook of National Society for the Study of Education in Public Schools, Part II (Bloomington: Public Schools Publishing Company, 1938), pp. 187-189.

[17]John A. Jones, "Grading, Marking, and Reporting in the Modern Elementary School," Educational Forum, XIX (November, 1954), p. 48.

of the evidence used in arriving at grades.[18] This he did not attempt to do. Travers said that marks are not unreliable in themselves, but different teachers are appraising different outcomes and expectations. For example, in arithmetic, one teacher may score a paper for correctness of answer, while another teacher may be concerned only with the suitability of the plan of attack or academic strategy.[19] Teachers do have different expectations, but no one has determined what these expectations are and on what they are based. A part of this study is to ascertain if the expectations of teachers affect marks. Sharp conducted a study by reproducing fifth grade arithmetic papers and asking a group of teachers to score them. The range of scores was from 48-88. He then set very precise standards by which these same papers were to be graded, and he asked the same teachers to mark the paper again. The second time the range was given at 58-77. Although the range presented a smaller deviation, the means were not compared. It was evident that, although standards for grading were detailed, only about 80 per cent

---

[18]Benjamin S. Bloom and Frank R. Peters, _The Use of Academic Prediction Scales for Counseling and Selecting College Entrants_ (New York: The Free Press of Glencoe, Inc., 1961), p. 4.

[19]Travers, _op. cit._, pp. 369-374.

of the teachers followed instructions.[20] The variability as measured by the range was reduced by twenty-one points. Guilford said that, although the total range is the indicator of variability that is easiest and most quickly ascertained, it is also the most unreliable. It is almost entirely limited to the purpose of preliminary inspection.[21] The standard deviation is the most commonly used indicator of degree of variability and is usually the more reliable, hence would have been more meaningful. Modern statistical practices rely even more heavily upon the variance, or the square of the standard deviation.

II. DETERMINANTS OF MARKS

A survey of the literature on marking reveals that marks are supposedly determined both by achievement and by other factors. Exactly what the other factors are and how marks are affected by them has not been determined in previous studies. According to Liggett, no one knows the meaning of a mark, with the exception, perhaps, of the person

---

[20] L. A. Sharp, "The Value of Standards in Grading Examination Papers," Peabody Journal of Education III (January, 1925), pp. 38-45.

[21] Guilford, *op. cit.*, p. 78.

who assigned it.[22] He may not know it either. However, even if he does, it is quite likely that in a brief period of time even he will forget the basis of his judgment.

Ross said: "It seems too bad that the marks received by certain individuals are conditioned more by the contours of the face than by the contents of the head."[23] The inference is that teachers may sometimes base their marks on the facial features of a student rather than on the knowledge that students may have. Since Ross offers no experiment to justify his statement, it is assumed that the quotation is based on opinion.

Norsted said that a large fraction of teachers consciously consider effort, attitude, and other factors in assigning marks in school subjects; and many other teachers are affected by these factors without realizing it.[24] This is probably a supposition on his part, since no evidence is presented to justify his statement.

A study by Hadley found that the tendency is for most-liked students to be marked higher than their

---

[22]W. A. Liggett, "Are There Better Ways of Evaluating, Recording, and Reporting Pupil Progress in the Junior and Senior High Schools?" National Association of Secondary School Principals Bulletin, XXXIV (March, 1950), pp. 79-89.

[23]C. C. Ross, op. cit., p. 404.

[24]R. A. Norsted, "To Mark or Not to Mark?" Journal of Education, CXXI (February, 1938), pp. 81-84.

accomplishment would justify. He noted a tendency also to assign higher marks to girls than to boys, although on standardized achievement tests boys' and girls' performances were not significantly different.[25] The study by Hadley showed that most-liked students were given higher marks and least-liked students were given lower grades; but it did not account for the students who were neither least-liked nor most-liked, who were the majority of the students. The study was not designed to test whether higher marks were given to girls or boys, so he could only say that it appeared that such a tendency existed. In the absence of an experimental design with its attendant treatments, little validity can be attributed to the findings.

Volberding noted that successful eleven-year-old pupils were more often girls than boys, were from middle-class homes, were more intelligent, and were better adjusted socially and personally. No mention was made of the sex ratio as recorded in the school census. This could have accounted for the findings that were presented. The factors that seemed to have the most influence on teacher marks were teachers' estimate of the pupil's intelligence, pupil's attitude toward school, pupil's energy, and chronological

---

[25] S. T. Hadley, "School Marks--Fact or Fancy?" Educational Administration and Supervision, XL (May, 1954), pp. 305-312.

age.[26] Her study was an assessment of the characteristics of successful students by a survey technique. The study did not present data to substantiate factors which were determinants of teacher marks.

It was the opinion of Farwell that girls get higher marks than boys, but he did not attempt to explain why this was true. He further stated that, because of the influence of the I.Q. on some teachers, grades might just as well be assigned at the beginning of a course rather than at the end. His belief is that the sex of the student, as well as the I.Q. score, is important to the teacher in assigning marks. He offers no evidence to substantiate his beliefs.[27]

Findings by Herron indicated that, while teachers' marks represented achievement by the pupil, they gave evidence of the effects of personality, intelligence, the socio-economic status, and the age of the student upon the teacher. The sex of the student was more a determinant of the marks assigned by teachers than the sex of the teachers. Girls were said to be given better marks than boys. An

---

[26]Eleanor Volberding, "Characteristics of Successful and Unsuccessful Eleven Year Old Pupils," Elementary School Journal, XLIX (February, 1949), p. 410.

[27]G. H. Farwell, "Pressures Behind the Grade," Clearinghouse, XXXIX (April, 1964), pp. 462-466.

observation was made that boys' marks may suffer as a result of some anti-social behavior on their part.[28]

Carter also stated in a study of comparison of algebra achievement and marks assigned that, although no significant difference existed among various groups in intelligence and algebra achievement, significant differences were found in marks assigned by teachers of beginning algebra. Teacher marks reflected intelligence and achievement, but some other factors entered into the assignment of marks by teachers.[29] This study was somewhat limited in that only six teachers constituted the sample. These six teachers may or may not be a random sample of the total population of algebra teachers. He also recommended that the question of whether non-intellectual factors are of significance when teachers assign marks be answered in future studies.

Personality factors of students enter into marks assigned, according to Hadley. However, he mentioned that the extent to which factors other than achievement indirectly affect marks assigned is not known. In his

---

[28]John S. Herron, "How Teachers Rate Their Pupils," Department of Elementary School Principals Bulletin, VIII (October, 1929), pp. 235-239.

[29]R. S. Carter, *op. cit.*, pp. 218-228.

summary he indicated that the complex of components other than actual quality of performance which enters into marks such as sex, appearance, conduct, and attendance, has not been studied through any precise procedures.[30]

A conclusion was reached by Crooks that teachers intended to base marks primarily on achievement; but other factors entered into the grades that were given, indicating again that teachers' intentions in marking were not followed.[31] Johnson was definite in his stand that the only justifiable base for marks was absolute achievement.[32] This belief was also shared by Rugg.[33] As has been previously mentioned, these are subjective opinions as to what factors determine marks. Through observations and reasoning, many writers have arrived at beliefs concerning this question. This is indicated by the scarcity of studies which used empirical evidence to support the contentions of the authors.

Bills pointed out that agreement or disagreement between what an instructor and a student consider important

---

[30]Hadley, op. cit., pp. 305-312.

[31]A. D. Crooks, "Marks and Marking Systems," Journal of Educational Research, XXVII (March, 1933), pp. 259-272.

[32]Mauritz Johnson, Jr., "Solving the Mess in Marks," New York State Education, XLIX (February, 1962), pp. 12-14.

[33]Rugg, op. cit., pp. 117-142.

is a determinant of marks.[34] In testing students the instructor asks questions about the concepts or facts which he considers important. If the student considers the same concepts and facts important, he is better able to prepare himself for the test that the teacher gives and should get a better grade. If the student does not regard as important the things the instructor deems of consequence, the pupil will not be able to prepare himself for the teacher-made tests. Consequently, his grades suffer.

Lobaugh, in an investigation of the relationships of achievement to marks assigned by teachers, reported in his study that girls had a grade point average of 2.19 compared to that of boys, which was 1.97. When he compared scores made on the Myers-Ruch High School Progress Test, the boys' median score was 46, compared with a median of 36 for the girls. In 1940 the girl who became valedictorian of her class on the basis of marks given by teachers ranked thirty-sixth in the class on the Myers-Ruch Test. The salutatorian, who was also a girl, ranked 105th. The total number of students in the class was 250. In 1941 the boy who led the class on the test was classified as a failure in classroom performance. Although girls completely

[34]Robert E. Bills, "The Effects of a Value on Learning," Journal of Personality, XXI (December, 1952), pp. 217-222.

dominated the National Honor Society in the school studied, in 1940 the top fourteen students on the Myers-Ruch Test were boys. In 1941 the top seven were boys, and in 1942 the top four were boys. In 1941 and 1942, only three girls ranked in the first fifteen students as shown by the standardized test.[35] This study was more of a collection of data to raise questions than to answer them. No indication was given of whether the study was designed to be heuristic in nature or whether it simply worked out that way.

Garner found also that girls got higher marks than did boys, although the achievement as measured by standardized scoring procedures showed no significant difference.[36] The study was a comparison of marks given by male and female teachers. No differentiation was made for school subjects; therefore, the procedures are open to criticism.

Swenson concluded after investigating the National Honor Society membership in Lindsborg, Kansas High School for the years 1932-41 that, although he could not find substantial difference in the intelligence of boys and girls, there were more girls than boys in membership. He concluded that membership was gained through inequalities

---

[35]Dean Lobaugh, "Girls' Grades and I.Q.'s," Nations Schools, XXX (January, 1942), p. 42.

[36]C. E. Garner, "Survey of Teachers' Marks," School and Community, XXI (October, 1935), pp. 116-117.

in teacher marks, since teacher marks served as a criterion for invitation to membership.[37] The conclusion could or could not be valid, since membership is based on several criteria that were not a part of his investigation.

Several studies confirmed the idea that girls received better marks in school than do boys. Day reported that girls had a consistent and substantial advantage in obtaining honor marks. There was no evidence that this was the result of intelligence. In his study he noted that the percentage of boys on the honor roll for a twelve-year period was 12.1. For a similar period 29.0 per cent of the girls in the schools surveyed made the honor group.[38] The study presented data in the form of such meaningless statistics as girls' advantage and boys' disadvantage and fulfilled its aim of presenting numbers to indicate that there are more girl honor students than boys. The only evidence that was presented was that girls were getting better marks and receiving more honors.

In 1944 Shinnerer indicated that a boy's chance of failing in school is 2¼ times that of a girl, although

---

[37] Clifford Swenson, "Packing the Honor Society," Clearinghouse, XVI (May, 1942), pp. 521-524.

[38] L. C. Day, "Boys and Girls and Honor Ranks," School Review, XLVI (April, 1938), pp. 288-289.

boys average 2½ points higher on I. Q. tests.[39] The writer freely admits that much of his writing is speculation. He presents failure ratios to prove that more boys fail than girls. The study was similar to other studies presented in that no one explained why this was true.

Conduct, conformity, punctuality, and diplomacy were enumerated among the influencing factors on grades by Douglas.[40] He stated that marks are determined by factors other than achievement, especially those given by women teachers. He noticed a tendency toward overrating girls and underrating boys.[41] In his interpretation of his study, he stated that he did not preclude the possibility or even the probability that girls are more industrious or able and that they actually achieved more than boys. He was not saying positively that girls get better marks because they are girls, because he did not know whether this was the reason or not. In order to determine if girls are assigned better marks because they are girls, it would have been necessary to construct an experiment in which all conditions were controlled

---

[39]M. C. Shinnerer, "Failure Ratio: Two Boys to One Girls," Clearinghouse, XVIII (February, 1944), pp. 264-270.

[40]Douglas and Olson, op. cit., p. 288.

[41]Ibid.

except the independent variable. The effect of the independent variable on the grades then could have been observed.

Wrinkle says that marks may represent an almost unlimited number of factors and indicates that these factors may vary with different classes and different teachers.[42]

It seems apparent that boys and girls are meeting success or failure in our schools on the basis of marks that teachers give to them. However, the marks are based upon some unknown quantities or qualities of unknown ingredients or characteristics.[43] It is not known whether marks are based upon the information factors that a teacher has about a child or to what extent, if any, these factors influence marks. No one has furnished enough information to show why girls apparently receive better grades from teachers than boys when intelligence is measured as equal, nor is the evidence conclusive that they do. Whether effort or attitude of the student affects the mark given by the teacher has never been studied by a precise procedure. No one is even sure what a grade means. Until

---

[42]W. L. Wrinkle, Improving Marking and Reporting Practices in Elementary and Secondary Schools (New York: Holt, Rinehart, and Winston, 1947), p. 92.

[43]Ivan L. Russell, "Personality--Does It Influence Teacher Marks?" Journal of Educational Research, XLVIII (August, 1955), pp. 561-641.

such information becomes available the vital process of administration of schools will be based upon mystical grading procedures. In the following chapter, a design will be presented that will generate data to determine if the information which a teacher has about a student influences the mark assigned to a paper.

III. SUMMARY

Although it is claimed by a few writers that teachers' marks are reliable and do not vary excesssively, the preponderance of studies indicates that marks are unreliable and variable.

Some teachers intend for their marks to be based on achievement only; this is the fixed point system, as defined in Chapter I. Other teachers, using the floating point system, consciously grade with the idea of letting other factors influence the grade. Some of these factors are purported to be personality, socio-economic status, sex, chronological age, grade, ability, effort, and past performance. These allegations are based mainly on beliefs, opinions, philosophic discussions, questionnaires, or observations, rather than investigations based on well-planned research designs. Since statistical procedures have not been followed, empirical evidence is not available to substantiate the validity of their claims. Only a few investigators have controlled conditions so that the independent variables or information factors are manipulated and the effect of the manipulations observed on the dependent variables, or pupils' scores.

CHAPTER III

PROCEDURES

An experimental design was developed to identify some determinants of teacher marks. The design, the procedures used for generating appropriate data, and the statistical procedures used in analyzing the data are presented in this chapter.

I. GENERAL DESIGN

The literature suggests that the following data are considered by teachers in scoring the papers of students: (1) identification of the pupil in terms of his age and grade placement; (2) his sex; (3) his ability: (4) his past achievement; (5) the behavioral background of the child. They may constitute the determinants of teacher marks or may represent only a few of the determinants. The information factors shown here are not given in any particular order. If they can be demonstrated as determinants, further experimentation can evaluate their relative potency.

The purpose for which the work is assigned has also been suggested as a determinant of the mark. Accordingly, if the work were given as a daily paper, the grade should be different, i.e., elevated in comparison to the grade if the work were given as a test.

Most writers agree that some subject areas lend themselves to more valid marking than others. The study also attempted to determine whether or not the structure of the subject, as defined in Chapter One, contributes to variance in marks.

II. RATIONALE

The terms fixed point and floating point have been used in this study in relation to grading. The terms have mathematical derivations. The fixed point term is derived from the practice of working with numbers in mathematics as they are written. For example, in multiplication, using a 10 without exponents, the product is found by moving the decimal one point to the right, e.g., 30 x 10 = 300. This is a fixed procedure, and the process is always the same.

The term floating point is derived from observing a system such as the exponential system of working with numbers. The exponent that is used with the base causes the decimal point to float or vary in either direction. For example, 30×10^2 would cause the decimal to be moved to the right two places, and the product would be 3,000. If the exponent were changed to -2, the decimal would float to the left two places; and the product would be .30.

Similarly, some teachers grade on the basis of an absolute standard of achievement whereas others grade on the basis of a comparison of achievement with the student's ability or other informational factors the teacher possesses. The teachers who mark on an absolute standard may be said to use a fixed point method of grading. Those who mark on a comparative basis may be said to use the floating point system of assigning marks. Certain characteristics distinguish the two methods. The fixed point method presumes a single standard for all. The standardized test is a good example of this method. All students take the test under standard conditions as described in the test directions. Standardized procedures are used in scoring, and the results are compared with a national or local norm for the particular grade. No allowance is made for any characteristic of the child other than age and grade. This allowance is on a group basis.

The floating point method may allow consideration of one or many factors in determining marks. The mark may vary because of the structure of the subject, the nature of the assignment, or informational factors the teacher has about the student. If the marks vary in one direction more than in another, systematic variance is included.

In the areas of information that are said to determine marks, the expectations of the teacher concerning the child are pertinent. If the expectations of teachers are comparatively great for a student, his work will have to be well above average if he is to receive an excellent mark. Conversely, if expectations are low for a student, his work does not have to be so good to receive a good mark. If a floating point system were used, the results would be as indicated above. In the following discussion the floating point system is presumed, along with the presumption that the paper being graded is the same for all teachers. If a fixed point system is used, the grades will not be affected.

According to the literature, the information which teachers have with regard to the age, grade, sex, ability, past achievement, and behavior of students affects the marks assigned by teachers. Certain informational factors should depress the grades given by teachers who use the floating point method:

A. The student who prepared these papers was in the fifth grade and was of the normal age for his grade.

B. The student who prepared these papers was a boy.

C. The student who prepared these papers has consistently scored above 112 on I. Q. tests.

D. The student who prepared these papers has consistently done above-average work.

E. The student who prepared these papers has generally conformed to the rules of the school and has been courteous to teachers.

Other factors of information should increase or enhance the grades assigned by teachers who assign marks on the basis of a floating point method:

A. The student who prepared these papers was in the fifth grade but was older than the ordinary fifth grader.

B. The student who prepared these papers was a girl.

C. The student who prepared these papers has consistently scored below ninety-one on I.Q. tests.

D. The student who prepared these papers has consistently done below-average work.

E. The student who prepared these papers has frequently violated the rules of the school and has been discourteous to teachers on numerous occasions.

Generally, the age and grade of a child are considered together; the child of modal age would be expected to perform at a higher level of efficiency than one who was above the normal age for his grade. The over-age child, by virtue of his age, is thought of as having had difficulties in learning in school and, perhaps, as having been retained in grade for one cause or another. Teachers would expect more from the modal-age child than they would from an older child and, therefore, the grade of the former child would be reduced in a floating point system. In a fixed point system the expectations would be the same for the modal-age pupil and the older.

The literature frequently states that girls receive better grades than do boys. If this is true, the expectations of the teacher would be reduced for girls, and the teacher would assign higher marks than when given the information that the student was a boy.

If a teacher estimates the ability of a pupil as high, the expectations for the pupil are greater; and the grades are depressed by these expectations, unless the work is excellent. The reverse would apply to a low-ability student. The teacher would be less demanding in the marking of a low-ability student, and the grade would be increased.

A student who has achieved satisfactorily in the past would be expected to continue such achievement while a student who has achieved less than satisfactorily would be graded in such a way as to increase his grade. Teachers begin to think the work that is being marked is "good" work for the student whose work consistently has been poor, and he compares current papers with the work he has done previously.

In a floating point system the grade would be increased for those students whose behavior is characterized by nonconformity to rules and standards set by the school. The expectations for such a child would not be so great as for one who has not been a behavior problem. The reasoning here is that a pupil who has been in trouble or who violates rules has probably not been consistent in fulfilling assignments that the teacher has given. The teacher gradually comes to expect less from this student than he does from the student who does as he is directed.

Following the same mode of thought, the teacher would expect the student to do better on a test than he would on daily papers. The higher expectations of the teacher would cause a lowering of the marks assigned to test papers as compared with marks assigned to the same paper designated as a daily paper.

Certain subject areas are more compatible with the fixed point system than others. Subjects which have a rigid structure apparently lend themselves to a fixed point system; whereas, subjects which have a loose structure apparently are more adaptable to a floating point system.

III. RESEARCH DESIGN AND DATA

Data were gathered by requesting the cooperation of administrators in the participating districts. The teachers who were nominated by their school administration and who agreed to participate were sent an exact copy of an arithmetic and an English paper which had been selected from a fifth grade classroom. The papers had been reproduced by an electronic stencil-maker. (See Appendix A for samples of papers.)

The teachers were instructed to grade the papers on the basis of 0-100 points, with 70 considered passing. Each teacher was supplied a copy of the two papers (one English and one arithmetic), along with information about the student who prepared the paper. The information was designated as Categories A, B, or C. The teachers who received papers in Category A were given no information about the student who prepared the papers. These teachers were asked merely to grade the papers on the scale described above. (See Appendix A for sample instructions.) The information in Categories B and C was fictional and was generated as experimental stimuli.

One-half of the teachers who received Category B information and one-half of the teachers who received Category C information were told that the papers were daily papers. An equal number in each group was informed that the papers were test papers.

Teachers who received Category B information were given information about the student which, according to the rationale, should reveal a floating point system if it were used. The information in Category B consisted of the information listed in the rationale that was designed to depress the assigned grades. It was reasoned that teachers would score these papers in such a manner that the average grades would be lower than for papers designated as Category C papers.

Teachers who received Category C information were given information about the student who prepared the papers which, in a floating point system, should enhance the scores. Category C information consisted of the information listed in the rationale that was supposed to increase the scores so that the average scores would be higher than the scores assigned by teachers who were given Category B information.

With regard to the designation of papers as daily and test papers, it was reasoned that those papers that were designated as test papers would be considered more important; hence, their average score would be lower. The design in Figure 1 generated data to test the null hypotheses which follow, concerning papers in English and arithmetic:

Ho_1 No statistically significant difference exists at the .05 level of confidence among the means of scores of arithmetic papers graded by teachers whether they were given Categories A, B, or C information about the student.

Ho_2 No significant difference exists at the .05 level of confidence between scores of papers marked by teachers whether they were given information that the papers were test papers or were given the information that the papers were daily papers.

When Ho_1 or Ho_2 was rejected, Stage II, presented in Figure II of the design, became operative.

The following hypotheses were tested for arithmetic and English:

Ho_3 No statistically significant difference exists at the .05 level of confidence between marks given by teachers whether they were given Category B or Category C information as to the age and grade of the student.

Ho_4 No statistically significant difference exists at the .05 level of confidence between marks given by teachers whether they were given Category B or Category C information with relation to the sex of the student.

Ho_5 No statistically significant difference exists at the .05 level of confidence between marks given by teachers whether they were given Category B or Category C information with relation to the ability of the student as indicated by I. Q. scores.

Ho_6 No statistically significant difference exists at the .05 level of confidence between marks given by teachers whether they were given Category B or Category C information with relation to past achievement of the student.

Ho_7 No statistically significant difference exists at the .05 level of confidence between marks given by teachers whether they were given Category B or Category C information with relation to the behavior of the student.

FIGURE I

RESEARCH DESIGN TO TEST HO_1 AND HO_2 FOR ARITHMETIC AND ENGLISH PAPERS

CATEGORIES OF INFORMATION	MATERIAL GRADED	
	Daily	Test
A (No information)		
B (Information to depress grades)		
C (Information to elevate grades)		

FIGURE II

RESEARCH DESIGN TO TEST HYPOTHESES
THREE THROUGH SEVEN

TYPES OF INFORMATION	CATEGORIES OF INFORMATION	MATERIAL GRADED Daily	Test
Age-Grade	B (Depress)		
	C (Elevate)		
Sex	B (Depress)		
	C (Elevate)		
Ability	B (Depress)		
	C (Elevate)		
Past Achievement	B (Depress)		
	C (Elevate)		
Behavior	B (Depress)		
	C (Elevate)		

<u>Statistical treatment of data</u>. The data were statistically treated by a multiple classification analysis of variance which provides a test of differences among the means for several groups simultaneously. Extension of the basic procedures permits the analysis of the interaction among the variables as well as for the main effects.

The basic concept or assumption underlying analysis of variance is that sums of squares of between groups and within groups are additive. By means of algebraic manipulation it can be shown that the total sum of squares is equal to the within-groups sum of squares plus the between-groups sum of squares. By dividing each sum of squares by the number of scores or frequencies which can vary around a parameter, the mean square or sample variance is found. It has been shown by Snedecor that the ratio of the mean square between groups to the mean square within groups follows a specific sampling distribution which he labeled the distribution F.[1] In analysis of variance, generally, the numerator of the F ratio is the estimate of variance that arises from the variations in the treatments being tested. The denominator is sometimes referred to as the error term. If the obtained F ratio is larger than the one required for the 5 per cent level of significance, the null hypothesis is rejected, permitting one to conclude that, in 95 out of 100 similar cases, the treatments resulted in measurable differences.[2]

A significant F would indicate that non-chance variations exist among means somewhere in the cells. It does not give information as to how many means are significantly different.[3]

---

[1]Deobold B. Van Dalen and William L. Meyer, Understanding Educational Research (New York: McGraw-Hill Book Company, Inc., 1962), p. 322.

[2]Ibid., p. 323.

[3]Guilford, op. cit., p. 262.

In the data that were explored, if a null hypothesis was rejected, Scheffe's test was applied to determine which means varied significantly from the pooled data. In this test the standard error of a difference between means was computed. Significant t's at the .05 and .01 levels of confidence were found for the appropriate degrees of freedom. The products of t.05 and the S.E. and t.01 and the S.E. were computed and compared with the differences between means to test for significance.

In testing whether observed variance was reduced, the significance of the ratio of two variances was tested, instead of testing the differences in standard deviations. An estimate of the population variance from each of the two samples was found and the ratio was computed for these values. The null hypothesis was assumed that the two samples came from the same population, and the question was asked if two estimates of that population could differ as much as the obtained ratio indicated.[4]

IV. THE SAMPLE

The sample was selected to satisfy the research design. The total sample size was 225 teachers. The sample size was determined after a pilot study revealed that each sub-group would require seventy-two teachers

---

[4]*Ibid.*, p. 224.

in order to allow detection of a statistically significant difference among the information factors. The required sample size, or n, was computed by the formula $Sdm = \sqrt{\frac{2Vw}{n}}$ which was transformed algebraically into its equivalent: $n = \frac{2Vw}{(SDm)^2}$ in order to secure an estimate of minimum sample size.[5]

Criteria and rationale for selection of cities. Cities were selected on the following criteria:

1. Each school was located in a city or a metropolitan area of 100,000 to 400,000 population.
2. At least one city was invited to participate in the study from each state that had metropolitan areas or cities of 100,000 to 400,000 population.

Cities or metropolitan areas of this size were selected because of the probability of response. Schools in these cities usually will have a research department and have responsibilities defined so that studies can be facilitated. Since smaller districts are seldom organized for research, it is doubtful that they would have responded.

An example is a reply received from a superintendent of schools who was invited to participate in the study through his director of research: "For the low pay scale your request offers, and with the big work load our teachers report, do you think the Director of Research will get any

---

[5] Ibid., p. 264.

enthusiastic takers? Since we have no Director of Research, I am talking about me!"

Many of the larger districts have projects of their own, and regulations that make it difficult for them to participate in other research. An example is a reply from a large city school: "In accordance with current regulations of the board of education, it is requested that you complete the enclosed application form to conduct research in the schools."

Size of sample. Twenty schools from the original sample responded by submitting the names of ten teachers who agreed to cooperate in the study. Of the 200 teachers who were mailed papers to score, 164 responded in an acceptable manner. The sixty-one participating teachers that did not come from the original sample met the criteria for the sample in every respect. Their cooperation was solicited in a second sampling. All geographic areas of the continental United States were represented. (See Appendix B for participating schools.)

Criteria and rationale for selection of teachers.

1. Each teacher was a female.
2. Each teacher was a fifth grade teacher at the time of marking the paper.
3. Each teacher had two to fifteen years of experience.
4. Each teacher was rated good to excellent by her supervisor on the basis of records in the

administrative offices of the participating district.

Female teachers were specified because 87.8 per cent of all elementary teachers are female.[6] A sample balanced for sex would have been difficult to acquire, especially since the sex-ratio varies widely for different sections of the country. Since the literature suggested that sex differences among teachers probably exist in scoring of papers by teachers, the entire sample was comprised of one sex. An apparent difference in the way teachers mark is that female teachers tend to favor girls in the assignment of marks more than do male teachers.[7]

The fifth grade was chosen because of its place in school organization. It is considered a part of the upper elementary school in most districts. The fifth grade is usually organized on a self-contained basis in most schools; thus, the sample would include the teachers from self-contained classrooms, rather than subject-matter specialists interspersed with elementary teachers. The fifth grade teacher scores papers on a basis which fits the purpose of this study better than a teacher of a lower grade. Teachers in lower grades often grade only on a basis of satisfactory or

---

[6] National Education Association Research Memo, XXI (July, 1964), p. 1.

[7] Douglas and Olson, op. cit., pp. 282-288.

unsatisfactory; while fifth grade teachers are accustomed to marking on a scale of at least five symbols.

In order to secure a more homogeneous sample, neither beginning teachers nor more experienced teachers were included. Also, the teachers who were rated below good by their supervisors were eliminated from the sample by the cooperating districts. Hopefully, the sampling by teachers' age, experience, and proficiency should have reduced variability in grading caused by these factors. The variability in grading caused by the information that teachers had about the student became more meaningful in terms of this study.

The statistical results of the study along with a discussion of the results, are presented in the ensuing chapter.

V. SUMMARY

An experimental design was developed to generate data to test first the null hypothesis that only chance differences exist among the scores of papers graded by teachers whether the teacher was given Category A information, Category B information, or Category C information. Since this was an exploratory study, the .05 level of confidence was selected as evidence that the factors being compared were homogeneous. The design was a sequential one consisting of two stages. In cases where the null hypothesis was rejected, other null hypotheses were tested to determine if only chance differences existed at the .05 level of confidence between scores of papers marked by teachers whether a teacher was given Category B information or Category C information with regard to the student's age-grade status, sex, I.Q., past achievement, and behavior at school. This was to determine if any of the information factors were a determinant of teacher marks.

The design also generated data to test the null hypothesis that only chance differences exist at the .05 level of confidence between scores of papers marked by teachers whether they were given information that the papers were test papers or that the papers were daily papers.

A homogeneous sample of teachers representing all geographic areas of the United States was selected to score a paper in arithmetic and a paper in English. The papers to be graded were reproductions of a single arithmetic and an English paper done by an average fifth grade student.

The data were analyzed by a multiple classification analysis of variance. When a significant F was found, Scheffe's test was applied to determine significant differences between means.

CHAPTER IV

RESULTS AND DISCUSSION

The objectives to which the study was devoted were the following:

1. To determine if the information that a teacher has about a student affects the marks assigned by a teacher, either by depressing or elevating the score.

2. To determine if the information that a teacher has about a student reduces observed variance in assigned marks.

3. To determine whether teachers grade on a fixed point or floating point basis.

4. To determine if variance in marking is greater in the more loosely structured subject areas such as English or in the more rigidly structured subject areas such as arithmetic.

5. To determine if the nature of the assignment affects the grades that are given in arithmetic and English, i.e., the designation of a paper as a test paper or a daily paper.

I. RESULTS OF THE STUDY

The data that were treated first consisted of the scores which were generated by the research design shown in Figure I, Chapter III, separately for arithmetic and for English. In each case, the data were treated by a multiple classification analysis of variance, which allows the examination of the potency of different treatments simultaneously.

The categories of information were as follows: (1) Category A teachers had no information about the student; (2) Category B teachers had information presumed to depress the grade; (3) Category C teachers had information presumed to elevate the grade.

It had been reasoned that the mean score for the group of teachers who scored papers having Category B information would be below that of the Category A group and that the Category C information group would have been above the Category A group. As can be seen in Table I, this was not the case. Both Category B and Category C information groups had means above the Category A or no-information group, with Category C having the highest mean.

<u>Arithmetic scores: gross information versus no information</u>. The null hypothesis concerning the influence that information has on teachers in assigning marks was expressed:

Ho_1 No statistically significant difference exists at the .05 level of confidence among the means of scores of arithmetic papers graded by teachers whether they were given Categories A, B, or C information about the student.

The F ratio was .92, which was not significant at the .05 level of confidence. The null hypothesis was not rejected. It should be pointed out that the means of the three information categories ranged from 74.61 in the category of no information to a mean of 76.85 in the category of information that was thought to increase grades.

TABLE I

MEANS AND STANDARD DEVIATIONS FOR ARITHMETIC PAPERS SCORED BY TEACHERS UNDER EXPERIMENTAL CONDITIONS

INFORMATION CATEGORY	N	MEAN	STANDARD DEVIATION	VARIANCE
A (No information)	75	74.61	12.61	158.61
B (Depress)	75	75.82	9.15	83.80
C (Elevate)	75	76.85	7.90	62.47

<u>Arithmetic scores--reduction of variance</u>. The significance of the ratio of two variances was tested instead of testing the significance of difference between two standard deviations. An estimate of the population variance from each of two samples was found and the ratio computed for these values. The null hypothesis assumed that the two samples came from the same population, and the question was asked if two estimates of that population could differ as much as the obtained ratio indicated. The ratio has been given the symbol F and was computed from the formula: $F = \frac{\text{larger variance}}{\text{smaller variance}}$

The estimated variance was computed by the usual method.

The application of the F test rests upon the assumption that the population is normally distributed.[1] The variance was examined even though the means were not significantly different. A reduction in observed variance would show a decrease in the spread of the scores and indicate an increased reliability.

The observed variance as shown in Table I was found to be greater in the scores given by teachers grading without information about the student. Variance was significantly reduced when information was given to the teacher about the student whose arithmetic paper he was scoring. In Table II the F ratio was significant at the .01 level of confidence when variance of negative and positive information categories were compared with the no-information category. When negative and positive information variances were compared, the difference was not significant.

English scores: gross information versus no information. The same null hypothesis was presented with regard to the scoring of English papers:

Ho_1 No statistically significant difference exists at the .05 level of confidence among the means of scores of English papers graded by teachers whether they were given Category A, Category B, or Category C information about the student.

---

[1]Guilford, *op. cit.*, p. 224.

The differences among means for these categories were also explored by analysis of variance. The F ratio was 4.98, which was significant beyond the .01 level of confidence.

TABLE II

REDUCTION IN OBSERVED VARIANCE BETWEEN INFORMATION CATEGORIES IN SCORING OF ARITHMETIC PAPERS

CATEGORIES	F
$\frac{\text{A (No Information)}}{\text{B (Negative Information)}}$	1.89**
$\frac{\text{A (No Information)}}{\text{B (Positive Information)}}$	2.54**
$\frac{\text{B (Negative Information)}}{\text{C (Positive Information)}}$	1.34

** Denotes significance at .01 level of confidence.

The null hypothesis was rejected; non-chance variations apparently existed among the means for the categories that are shown in Table III.

<u>English papers</u>: <u>Scheffe's test</u>. Scheffe's test was applied to determine where differences among the means were significant. The standard error of the means was computed to be 1.51. Through the application of Scheffe's test it was found that a difference between means of 3.01

was significant at the .05 level of confidence, and a difference between means of 3.99 was significant at the .01 level of confidence.

TABLE III

MEANS AND STANDARD DEVIATIONS FOR ENGLISH PAPERS
SCORED BY TEACHERS
UNDER EXPERIMENTAL CONDITIONS

INFORMATION CATEGORY	N	MEAN	STANDARD DEVIATION	VARIANCE
A (No Information)	75	71.72	10.10	101.96
B (Depress)	75	75.53	8.12	66.00
C (Elevate)	75	71.15	9.24	85.32

Table IV shows that the differences between the means of B and A were significant at the .05 level of confidence. The difference between the means of B and C was at the .01 level of confidence.

English papers: subsequent analysis. Since Ho_1 was rejected for the scoring of English papers, it became proper to use Stage II of the design shown in Figure II, Chapter III. Null hypotheses three through seven, as shown below, were tested for difference by analysis of variance for the scoring of English papers:

Ho_3 No statistically significant difference exists at the .05 level of confidence between the means of marks given by teachers whether they were given Category B or Category C information as to the age and grade of the student.

Ho_4 No statistically significant difference exists at the .05 level of confidence between the means of marks given by teachers whether they were given Category B or Category C information with relation to the sex of the student.

Ho_5 No statistically significant difference exists at the .05 level of confidence between the means of marks given by teachers whether they were given Category B or Category C information with relation to the I.Q. of the student.

Ho_6 No statistically significant difference exists at the .05 level of confidence between the means of marks given by teachers whether they were given Category B or Category C information with relation to past achievement of the student.

Ho_7 No statistically significant difference exists at the .05 level of confidence between the means of marks given by teachers whether they were given Category B or Category C information with relation the behavior of the student.

The analysis was performed using four random samples of Category A information as controls along with the five types of information found in Category B. The means, standard deviations, and variances are shown in Table V. The F was computed as 1.45. An F of this magnitude did not permit rejection of the hypotheses of no difference among means; whether the teachers were given information or not had no measurable effect upon the average grades awarded to the papers in this study.

TABLE IV

COMPARISON OF MEANS TO DETERMINE SIGNIFICANT DIFFERENCES

DIFFERENCES BETWEEN MEANS	
A - C	.57
B - A	3.81*
B - C	4.38**

*Denotes significance at the .05 level of confidence.
**Denotes significance at the .01 level of confidence.

An interpretation of the data is that the information given to teachers did cause scores in Category B to be significantly higher than Categories A and C. However, subsequent analysis using five informational components against random samples of Category A did not show that a single item of Category B was potent enough to be significantly different.

English papers: reduction of variance. In Table VI the exploration of the observed variances among the three treatment groups is presented. The finding was that the information given to teachers reduced the variance in the scoring of English papers between Category A and Category B information. The observed variance was significantly

TABLE V

MEANS AND STANDARD DEVIATIONS FOR ENGLISH PAPERS SCORED BY TEACHERS IN CATEGORY B BY TYPES OF INFORMATION

	CONTROL GROUPS				EXPERIMENTAL GROUPS				
	A No Information	A No Information	A No Information	A No Information	B Age-Grade	B Sex	B Ability	B Past Achievement	B Behavioral Background
N	9.00	9.00	9.00	9.00	7.00	9.00	9.00	9.00	8.00
Mean	65.22	67.67	68.55	70.44	70.71	71.88	70.33	74.77	79.50
Standard Deviation	9.85	10.95	12.07	8.59	6.86	11.19	9.82	8.46	3.64
Variances	97.06	120.00	145.58	73.80	47.06	125.21	96.44	71.51	13.25

reduced as teachers were given information designed to depress the grades of English papers when compared to the variance of teachers' grades who were given no information.

TABLE VI

REDUCTION IN OBSERVED VARIANCE BETWEEN INFORMATION CATEGORIES IN SCORING OF ENGLISH PAPERS

CATEGORIES	F
$\frac{\text{A (No Information)}}{\text{B (Negative Information)}}$	1.57*
$\frac{\text{A (No Information)}}{\text{B (Positive Information)}}$	1.20
$\frac{\text{C (Positive Information)}}{\text{B (Negative Information)}}$	1.29

*Denotes significance at .05 level of confidence.

Of the five types of information given in Category B, only one reduced the variance significantly, and that beyond the 1 per cent level of confidence. (See Table VII.) The information which was given to teachers concerning the student who composed the paragraph in English reads as follows:

> "The student who prepared these papers has generally conformed to the rules of the school and has been courteous to teachers."

It had been reasoned that this information would increase the expectations for teachers who used a floating point method of grading. Actually, while the grade did not increase significantly, the variance reduction was significant. The information caused the teachers to agree more closely that a student who is not a behavior problem should receive a good score.

TABLE VII

REDUCTION IN OBSERVED VARIANCE BETWEEN TYPES OF INFORMATION IN SCORING OF ENGLISH PAPERS

TYPES OF INFORMATION	F
$\frac{\text{A (No Information)}}{\text{B Age-Grade (Negative)}}$	2.17
$\frac{\text{B Sex (Negative)}}{\text{A (No Information)}}$	1.23
$\frac{\text{A (No Information)}}{\text{B Ability (Negative)}}$	1.06
$\frac{\text{A (No Information)}}{\text{B Past Achievement (Negative)}}$	1.43
$\frac{\text{A (No Information)}}{\text{B Behavior (Negative)}}$	7.70**

**Denotes significance at .01 level of confidence.

<u>Arithmetic papers</u>: <u>test versus daily</u>. In order to determine if the nature of the assignment affects scores

assigned to papers, the following hypothesis was tested for arithmetic papers:

Ho_2 No significant difference exists at the .05 level of confidence between means of scores of arithmetic papers marked by teachers whether they are given information that the papers were test papers or daily papers.

By analysis of variance it was determined that, for the mean scores on arithmetic papers between the two gradients, the F ratio was 2.68. This was not significant at the .05 level of confidence; hence, the null hypothesis was not rejected. Means and standard deviations are shown in Table VIII.

TABLE VIII

MEANS AND STANDARD DEVIATIONS FOR ARITHMETIC PAPERS SCORED BY TEACHERS WITH INFORMATION AS TO THE NATURE OF THE ASSIGNMENT

	NO INFORMATION	TEST PAPER	DAILY PAPER
N	75.00	75.00	75.00
Mean	74.61	74.60	77.43
Standard Deviation	12.61	12.62	7.81
Variance	158.61	146.64	61.00

<u>Arithmetic papers: test versus daily; reduction in variance</u>. The observed variance in daily-paper scores was compared with variance in no-information paper scores and test-paper scores, and was found to be significantly different beyond the .01 level of confidence. (See Table IX.)

TABLE IX

REDUCTION IN OBSERVED VARIANCE BETWEEN ARITHMETIC PAPERS DESIGNATED AS TO NATURE OF PAPER

NATURE OF PAPER	F
Test Paper / No Information	1.01
No Information / Daily Paper	2.60**
Test Paper / Daily Paper	2.61**

**Denotes significance at .01 level of confidence.

<u>English papers: test versus daily</u>. The same hypothesis was tested for the scoring of English papers and an F ratio of 2.85 was computed, which was not significant at the .05 level of confidence; hence, the null hypothesis was not rejected. Means and standard deviations are shown in Table X.

TABLE X

MEANS AND STANDARD DEVIATIONS FOR ENGLISH PAPERS SCORED BY TEACHERS WITH INFORMATION AS TO THE NATURE OF THE ASSIGNMENT

	NO INFORMATION	TEST PAPER	DAILY PAPER
N	75.00	79.00	79.00
Mean	71.72	72.16	75.06
Standard Deviation	10.10	9.24	8.97
Variance	101.96	85.38	80.46

English papers: test versus daily; reduction of variance. The observed variance was also analyzed for English papers scored by teachers who were given the information that the papers were daily papers or test papers. The information that the English papers were test papers or daily papers did not reduce the observed variance significantly. (See Table XI.)

Structure of the subject: English versus arithmetic; reduction of variance. The observed variance was analyzed to determine if variance was greater or less in the more rigidly structured subject areas or in the more loosely structured subject areas. Table XII shows the result of this comparison. When teachers had no information about the student, the variance was significantly greater in the more rigidly structured subject, i.e., arithmetic. When teachers were given information about the student, variance was reduced for both subject areas and no difference was shown.

TABLE XI

REDUCTION IN OBSERVED VARIANCE BETWEEN ENGLISH PAPERS DESIGNATED AS TO NATURE OF PAPER

NATURE OF PAPER	F
$\frac{\text{No Information}}{\text{Daily Paper}}$	1.20
$\frac{\text{No Information}}{\text{Test Paper}}$	1.27
$\frac{\text{Test Paper}}{\text{Daily Paper}}$	1.06

TABLE XII

COMPARISON OF OBSERVED VARIANCE BY SUBJECTS

STRUCTURE OF SUBJECT	F
Arithmetic (No Information) English (No Information	1.56*
Arithmetic (Depress) English (Depress)	1.27
English (Elevate) Arithmetic (Elevate)	1.37

*Denotes significance at .05 level of confidence.

II. DISCUSSION OF RESULTS

Arithmetic scores: gross information versus no information. The hypothesis that was tested concerning information about the student as opposed to no information about the student was not rejected. This means that the obtained difference between scores assigned to arithmetic papers by teachers who were given no information, information designed to depress the scores, and information designed to enhance the scores, was not significant. The average scores for the three categories of information ranged from 74.61 to 76.85; the higher average score was from the category of information that was designed to increase the score. There were more than five chances in one hundred that a difference as large as shown here could have occurred by random sampling from the same population.

It is improbable that the information a teacher has concerning a student's age-grade, his sex, his ability, his past achievement, and his behavior affects the marks assigned by teachers to arithmetic papers, either to elevate or depress the grade.

A reduction of observed variance in the scoring of arithmetic papers was shown when any information was given about the student who prepared the paper. The range of arithmetic scores assigned by teachers with no information about the student ranged from 25 to 100. In the other categories in which teachers were given information about the student, the scores ranged from 50 to 98. The reduction in observed variance shows that the scores which teachers assigned to the same paper did not vary so much from the mean. Teachers agreed more closely on the scores given to the papers. This increased agreement suggests a decrease in error and a gain in reliability when teachers possess information about the student.

Arithmetic scores: daily versus test. A similar explanation to that given for the null hypothesis concerning information about the student would apply to the hypothesis as to the nature of the assignment. The obtained difference between scores assigned to arithmetic papers by teachers who were told that the paper was a test paper and those who were told that the paper was a daily paper was not significant. It was probable that the nature of the assignment did not influence the marks given by teachers.

The information that a paper was a daily arithmetic paper reduced variance significantly in comparison to the information that the paper was a test paper and to the condition of no information. Apparently there is more agreement among teachers in the scoring of a daily paper than in scoring a test paper in arithmetic. Possibly teachers in grading daily arithmetic papers are not so concerned about the grade given and do not scrutinize the paper so closely as they do test papers. It is also possible that the teachers grade more for motivation in scoring daily work than when tests are graded. When teachers grade tests, they may grade more for information about what the student has learned.

English scores: information versus no information. In the exploration of the English scores, the hypothesis concerning the information as opposed to no information was rejected. The rejection of the null hypothesis indicated that the information the teachers were given about the student did cause a significant difference in the scoring of an English paper. It had been reasoned that

Category B information would cause the mean scores of the English papers to be decreased because the information would lead the teachers to expect more. The mean scores were increased significantly for Category B as compared to Category A and Category C. This could indicate that teachers were not using a floating point method of grading geared to expectations of the student. On the contrary, they graded the way the category suggested. If the age-grade, the sex, the ability, the past achievement, and the behavior of the student were favorable, expectations were not increased for those students. The suggestion that they were good students caused the teacher to react favorably and to award a higher grade.

The analysis of variance that was subsequently used to test null hypotheses three through seven failed to reject the hypotheses that no significant difference existed among the means by types of information. It is improbable that any one type of information that was given the teacher about the student affected the marks assigned by teachers. The combined negative information about the student affected the marks assigned by the teachers significantly; but no one item of information was potent enough to distinguish it from the other items of negative information. It was noted that means ranged from 70.33 to 79.50. The lowest mean score was assigned by the teachers who were given information that the student had consistently scored above 112 on I.Q. tests. Had this been a significant difference the rationale of the study would have been supported. The highest mean score was assigned by teachers who were given information that the student had conformed to the rules of the school and had been courteous to teachers. Had this been a significant difference, the rationale of the study would not have been supported.

The observed variance was reduced when teachers were given negative information about the student as compared to no information about the student. Positive information about the student apparently failed to reduce variance. This supports the contention that, when teachers see information about a student that indicates he is a good student, they agree more readily on the grade to be assigned.

A further analysis of the five types of negative information about the student revealed that the observed variance among these types of information was reduced significantly when the information was given the teacher that the student conformed to the rules of the school and was courteous to teachers. The agreement was great and was in the direction of elevating the grade.

English scores: test versus daily papers. Null hypothesis two was not rejected in the exploration concerning the scoring of English papers on the basis of the nature of the assignment. It was probable that no differences existed in the scoring of English papers because of the nature of the assignment.

Variances were compared and no reduction in variance was observed in the scoring of English papers through giving the information that the paper was a daily paper or a test paper.

Structure of subject. The question of whether greater variance existed in the scoring of English papers or the scoring of arithmetic papers was not tested by a null hypothesis. The range of scores in arithmetic was from 25 to 100; whereas, the scores in English ranged from 50 to 90. In both cases the extreme scores were in the no-information category.

The variances of the distributions provided an idea as to whether the structure of the subject that was being graded was important in the assigning of marks. It was not apparent that a reduction in variance was a function of the structure of the subject, except when no information was given in either subject. The variance then was greater in the more rigidly structured subject.

Floating point versus fixed point. Although many teachers insisted that the information about the student was important and that they could not grade papers without knowing such things as the age-grade of the student, his ability, etc., little evidence was shown of teachers' use of the floating point system. Regardless of the information given about the student and the information as to the nature of the assignment, the only information that caused the means to differ was negative information in the scoring of English papers. The mean in that case was significantly higher, indicating the probability that teachers did not gear expectations to the information about students as the literature and teachers had indicated.

It had been anticipated that if the expectations for a student were great and he presented an "average" paper, the score would be reduced. If the expectations for a student were meager, the score for an "average" paper would be increased. In the cases where a floating point system was observed, the point apparently floated with the reputation of the student rather than the expectations for the student. If the student's reputation was good, the mean score increased instead of decreasing. If the student's reputation was not good, the mean score was not increased as had been theorized.

In the succeeding chapter an analysis of the comments of teachers who participated in the study is recorded.

III. SUMMARY

The data were treated by a multiple classification analysis of variance and by variance ratios with the following results:

Arithmetic. It was improbable that the information a teacher had concerning a student's age-grade, his sex, his ability, his past achievement, and his behavior significantly affected the marks assigned by the teacher. It was also improbable that the information the teacher had as to the nature of the assignment, i.e., whether it was a test or a daily paper, affected the scores assigned by the teacher.

English. The information a teacher had concerning a student's age-grade, sex, ability, past achievement, and behavior collectively affected the mark given by a teacher. When each of the factors was considered separately, it appeared improbable that any of them alone was potent enough to significantly affect the marks given by teachers.

It was improbable that the information that the teacher was given, with reference to the nature of the assignment, affected the mark which the teacher assigned.

No apparent difference in observed variance existed in the English paper scores because of more permissible alternatives than existed in the arithmetic paper scores.

The observed variance was reduced in scoring arithmetic papers when information about the student was given to the teacher. Variance was also reduced when the teacher was told the nature of the paper, i.e., whether it was a test or daily paper.

The observed variance was reduced in scoring English papers if negative information about a student was given to the teacher. Especially was this true if negative information was given as to the conduct of the student.

CHAPTER V

REVIEW OF TEACHERS' WRITTEN COMMENTS

The information and analysis contained in this chapter were not a part of the original purpose of the study. It became evident as the data were collected that the gratuitous notations of participating teachers could make a contribution to the study. According to the instructions given the teachers, only a grade was to be assigned. This did not prevent the issuance of the notes which are presented in this chapter.

As has been previously stated, the 225 teachers who participated in the study were female teachers with from two to fifteen years of teaching experience and, who were rated from good to excellent by their supervisors. Many of them expressed appreciation for being given the opportunity to participate in a study concerning the marking of students' papers. They indicated that a need existed for standardization of grading procedures and agreed that confusion does exist in this area.

I. ANALYSIS OF NOTATIONS BY TEACHERS

The fact that written comments were elicited from the participating teachers was unexpected in view of the full schedule of activities required of them. No suggestions were made to the teachers concerning the writing of notes to the mythical student about whom information was given; yet many of the teachers wrote notes to this student offering help if the marking was not clear. They also wrote other comments which were designed to be helpful to the child. Some were in the nature of praise for the work which had been done; others were critical. Nearly all of the teachers wrote to the student or to the researcher. The comments of teachers, in some cases, substantiated the statistical findings of the study and, in other cases, contradicted the findings.

Teachers who returned papers not graded. Several teachers who agreed to participate in the survey sent the papers back ungraded. A variety of reasons was given for their actions.

One teacher marked the English paper, but said that she could not score the arithmetic paper because arithmetic was not in her field. This indicated a degree of specialization that does not ordinarily exist at the fifth grade level. Although the arithmetic problems were simple fifth grade problems, she did not feel capable of assigning a mark to them. Teachers ordinarily score an arithmetic paper with more confidence than they score an English paper because of the relatively few alternatives which a teacher of arithmetic has.

One teacher refused to score the papers on the basis that she had never used a 100-point scale before. She also indicated that 0 and 100 were not defined; therefore, all points between were meaningless. This, of course, is in keeping with the contention of some of the writers in the field of grading. This teacher rebelled at the thought of putting a point value on an idea belonging to another person, as in the English paper. She stated also that she would have to know the child's ability. In her review of the papers she said that the English paper was below average in sentence structure and above average in spelling. She did not, however, define average. She would not make judgments beyond that. With regard to the arithmetic paper, she ascertained readily that one of the problems was incorrect. However, that is all she would say about the arithmetic paper even though she gave the problems to her class and the paper fell in the lower quarter of her class. Although she did not put a score on the papers, she was eager to get the results of the study.

Another teacher wanted criteria for evaluation such as these: (1) Am I grading for correctness of the problem? (2) Is form and neatness a factor? (3) Do I grade on originality? She returned the papers without grading them.

A teacher reported that she could not score the papers because she had not used points in years. She stated that, in arithmetic, students should be told only the number missed. This does not give the student much information about his achievement. She indicated that grades are given in her class only on a test paper.

Some teachers gave no reason for not scoring the paper. They simply stated that they could not, although they were fifth grade teachers and the papers were taken from a regular fifth grade classroom. They probably were required to give grades in their own classes. Other teachers in the same school district did not mention experiencing difficulties but scored the papers and returned them.

One teacher said that she would have to know the following information before she would be able to score the papers: (1) the teacher's purpose, (2) what the teacher had stressed with the class, (3) how the writing was stimulated, and (4) the student's ability and past achievement.

Teachers who had difficulty in grading. One teacher noted that no mention was made on her paper of I.Q., socio-economic level, or status with peer group of the student whose paper she was marking. The inference was that she considered all of these factors in assigning marks.

One teacher was frank in saying that the job would be less difficult if she knew the author of the paper.

Another teacher expressed herself as being "unfair" to the student because she did not know what skills had been taught, or what was being tested. She gave the student a grade of seventy-five in arithmetic and English. This teacher's suggestion that she was "unfair" in grading a student without knowing what was being tested reveals insight into the process of marking. She evidently understood that the first step in evaluation occurs in the selection of objectives for the course. This teacher seemed to understand that, if she were grading for achievement in learned skills, the score would be different from a score that was given on the basis of concepts learned.

Several teachers said that they needed more information about everything. The need for "everything" was a difficult remark to classify. It is possible that the teachers felt that they needed more information but could not specify what information they needed. Thus, it may be that the request for information reflected a general anxiety or discomfort in grading.

The possibility that there were Mexican children enrolled in Texas schools caused one teacher to suggest that scoring on a curve would be more realistic. She apparently thought in terms of a floating point system, which would allow for the language handicap.

Several teachers voiced difficulties in scoring due to the lack of knowledge concerning the circumstances of the preparation and working periods. This indicated a need to know how much time the student had spent in learning the skills that were being graded, whether the student had been helped by other children, whether the student was ill at the time he prepared the paper, or whether there were other modifying circumstances. These teachers also subscribe to the floating point method of grading.

One teacher mentioned that grade level was not stated. She thought this knowledge would be helpful.

Along this line of reasoning, another teacher said that she could have done better if a complete set of papers had been sent to her. This would indicate that she uses a comparative basis for marks with the rank-order system of grading, which is another version of the floating point method.

Several teachers commented that they needed to know the teacher's objectives in assigning the work, previous standards that had been set up, and information as to the kind of motivation given prior to the assignment. However, each of these teachers cooperated in the study. It seemed that the teachers who expressed difficulty in grading the papers used a floating point system of grading. They were asking for more information in order to follow their procedure of grading according to the student's characteristics.

Teachers who reflected insecurity in marking. It was noted that many teachers who refused to score the papers, and others who did score the papers, felt insecure in their scoring. Many were apologetic in their comments:

"I wish I could be of more help to you in your search for a solution to the problem of how to grade. It has troubled me for years."

"I feel that the enclosed papers that were sent to me by you would almost have to be graded on a subjective rather than on an objective basis. . . ."

"Frankly, as far as I am concerned, grading would be done away with, and reports of the child's progress would be given to parents."

The teacher did not state in what manner the child's progress would be reported. It is possible that she favored parent conferences or informal letters as a means of conveying information to the parents about their child. The conference method of reporting does allow teachers to interpret facts about the child as well as learn from the parent facts about the child at home. Such conferences take an extra two or three hours of teacher time per pupil per year and also involve the time of parents. The informal letter is probably less time-consuming, and parents of each child can be reached more surely. This task also can become burdensome to the teacher and can be costly at five cents per letter for stamps alone, plus the cost of the stenographer and supplies.

One teacher apparently felt that insecurity of teachers would continue, for she said, "I sincerely believe that it is impossible to keep grades from being confusing." She hoped that the current study would be of help.

Teachers who exhibited confusion in marking. Evidence of the confusion that exists concerning grading is shown by the remarks of several teachers. One stated that it would be possible to score "accurately" without information about the student, but she said that grading would depend on a great number of factors. She then gave a score of seventy-five in arithmetic and eighty-four in English.

Another teacher said that she scored arithmetic "all the same," but she did not fail some of the poorer pupils on report cards if they were working up to their ability. She indicated a different standard for English when she said that if the paper had been done by a high-ability student, the score would be eighty. If done by a low-ability student, the score would be ninety. She gave the student eighty-five, since she did not know the ability of the student. The remarks of this teacher substantiated the statistical findings of the study that showed no significant differences in the grades assigned to the arithmetic paper. She indicated that arithmetic was scored on a fixed point basis, but English was scored on a floating point basis. She stated that the student with high ability would receive a lower grade than the student with the lower ability. This agreed with the rationale for the study; but the information, such as low ability, which was designed to enhance the grade, did not significantly affect it.

Another teacher objected to grading on a numerical basis, but she said she could give a letter grade and designate where the letter falls numerically. The letter grade she assigned was a "B," which she said would represent numbers 80-89. It is interesting to note that she felt capable of figuring a grade to within a range of ten points, but she did not feel capable of assigning a definite mark to the paper.

One teacher said, "As you note, I'm a stickler for perfection. Of course, the problems are worked correctly. I am giving a grade of eighty; but if the child were working for me, he would get an 'F.' I do not permit so much writing." It is interesting to note that the paper did have errors in it which she did not find. For example, in arithmetic she failed to find the multiplication error. She did find numerous errors in the English composition and noted them thoroughly.

Another teacher found the organization of the arithmetic paper confusing and was able to locate only three problems of the four that made up the paper. She graded on the basis of the number of problems that she found. In all fairness to the teacher, the numbering on the paper was rather poor and could easily cause confusion. The error by the teacher did have the effect of lowering the student's grade from passing to failing. Another teacher found five problems and counted one wrong, which yielded a score of eighty.

II. ANALYSIS OF NOTATIONS ACCORDING TO SUBJECTS

English paper review. It is apparent from reviewing the papers returned that while teachers were not asked to substantiate the grade which they assigned to the papers, many did. Thirty-seven teachers out of the 225 participating put more than ten critical marks on the English papers. Almost one-fourth of the teachers were critical of the student's beginning the sentences with "He." About 10 per cent objected to the frequent ending of sentences with "to me." A like number were critical of the handwriting in the paper. Only four teachers put a number grade on the papers without some comment.

Some of the criticism was vague, which would also leave a student in doubt as to how to correct his errors. The teachers in these cases would circle letters, words, and punctuation, and draw dashes and lines in places throughout the composition.

Several teachers let loose a barrage of suggestions that possibly would have bewildered a fifth grade student:

(1) Indent the first line

(2) Suggest the paragraph in the first sentence

(3) See that every sentence keeps to the topic

(4) Do not join the sentences with and's

(5) Begin and end each sentence correctly

(6) Watch your use of capital letters

(7) Check punctuation marks

(8) Use different words at the beginning of each sentence

(9) Make the paragraph interesting

Another teacher wrote in the margins concerning the same paper:

(1) Write on the lines!

(2) Stay close to margin!

(3) Follow writing instruction!

(4) Sentences too short and choppy.

(5) Name your friend.

(6) Tell why you admire him.

(7) Rewrite.

Disagreement in selecting what was wrong and what was right about the English paper was much more evident than in the scoring of arithmetic papers, although certain repetitions and short sentences seemed to bother the majority of the teachers. The question of whether the student wrote about his topic was one in which there was disagreement. Some indicated that he had; others indicated that he had not. As one teacher put it, essay material is always difficult to grade. Although there were more differences in what should have been marked wrong, the range of scores was less than the range of thé arithmetic papers. The range for scores in English was from fifty to one hundred, and in arithmetic was from twenty-five to one hundred. The observed variance was significantly greater in arithmetic than in English in the category of no information.

Methods of scoring English papers. It made little difference as to whether the paper was a daily paper or a test paper; this is shown in the similarity of comments that teachers gave in scoring English papers. An examination of the test papers and daily papers showed no apparent difference in the methods used by the teachers regardless of the nature of the assignment. A number of methods were presented substantiating scores that were assigned. One teacher set up the following scale in grading the paper:

	Possible Score	This Paper
Idea	20 points	17
Sentence formation	20 points	10
Correct usage	20 points	18
Punctuation	10 points	5
Capitalization	10 points	8
Spelling	10 points	8
Penmanship	5 points	3
Form	5 points	4
Total	100 points	73

The teacher commented that this procedure takes much time; but she preferred to do it because a child has a chance to see his weaknesses.

The scale that another teacher presented was similar, although the two teachers lived in different geographical areas:

	Possible Score	This Paper
Expression of thought	25 points	15
Sentence structure	20 points	10
Correct usage of words	20 points	20
Punctuation	20 points	10
Handwriting	10 points	9
Effort	5 points	5
Total	100 points	69

With a score of seventy required for passing, the student would fail on this scale, but he would pass on the previous scale; notwithstanding the scores were nearly the same.

Another teacher based the grade on a loss of one point for each mistake. She found thirteen errors and scored the paper eighty-seven. Still another teacher gave as the basis for the score she assigned two points for each error. She gave a score of eighty-six because she found seven errors.

If written at the beginning of the year, the paper would receive a "C," according to one teacher. If written after a study of paragraphs had been made, a grade of "D" would result.

Arithmetic paper review. Of the 225 teachers who scored the arithmetic paper, only three failed to find the error in multiplication in the second problem. Many of the teachers indicated the actual error that was made and gave the basis for their grading. One teacher who corrected the problem also erred in her computation. She found the answer for problem one to be $240.25 instead of $239.25, which was the correct answer.

Several teachers who scored the arithmetic paper were concerned with its appearance. They also criticized the inconsistency of the use of dollar signs and cent symbols in the last problem. The main idea that they tried to project here was consistency of the signs. The student was also reprimanded for his writing of fractions. The denominator was written below the imprinted line on the paper. There was general agreement that this was not the proper way to write fractions.

Methods of scoring arithmetic papers. It was seen in looking through the papers that some teachers had a simple way of scoring an arithmetic paper. They counted the number of problems and determined the number to be four. They decided that, if there were four problems, and a score of 100 was considered to be perfect, each problem would count 25. If the student failed to get an acceptable answer to a problem, he was penalized 25 points. In the paper graded by the teachers, one of the problems had a computational error, so according to these teachers, the grade should have been 75. In fact, 107 teachers of 239 teachers who scored the arithmetic paper gave a score of 75. This is typical of the fixed point system. Other teachers who scored the paper from 25 to 100 were influenced in their decision by factors other than whether the problem was right or wrong.

One teacher based her scoring of the arithmetic paper on the scale shown below:

1st. problem---24 points
2nd. problem---13 points
(11 off for inaccurate working process).
3rd. problem---24 points
4th. problem---24 points
(4 off for incorrect numbering of problems and minor grammatical errors).

Total received: 85 points

Another teacher used about the same process in reasoning that the problems could be divided into two parts, i.e., the computation, and the labeling and explanation of the problems. She gave a score of 88 because of inaccurate computation of one problem.

Another method was devised that gave ten points for procedure, ten points for answer, and five points for statement of the problem. The paper was graded 80 by this method.

The teachers who gave a score of below 75 on the arithmetic papers seemed to have subtracted points for the statement of the problem or neatness. There was little evidence to suggest that the information the teacher had about the student influenced the mark assigned to the arithmetic paper.

III. ANALYSIS OF NOTATIONS IN THE LIGHT OF PRINCIPLES OF MOTIVATION AND GUIDANCE

Motivation in grading. Evidence that some teachers intend to grade for motivation was a statement by a teacher who gave the student 90 in arithmetic to encourage him. She evidently did not want to encourage him in English, for she gave a score of 70 on that paper. The possibility also exists that the grade of 70 to that teacher was an encouraging grade when she considered the quality of the English paper.

Use of grades to motivate was indicated by another teacher who wrote this note to the student: "Do over for better grade." The note to the mythical student emphasizes the identification of the teacher with her task of grading. Many teachers gave evidence of this identification by comments they made to the student.

Testing and evaluation enter into pupil motivation in a number of ways. The teacher appraisal, made known to the pupil through corrections on papers and through oral comments, provides a continuing set of cues on success or progress that motivates daily activities.[1]

Teachers who graded the papers gave a great number of comments obviously designed to motivate students. Some comments would best be classified as positive attempts, and other comments would be construed as negative techniques of motivation. Although many of the teachers were kind in their notes to the student who prepared the paper, only nine actually praised the student for the work that he had done. One of them said, "The idea was excellent!"

Several teachers gave concrete evidence to the student that he had been successful in gaining their approval. One teacher wrote: "Good computing! Keep up the good work!" She gave the student 90 in arithmetic.

---

[1]Robert L. Thorndike and Elizabeth Hagen, Measurement and Evaluation in Psychology and Education (New York: John Wiley and Sons, Incorporated, 1961), p. 459.

In another note not designed for the student, she wrote, "I may be an 'easy' grader in math. It would be so disheartening to take off 25 points for one inaccuracy, especially when the reasoning was good."

Another teacher wrote, "I think this is a good paper." Still another teacher indicated, "The last problem was very well-spaced and solved."

Another method that was used to help students improve was the writing of a series of questions in the margin which was designed to help the child write a "better" composition. Five of the teachers wrote notes to the student asking him to see them for help. One wrote, "I've made some corrections, however, see me to go over the entire paper to talk over correlation and longer sentences. Your penmanship is messy in spots."

Some of the teachers' messages to students were somewhat negative in nature, and they could possibly give rise to feelings of frustration or rejection. A teacher wrote, "You have too many careless mistakes." In this case, she pointed out one mistake in computation. Another wrote, "Your thoughts are all the same--makes your theme rather boring." A student who was identified as one who had consistently done below-average work was informed: "This is terrible! You need to spend more time thinking about what you want to say." These comments reflect to the pupil the teacher's evaluation of the pupil's performance. Clinical research has shown that students need to experience success.[2] Teachers probably need to set more flexible standards in order for students to feel that they have succeeded.

Guidance in marking. Some teachers were concerned with the teaching of attitudes and values as part of the marking of the English paper. They evidently felt that the paragraph gave some insight as to the student who prepared the paper.

One teacher stated, "You are interested only in what he can do for you, not in what he can contribute to himself and society." This remark called to the student's attention that he might be somewhat selfish in his attitude. This same concern was evidenced by a teacher who asked, "What do you do for him?"

Several teachers wondered if the friend was really the best friend for the student. One said, "The student's values are questionable. His best friend is admired for anti-social behavior." The student could have meant that, regardless of the trouble in which his friend finds himself, the student is still his friend.

One teacher attempted to arrive at a deeper understanding of the student who prepared the English

---

[2] Ibid., p. 487.

paper. She concluded that the student who prepared the paper was either socially or mentally retarded. This teacher, in the absence of information as to sex, assumed that the student was a girl. It is seen that the sex of the student who prepared the paper could affect the teacher's concept of the student, as far as the guidance function is concerned. As a matter of fact, the student who prepared the paper is a well-adjusted, "average," male student, according to his teacher and principal.

IV. ANALYSIS OF NOTATIONS AS TO SPELLING ERRORS

Teachers have sometimes been critized for their spelling and writing; but in the notes which they wrote to the researcher and to the student, errors were few. The word "to" was used for "too" in one case, and "titile" was used for "title." In one of the rare cases where praise was given, a teacher said, "Your spelling is good." Immediately following this comment was, "Check rules for comas."

An analysis of teachers' notations is summarized below. The following chapter summarizes the study and its results along with the conclusions that are supported by the findings of the investigation. The implications of the conclusions are also presented.

V. SUMMARY

In summary, the teachers' notations tended to indicate general agreement as to what was wrong with the arithmetic paper; but as far as assigning a grade to the paper, they had great differences of opinion. The scores ranged from 25 to 100.

Generally speaking, they used two methods of scoring the arithmetic paper. One method was scoring the problem all wrong if any error was made. The other method was giving partial credit if the process was correct and the computation wrong.

There was little evidence in the notations of the teachers to suggest that the information given about the student influenced the mark assigned to the arithmetic paper. This is in accord with the statistical findings of the study which found no differences in means because of information.

The English paper seemed to be more difficult to score in that more alternatives were permitted by the scorers than in arithmetic. Teachers attempted to score on the basis of content, sentence structure, spelling, punctuation, capitalization, creativity, and other bases. Some teachers indicated that they used information factors about the student in order to determine a grade. This agrees with the findings in Chapter IV concerning the influence of information factors on grades. The mean of the scores given by teachers who were given information designed to depress the scores was significantly different. However, the information designed to depress the scores actually elevated them. The notations of teachers reflected that the information designed to decrease scores caused them to lower scores. This did not actually happen.

That teachers have some feelings of insecurity and confusion in marking is reflected in comments written on the papers that they were asked to grade.

It seems evident that teachers grade for a variety of purposes, among which are to motivate and to inform students of their progress. That teachers are alert to opportunities of teaching attitudes and values is shown quite clearly in the notations which they addressed to the student whose paper they marked.

CHAPTER VI

SUMMARY, CONCLUSIONS, RECOMMENDATIONS, AND IMPLICATIONS

One of the activities of teachers that affects the lives of students in a significant manner is the assigning of marks. Marks are given for a variety of purposes, among which are information and motivation. The basic determinants of marks have not been satisfactorily explained.

I. SUMMARY

The literature set forth the notion that non-intellectual factors of information about the student, such as age-grade, sex, teacher's estimate of ability, past achievement, and behavior of the student affected the marks assigned by the teacher. It was also theorized that the nature of the assignment, i.e., whether the paper was a daily or test paper, affected marks.

If teachers used a fixed point method of grading, marks should not be affected by the information as to the student or to the nature of the assignment. If the floating point method were used, marks should be affected.

The purposes of the study were these:

1. To determine if the information that a teacher has about a student affects the marks assigned by the teacher, either by elevating or depressing the score.

2. To determine if the information that a teacher has about a student reduced observed variance in assigned marks.

3. To determine whether teachers grade on a fixed point system or floating point basis.

4. To determine if variance in marking is greater in the more loosely structured subject areas such as English than in the more rigidly structured subject areas such as arithmetic.

5. To determine if the nature of the assignment affects the marks assigned by teachers.

A sample of 225 female, fifth grade teachers representing all of the geographical areas of the continental United States was selected. The same fifth grade arithmetic and fifth grade English paper were submitted to the teachers to score. One-third of the teachers were given no information; one-third were given information designed to depress the scores of the papers; and one-third were given information that was designed to elevate the grades assigned to the paper. A random half of those teachers who received information were also given the information that the paper was a test paper. The other half of the teachers who received information were informed that the paper was a daily paper.

The resulting scores were explored by a multiple classification analysis of variance with the .05 level of confidence set as the fiducial limits. The variances were compared by ratio of variances to determine if reduction in observed variance was a function of the information transmitted to the teacher.

II. CONCLUSIONS

Scoring of arithmetic papers. Only chance differences existed in the scoring of arithmetic papers whether teachers had information about the student or not. In scoring arithmetic papers the observed variance was reduced when either positive or negative information about the student was introduced. Although the means were not affected, the spread of scores was reduced, giving the grades more stability and utility.

Scoring of English papers. The information about the student did affect the average scores assigned to the English papers. The information that was supposed to depress the mean scores raised the scores significantly. Instead of the teachers expecting more of the student who was supposedly capable of doing better, the teachers followed the suggestion of the information. If the student had done well in the past, he was graded higher in English. In subsequent analysis to determine if a particular type of information caused significantly different mean grades, it was found that it was improbable that any one type of information had enough potency by itself to be significant.

In scoring English papers the observed variance was reduced when negative information about the student was given. In subsequent analysis, the variance by types of information was compared to no information, and the information concerning the behavior of the student was tested as significant.

Fixed point versus floating point. It was concluded that teachers used a fixed point system of grading more than a floating point system. This was especially true in the grading of arithmetic papers. In English, the teachers probably used more of a floating point system, but the floating of the point was apparently not geared to expectations concerning the student. According to the rationale, if the expectations for a student were great and he turned in an "average" paper, the score would be reduced. This study indicated that the reputation of the student may have caused the point to float rather than the expectation for the student. If students' reputations were good, their mean score increased.

Structure of the subject. The structure of the subject made little difference in the marking of the papers. English, the subject with more permissible alternatives in marking, was affected by information about the student enough to cause the mean grade in one category to be elevated. The arithmetic means were not affected by information. In comparing variances by structure of subject, the study showed that variance was greater in marking in arithmetic than in English, when no information was given about the student whose paper was being graded.

Nature of the assignment. The nature of the assignment as to daily or test papers did not affect the marks in such a way as to elevate or depress the grades. The variance was reduced in scoring arithmetic papers, when the information that the paper was a daily paper was given. In English, the observed variance was not affected by the introduction of the information that a paper was a daily paper or a test paper.

General conclusion. The study was exploratory in nature and did not find the major determinant or determinants of marks.

III. RECOMMENDATIONS

The following recommendations are made relative to further exploration in the field of marking.

1. The study should be replicated using a more heterogeneous sample in order to determine if the restrictions applied to the sample in this study, such as experience and ability, influenced the results unduly.

2. Further studies should be made in attempts to analyze the evidence used in arriving at grades.

3. The study should be replicated with more information supplied to the teacher about the student

such as the information a teacher has available from school records.

4. Further attempts should be initiated to identify some determinants of marks other than achievement.

5. Teachers should study rationale of marking in a course designed for that purpose and with standardization so that symbols carry about the same informational content.

IV. IMPLICATIONS

Since teachers apparently used a fixed point system in grading, standards for grading papers can be set up and adhered to by teachers. This could lead to standardization of marking procedures so that the marks given within a school district and among school districts would convey the same information.

Another implication is that teachers should be informed that information designed to increase expectations for students in reality may cause them to follow the suggestion of the information, especially in loosely structured subjects. They could then be aware of this and realize that, although reliability is increased, validity is not necessarily increased.

Where reliability is a goal, teachers should have ready access to cumulative records so that they can know as much as possible about students. In grading more rigidly structured subjects, the information apparently reduced the spread of scores and increased reliability.

BIBLIOGRAPHY

SELECTED REFERENCES

A. BOOKS

Ahmann, J. Stanley, and Marvin D. Glock. Evaluating Pupil Growth. Boston: Allyn and Bacon, Inc., 1963.

Ahmann, J. Stanley, and Helen L. Wardeberg. Evaluating Elementary School Pupils. Boston: Allyn and Bacon, Inc., 1960.

Bloom, Benjamin S., and Frank R. Peters. The Use of Academic Prediction Scales for Counseling and Selecting College Entrants. New York: The Free Press of Glencoe, Inc., 1961.

Davis, Frederick B. Educational Measurements and Their Interpretation. Belmont, California: Wadsworth Publishing Company, 1964.

Greene, Harry A. Measurement and Evaluation in the Elementary School. New York: Longmans, Green, and Company, 1953.

Gronlund, Norman B. Measurement and Evaluation in Teaching. New York: The Macmillan Company, 1965.

Guilford, J. P. Fundamental Statistics in Psychology and Education. New York: McGraw-Hill Book Company, Inc., 1956.

Kelly, F. J. Teacher Marks, Their Variability and Standardization. New York: Teachers College, 1914.

Kerlinger, Fred N. Foundations of Behavioral Research. New York: Holt, Rinehart, and Winston, Inc., 1964.

Ross, C. C. Measurement in Today's Schools. New York: Prentice Hall, Inc., 1942.

________. Measurement in Today's Schools. New York: Prentice Hall, Inc., 1947.

Ruch, G. M. Improvement of the Written Examination. Chicago: Scott, Foresman Company, 1924.

Ruch, G. M., and George D. Stoddard. Tests and Measurements in High School Instruction. Chicago: World Book Company, 1927.

Starch, Daniel E. _Educational Measurements_. New York: The Macmillan Company, 1916.

Symonds, P. M. _Diagnosing Personality and Conduct_. New York: The Macmillan Company, 1931.

________. _Measurements in Secondary Education_. New York: The Macmillan Company, 1934.

Thorndike, Robert L., and Elizabeth Hagen. _Measurement and Evaluation in Psychology and Education_. New York: John Wiley and Sons, Incorporated, 1961.

Traxler, A. E. _Techniques of Guidance_. New York: Harper and Row, Publishers, 1957.

Van Dalen, Deobold B., and William L. Meyer. _Understanding Educational Research_. New York: McGraw-Hill Book Company, Inc., 1962.

Webb, L. W., and A. M. Shotwell. _Testing in the Elementary School_. New York: Farrar and Rinehart, Incorporated, 1939.

Wrinkle, W. L. _Improving Marking and Reporting Practices in Elementary and Secondary Schools_. New York: Holt, Rinehart, and Winston, 1947.

B. PUBLICATIONS OF THE GOVERNMENT, LEARNED SOCIETIES AND OTHER ORGANIZATIONS

Heck, Arch O. _Thirty-seventh Yearbook of the National Society for the Study of Education, Part II_. Bloomington, Illinois: Public School Publishing Company, 1938, pp. 187-189.

National Education Association Research Memo. Volume XXI. July, 1964.

C. PERIODICALS

Bills, Robert E. "The Effects of a Value on Learning," _Journal of Personality_, XXI (December, 1952), pp. 217-222.

Carter, R. S. "How Invalid are Marks Assigned by Teachers?" Journal of Educational Psychology, XLIII (May, 1952), pp. 218-228.

Chansky, N. M. "X-Ray of the School Mark," Educational Forum, XXVI (March, 1962), pp. 347-352.

Crooks, A. D. "Marks and Marking Systems," Journal of Educational Research, XXVII (March, 1933), pp. 259-272.

Day, L. C. "Boys and Girls and Honor Ranks," School Review, XLVI (April, 1938), pp. 288-289.

Douglas, H. R., and N. E. Olson. "Relation of High School Marks to Sex in Four Minnesota Senior High Schools," School Review, XLV (October, 1937), pp. 282-288.

Farwell, G. H. "Pressures Behind the Grade," Clearinghouse, XXXIX (April, 1964), pp. 462-466.

Fensch, E. A. "Good Ol' A, B, C, D, and F," Clearinghouse, XXXVII (October, 1962), pp. 93-95.

Franzen, R. "The Accomplishment Quotient," Teacher College Record, XXI (November, 1920), pp. 432-440.

Garner, C. E. "Survey of Teachers' Marks," School and Community, XXI (October, 1935), pp. 116-117.

Hadley, S. T. "School Marks--Fact or Fancy?" Educational Administration and Supervision, XL (May, 1954), pp. 305-312.

Herron, John S. "How Teachers Rate Their Pupils," Department of Elementary School Principals Bulletin, VIII (October, 1929), pp. 235-239.

Hurlock, E. B. "The Value of Praise and Reproof as Incentives for Children," Archives of Psychology, LXXI (July, 1924), p. 462.

Johnson, Mauritz, Jr. "Solving the Mess in Marks," New York State Education Department, XLIX (February, 1962), pp. 12-14.

Jones, John A. "Grading, Marking, and Reporting in the Modern Elementary School," Educational Forum, XIX (November, 1954), pp. 45-54.

Liggett, W. A. "Are There Better Ways of Evaluating, Recording, and Reporting Pupil Progress in the Junior and Senior High Schools?" National Association of Secondary School Principals Bulletin, XXXIV (March, 1950), pp. 79-89.

Lobaugh, Dean. "Girls Grades and I.Q.'s," Nations Schools XXX (January, 1942), p. 42.

Mandel, B. "Appraising Ability and Achievement," High Points, XLIV (April, 1962), pp. 31-37.

Morlan, G. "He Gave de Maupassant a Grade of B," School and Society, XXXIV (September, 1931), pp. 321-322.

Muessig, R. H. "The Mystery of Grading and Reporting," Education, LXXXIII (October, 1962), pp. 90-93.

Norsted, R. A. "To Mark or Not to Mark?" Journal of Education, CXXI (February, 1938), pp. 81-84.

Rugg, H. O. "Teachers' Marks and Marking Systems," Educational Administration and Supervision, I (March, 1915), pp. 117-142.

Rusch, Reuben R., John A. Brown, and Arthur R. Delong. "Meaning of an Arithmetic Score," The Arithmetic Teacher (March, 1962), p. 145.

Russell, Ivan L. "Personality--Does It Influence Teacher Marks?" Journal of Educational Research, XLVIII (August, 1955), pp. 561-641.

Sharp, L. A. "The Value of Standards in Grading Examination Papers," Peabody Journal of Education, III (January, 1925), pp. 38-45.

Shinnerer, M. C. "Failure Ratio: Two Boys to One Girl," Clearinghouse, XVIII (February, 1944), pp. 264-270.

Starch, Daniel E., and E. C. Elliot. "Reliability of Grading Work in English," School Review, XX (September, 1912), pp. 442-457.

________. "Reliability of Grading Work in History," School Review, XXI (December, 1913), pp. 676-681.

________. "Reliability of Grading Work in Mathematics," School Review, XXI (April, 1913), pp. 254-259.

Swenson, Clifford. "Packing the Honor Society," Clearinghouse, XVI (Mar, 1942), pp. 521-524.

Travers, Robert M., and Norman E. Gronlund. "Meaning of Marks," Journal of Higher Education, XXI (May, 1950), pp. 369-374.

Volberding, Eleanor. "Characteristics of Successful and Unsuccessful Eleven-Year-Old Pupils," Elementary School Journal, XLIX (February, 1949), p. 410.

Vredevoe, L. E. "How May We Make the Recording and Reporting of Pupil Progress More Meaningful?" National Association of Secondary School Principals Bulletin, XXXVII (February, 1953), pp. 179-182.

Wood, B. D. "Measurement of College Work," Educational Administration and Supervision, VII (September, 1921), pp. 301-304.

Yauch, W. A. "School Marks and Their Reporting," National Education Association Journal, L (May, 1961), p. 50.

OTHER REFERENCES

A. BOOKS

Bradfield, J. M., and Stewart H. Moredock. Measurement and Evaluation in Education. New York: The Macmillan Company, 1957.

B. PERIODICALS

Allen, C. H. "A Statistical Study of Teachers' Marks," Peabody Journal of Education, IX (January, 1932), pp. 310-312.

Archer, N. S. "Weighted Question: Is Assigning Values a Useful Procedure?" Clearinghouse, XXXVII (February, 1963), pp. 360-362.

Banker, Howard J. "The Significance of Teacher Marks," Journal of Educational Research, XVI (March, 1927), pp. 271-284.

Barnes, K. F., and E. H. Barnes. "Realistic Approach to Grading," Clearinghouse, XXXVI (April, 1962), pp. 476-478.

Bolmeir, Edward C. "Principles Pertaining to Marking and Reporting Pupil Progress," School Review, LIX (January, 1961), pp. 15-24.

Bolton, F. E. "Do Teachers' Marks Vary as Much as Supposed?" Education, XLVIII (January, 1927), pp. 23-29.

Caudle, J. B. "Educators Look at Grading," Texas Outlook, LV (July, 1961), pp. 14-15.

Cummings, S. N. "How to Earn a Zero," Clearinghouse, XXXVIII (October, 1963), p. 110.

Cummins, R. E. "Evaluating and Grading," Education, LXXXII (March, 1962), pp. 403-405.

DePue, P. "Rubber Yardstick, and the Pursuit of Excellence," Clearinghouse, XXXVIII (December, 1963), pp. 248-250.

De Zouch, D. "The Wounds are Mortal," Clearinghouse, XIX (June, 1944), pp. 339-344.

Doak, E. D. "Grading--A Deterrent to Learning," Clearinghouse, XXXVII (December, 1962), pp. 245-248.

Garverick, C. M., and H. D. Carter. "Instructor Ratings and Expected Grades," California Journal of Educational Research, XIII (November, 1962), pp. 218-221.

Gould, Geroge. "Practices in Marking Examinations," School Review, XL (February, 1932), pp. 142-146.

Hoggard, J. K. "Grading the Accelerated Student," Education, LXXXXIII (January, 1963), pp. 286-288.

Jarrett, C. D. "Marking and Reporting Practices in the American Secondary School," Peabody Journal of Education, XLI (July, 1963), pp. 36-48.

Krasteff, K. "Grading," School and Community, LI (October, 1962), p. 29.

Leese, J. "Mess in Marks," New York State Education, XLVIII (May, 1961), pp. 9-11.

Lund, John. "More Truth About Marks," Journal of Education, CIX (November, 1929), pp. 609-619.

McCabe, G. E. "Evaluative Atmosphere: Evil or Asset," Journal of Teacher Education, XII (September, 1961), pp. 262-270.

McKeachie, Wilbert J. "The Improvement of Instruction," Review of Educational Research, XXX (October, 1960), pp. 351-360.

Moskowitz, S. D. "Guide Lines for Marking Procedures," High Points, XLIV (April, 1962), pp. 26-30.

Muessig, R. H. "How Do I Grade Thee? Let Me Count the Ways," Clearinghouse, XXXVI (March, 1962), pp. 414-416.

Phillips, M. F. "Two Practical Questions Involving Grading Papers," Clearinghouse, XXXVI (December, 1961), pp. 237-239.

Quinn, G. R., and C. A. Szuberla. "Relative Grading," Clearinghouse, XXXVII (April, 1963), pp. 490-494.

Stockard, J. "Case for Dual Grading," Texas Outlook, XLVI (May, 1962), p. 30.

Stroup, F. "Grade-Point Average is Obsolete," Journal of Higher Education, XXXIV (January, 1963), pp. 10-15.

Teaf, H. M. "Grades, Their Dominion Is Challenged," Journal of Higher Education, XXXV (February, 1964), pp. 87-88.

Walker, W. "Make Those Marks More Reliable," School and Community, L (January, 1964), pp. 12-13.

APPENDIX A

APPENDIX A

Included in this Appendix are the form letters and sample papers distributed to teachers who participated in the study.

The materials and instructions sent to Category A teachers are found on pages 131 to 134.

The materials and instructions sent to Category B teachers are found on pages 135 to 139.

The materials and instructions sent to Category C teachers are found on pages 140 to 144.

MATERIALS AND INSTRUCTIONS SENT TO TEACHERS IN CATEGORY A (NO INFORMATION)

Dear Teacher:

Your school administration has agreed to cooperate with me in a research project which has as its goal the finding of some determinants of teacher marks. You were selected by your district because of your high rating by supervisors and your professional attitude.

This project is only one in a series of steps that will be necessary to bring order from the confusion that currently exists in marking throughout the nation.

We are asking you to take approximately twenty minutes of your time and score the enclosed papers on a scale of 0-100 points. You should consider 70 or above passing and below 70 failing. Please do not discuss this paper with any other teachers.

Please send me the scored papers in the enclosed self-addressed envelope at your earliest convenience.

Thank you very much.

Sincerely yours,

Glenn L. Morris

Problems

1. Bob's mother bought a new washing machine. She paid for it in eight payments of $24.50 each. How much did the washing machine cost?
8 payments for the washing machine
$24.50 each payment How much did
$24.50 the washing machine cost?

$$\begin{array}{r} \$24.50 \\ \times\ 8 \\ \hline \$196.00 \end{array}$$

The washing machine cost $196.00

2. On Monday Uncle Jim sold 87 bu. of apples at $2.75 a bushel. How much did he get for them?
$238.85 is he got for them.

$$\begin{array}{r} \$2.75 \\ \times\ 87 \\ \hline 1885 \\ 2200 \\ \hline \$238.85 \end{array}$$

1. The weather bureau reported that the snowfall on Monday $3\frac{1}{2}$ inches, on Tuesday 4 inches, and on Wednesday $2\frac{1}{2}$ inches.
What was the average snowfall per day?

Problems

$3\frac{1}{2}$ inches on Monday
4 inches on Tuesday
$2\frac{1}{2}$ inches on Wednesday
What was the average snowfall per day?

$$\begin{array}{r} 3\frac{1}{2} \\ 4 \\ +\ 2\frac{1}{2} \\ \hline 11\frac{2}{2} = 12 \end{array} \qquad \begin{array}{r} 4 \\ 3\overline{)12} \\ \underline{12} \\ 0 \end{array}$$

4 inches was the average snowfall per day.

1. Susan is buying a large chocolate Easter egg for 45¢ and a small one for 20¢. Find her change from a dollar bill?

45¢ a large Easter egg
20¢ a small Easter egg

$$\begin{array}{r} 45 \\ +20 \\ \hline 65¢ \end{array} \qquad \begin{array}{r} \$1.00 \\ -65 \\ \hline 35¢ \end{array} \qquad \begin{array}{r} 65 \\ +35 \\ \hline \$1.00 \end{array}$$

Her change from a dollar bill is 35¢.

My Best Friend

My Best friend is nice to me. He does things for me. He gives things to me. He is not very pretty but it does not matter to me. He never combs his hair. ~~he~~ He always leans back in his chair anyway he is very nice to me and it does not matter if he gets in trouble.

MATERIALS AND INSTRUCTIONS SENT TO TEACHERS IN CATEGORY B (INFORMATION TO DEPRESS GRADES)

Dear Teacher:

Your school administration has agreed to cooperate with me in a research project which has as its goal the finding of some determinants of teacher marks. You were selected by your district because of your high rating by supervisors and your professional attitude.

This project is only one in a series of steps that will be necessary to bring order from the confusion that currently exists in marking throughout the nation.

We are asking you to take approximately twenty minutes of your time and score the enclosed papers on a scale of 0-100 points. You should consider 70 or above passing and below 70 failing. Please do not discuss this paper with any other teachers.

Please send me the scored papers in the enclosed self-addressed envelope at your earliest convenience.

Thank you very much.

Sincerely yours,

Glenn L. Morris

TEACHERS IN CATEGORY B RECEIVED ONE OF THE FOLLOWING ITEMS OF INFORMATION

The student who prepared this paper was in the fifth grade and was of the normal age for his grade.

The student who prepared this paper was a boy.

The student who prepared this paper has consistently done above-average work.

The student who prepared this paper has consistently scored above 112 on I.Q. tests.

The student who prepared these papers has generally conformed to the rules of the school and has been courteous to teachers.

A random half of the teachers were told the following:

> The attached papers are daily papers in arithmetic and English.

An equal number of teachers were told the following:

> The attached papers are test papers in arithmetic and English.

Problems

1. Bob's mother bought a new washing machine. She paid for it in eight payments of $24.50 each. How much did the washing machine cost?

8 payments for the washing machine
$24.50 each payment How much did
$24.50 the washing machine cost?
x 8
$196.00

The washing machine cost $196.00

2. On Monday Uncle Jim sold 87 bu. of apples at $2.75 a bushel. How much did he get for them?

$2.75 $238.85 is the got for them.
x 87
1885
2200
$238.85

1. The weather bureau reported that the snowfall on Monday $3\frac{1}{2}$ inches, on Tuesday 4 inches, and on Wednesday $2\frac{1}{2}$ inches.
What was the average snowfall per day?

Problems

$5\frac{1}{2}$ inches on Monday
4 inches on Tuesday
$2\frac{1}{2}$ inches on Wednesday
What was the average snowfall per day?

$$\begin{array}{r} 5\frac{1}{2} \\ 4 \\ 2\frac{1}{2} \\ + \quad \\ \hline 11\frac{2}{2} = 12 \end{array} \qquad \begin{array}{r} 4 \\ 3\overline{)12} \\ \underline{12} \\ 0 \end{array}$$

4 inches was the average snowfall per day.

1. Susan is buying a large chocolate Easter egg for 45¢ and a small one for 20¢. Find her change from a dollar bill?

45¢ a large Easter egg
20¢ a small Easter egg

$$\begin{array}{r} 45 \\ +20 \\ \hline 65¢ \end{array} \qquad \begin{array}{r} \$1.00 \\ -65 \\ \hline 35¢ \end{array} \qquad \begin{array}{r} 65 \\ +35 \\ \hline \$1.00 \end{array}$$

Her change from a dollar bill is 35¢.

My Best Friend

My Best friend is nice to me, He does things for me. He gives things to me, He is not very pretty but it does not matter, to me. He never combs his hair. ~~he~~ He always leans back in his chair anyway he is very nice to me and it does not matter if he gets in trouble.

MATERIALS AND INSTRUCTIONS SENT TO TEACHERS IN CATEGORY C (INFORMATION TO ELEVATE GRADES)

Dear Teacher:

Your school administration has agreed to cooperate with me in a research project which has as its goal the finding of some determinants of teacher marks. You were selected by your district because of your high rating by supervisors and your professional attitude.

This project is only one in a series of steps that will be necessary to bring order from the confusion that currently exists in marking throughout the nation.

We are asking you to take approximately twenty minutes of your time and score the enclosed papers on a scale of 0-100 points. You should consider 70 or above passing and below 70 failing. Please do not discuss this paper with any other teachers.

Please send me the scored papers in the enclosed self-addressed envelope at your earliest convenience.

Thank you very much.

Sincerely yours,

Glenn L. Morris

TEACHERS IN CATEGORY C RECEIVED ONE OF THE FOLLOWING ITEMS OF INFORMATION ALONG WITH THE PAPERS WHICH FOLLOW

The student who prepared this paper was in the fifth grade but was older than the ordinary fifth grader.

The student who prepared this paper was a girl.

The student who prepared this paper has consistently scored below ninety-one on I.Q. tests.

The student who prepared this paper has consistently done below-average work.

The student who prepared these papers has frequently violated the rules of the school and has been discourteous to teachers on numerous occasions.

A random half of the teachers were told the following:

The attached papers are daily papers in arithmetic and English.

An equal number were told the following:

The attached papers are test papers in arithmetic and English.

Problems

1. Bob's mother bought a new washing machine. She paid for it in eight payments of \$24.50 each. How much did the washing machine cost?
8 payments for the washing machine
\$24.50 each payment How much did
\$24.50 the washing machine cost?
$\underline{\times 8}$
\$196.00
The washing machine cost \$196.00
2. On Monday Uncle Jim sold 87 bu. of apples at \$2.75 a bushel. How much did he get for them?
\$2.75 \$238.85 is he got for them.
$\underline{\times 87}$
1885
$\underline{2200}$
\$238.85

1. The weather bureau reported that the snowfall on Monday $5\frac{1}{2}$ inches, on Tuesday 4 inches, and on Wednesday $2\frac{1}{2}$ inches.
What was the average snowfall per day?

Problems

$3\frac{1}{2}$ inches on Monday
4 inches on Tuesday
$2\frac{1}{2}$ inches on Wednesday
What was the average snowfall per day?

$$\begin{array}{r} 3\frac{1}{2} \\ 4 \\ + 2\frac{1}{2} \\ \hline 11\frac{2}{2} = 12 \end{array} \qquad 3\overline{)12}^{\,4} \quad \begin{array}{r} 12 \\ \hline 0 \end{array}$$

4 inches was the average snowfall per day.

1. Susan is buying a large chocolate Easter egg for 45¢ and a small one for 20¢. Find her change from a dollar bill?

45¢ a large Easter egg
20¢ a small Easter egg

$$\begin{array}{r} 45 \\ +20 \\ \hline 65¢ \end{array} \qquad \begin{array}{r} \$1.00 \\ -65 \\ \hline 35¢ \end{array} \qquad \begin{array}{r} 65 \\ +35 \\ \hline \$1.00 \end{array}$$

Her change from a dollar bill is 35¢.

My Best Friend

My Best friend is nice to me. He does things for me. He gives things to me. He is not very pretty but it does not matter to me. He never combs his hair. ~~he~~ He always leans back in his chair anyway he is very nice to me and it does not matter if he gets in trouble.

APPENDIX B

APPENDIX B

CITIES FROM WHICH SAMPLE WAS DRAWN

Akron, Ohio

Baton Rouge, Louisiana

Beaumont, Texas

Charlotte, North Carolina

Erie, Pennsylvania

Evansville, Indiana

Greensboro, North Carolina

Houston, Texas (metropolitan area)

Jackson, Mississippi

Kansas City, Kansas

Little Rock, Arkansas

Louisville, Kentucky

New Bedford, Massachusetts

Omaha, Nebraska

Richmond, Virginia

Rockford, Illinois

Salt Lake City, Utah

Savannah, Georgia

Tucson, Arizona

Worcester, Massachusetts

APPENDIX C

APPENDIX C

TABLE XIII

ANALYSIS OF VARIANCE FOR ARITHMETIC BY CATEGORIES OF INFORMATION

SOURCE	df	SUMS OF SQUARES	MEAN SQUARES	F
Groups	2	188.59	94.30	.92***
Error	222	22865.93	103.00	
Total	224	23054.52		

TABLE XIV

ANALYSIS OF VARIANCE FOR ENGLISH BY CATEGORIES OF INFORMATION

SOURCE	df	SUMS OF SQUARES	MEAN SQUARES	F
Groups	2	852.82	426.41	4.98**
Error	222	18,997.18	85.57	
Total	224	19,850.00		

*Denotes significance at .05 level of confidence.
**Denotes significance at .01 level of confidence.
***Denotes no significant difference.

TABLE XV

ANALYSIS OF VARIANCE FOR ARITHMETIC
TEST AND DAILY

SOURCE	df	SUMS OF SQUARES	MEAN SQUARE	F
Groups	2	397.62	198.81	1.55***
Error	222	28428.14	128.06	
Total	224	28825.76		

TABLE XVI

ANALYSIS OF VARIANCE FOR ENGLISH
TEST AND DAILY

SOURCE	df	SUMS OF SQUARES	MEAN SQUARE	F
Groups	2	514.32	257.16	2.85***
Error	230	20,746.68	90.20	NSD
Total	232	21,261.00		

***Denotes no significant difference.

TABLE XVII

ANALYSIS OF VARIANCE FOR ENGLISH
TYPES OF INFORMATION

SOURCE	df	SUMS OF SQUARES	MEAN SQUARE	F
Groups	8	1174.50	146.81	1.45***
Error	69	7001.90	101.48	
Total	77	8176.40		

***Denotes no significant difference.

APPENDIX D

APPENDIX D

FREQUENCY DISTRIBUTION OF ENGLISH SCORES ASSIGNED BY TEACHERS ACCORDING TO INFORMATION CATEGORIES

CLASS INTERVAL	FREQUENCY		
	CATEGORY A	CATEGORY B	CATEGORY C
96-100	0	2	0
91-95	0	0	0
86-90	3	4	2
81-85	8	12	3
76-80	13	15	9
71-75	18	34	14
66-70	14	9	20
61-65	5	2	15
56-60	7	4	4
51-55	0	1	4
46-50	7	2	6

FREQUENCY DISTRIBUTION OF ARITHMETIC SCORES ASSIGNED BY TEACHERS ACCORDING TO INFORMATION CATEGORIES

CLASS INTERVALS	FREQUENCY		
	CATEGORY A	CATEGORY B	CATEGORY C
96-100	2	0	0
91-95	1	2	1
86-90	4	4	5
81-85	6	10	8
76-80	9	10	10
71-75	33	38	38
66-70	14	10	7
61-65	1	4	4
56-60	0	1	0
51-55	1	0	1
46-50	4	5	2
41-45	0	0	0
36-40	0	0	0
31-35	0	0	0
26-30	0	0	0
21-25	2	0	0