STATEMENT OF PERMISSION TO COPY

In presenting this professional paper in partial fulfillment of the requirements for an advanced degree at Montana State University, I agree that the Library shall make it freely available for inspection. I further agree that permission for extensive copying of this paper for scholarly purposes may be granted by my major professor, or, in his absence, by the Director of Libraries. It is understood that any copying or publication of this paper for financial gain shall not be allowed without my written permission.

Signature Sarah Ann Cashmore Herald

Date May 14, 1970

EXPERIMENTS IN TEACHERS' EVALUATION OF STUDENTS' WORK

by

Sarah Ann Cashmore Herold

A professional paper submitted to the Graduate Faculty
in partial fulfillment of the requirements
for the degree

of

MASTER OF EDUCATION

in

Elementary Education

Approved:

Willis Vandmer
Head, Major Department

Gerald D. Sullivan
Chairman, Examining Committee

K. Goering
Graduate Dean

MONTANA STATE UNIVERSITY
Bozeman, Montana

June, 1970

ACKNOWLEDGEMENTS

The completion of this study depended upon the assistance and cooperation of many persons. The writer wishes to thank the teachers for the time and evaluations they gave. The cooperation and enthusiasm for the purposes of this study by the superintendents and principals who allowed the writer to move freely through their buildings was greatly appreciated.

The writer wishes to thank the two people without whose help this study would never have reached completion. Dr. Theodore Hoff, Professor of Economics and Statistics, Northern Arizona University, Flagstaff, Arizona, helped to set up the experiments so they would be valid statistically and evaluated the data for presentation in the study. Gerald Sullivan, Professor of Education, Montana State University, assisted and advised the writer repeatedly and graciously during the many months of compiling and writing.

S. H.

TABLE OF CONTENTS

Chapter		Page
1.	INTRODUCTION	1
	STATEMENT OF THE PROBLEM	2
	NEED OF THE STUDY	2
	NULL HYPOTHESIS	3
	GENERAL PROCEDURE	3
	LIMITATIONS	4
	DEFINITION OF TERMS	5
2.	REVIEW OF RELATED LITERATURE	6
	INFLUENCE OF EXTERNAL FACTORS	6
	Physical Appearance	7
	I. Q. Score	9
	Socio-economic Background	11
	Sex Factor	12
3.	STATISTICAL DATA AND INTERPRETATIONS	15
	POPULATION DESCRIPTION	15
	HYPOTHESES	16
	Major Hypothesis	16
	Sub-hypothesis 1.	16
	Sub-hypothesis 2.	16
	Sub-hypothesis 3.	17
	Sub-hypothesis 4.	17
	COLLECTION AND ORGANIZATION OF DATA	17
	STATISTICAL MODEL AND ANALYSIS OF MAJOR HYPOTHESIS	20
	TESTING THE SUB-HYPOTHESES	22
4.	SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS	25
	SUMMARY	25
	CONCLUSIONS	28
	RECOMMENDATIONS	31
	APPENDIX	34
	INTRODUCTION TO STUDY AND INSTRUCTIONS	35
	WORK ONLY	35
	PHYSICAL APPEARANCE	36
	I. Q. SCORE	36
	SOCIO-ECONOMIC BACKGROUND	36
	SEX FACTOR	37
	STATISTICAL INFORMATION	40
	REFERENCES	44

LIST OF TABLES

Table Page

Table 1. Analysis of Variance----------21

Table 2. Least Significant Difference Table----------22

Table 3. Analysis of Variance for Sub-hypothesis 4.----------24

APPENDIX

Table 4. Average of Grades of Ten Teachers Given on Basis of Work Only and Work Plus Other Knowledge----------41

Tabulation of Data for Table 1----------42

Table 5. Average of Grades Assigned by Five Male and Five Female Teachers When Sex of Student Was Known----------43

LIST OF PLATES

Plate Page

APPENDIX

Plate 1. Work Plus Physical Appearance----------38

Plate 2. Work Plus I. Q. Score----------39

ABSTRACT

Problem

The major null hypothesis stated, "Teachers' evaluations of students' work is not influenced by factors other than the work itself. Therefore, grades assigned by teachers on the basis of work alone will not differ significantly from grades assigned when other factors are known by the teacher." Acceptance on a statistical basis, of the major hypothesis must lead to acceptance of all four sub-hypotheses.

Population Description and Procedures

Data was gathered by personal interview from seventy teachers who had been actively teaching academic subjects within the last five years in the intermediate grades of Montana elementary schools.

The empirical data consisted of the teacher's evaluation, on an ascending scale of 1 to 10, of the work of each of twenty-five fifth-grade public school students.

Analysis of Data

Analysis of the data showed that factors other than work were significant in the grading of a student's work. Therefore, the major hypothesis was rejected. This led to the testing of each of four null sub-hypotheses.

Physical appearance factor. Analysis of the data when work alone was compared to work plus knowledge of the physical appearance of the student, indicated that physical appearance was highly significant in evaluating a student's work. Therefore, Sub-hypothesis 1. was rejected.

I. Q. score. Analysis of the data when work alone was compared to work plus the knowledge of the student's I. Q. score, indicated that an I. Q. score was highly significant in evaluating a student's work. Therefore, null Sub-hypothesis 2. was rejected.

Socio-economic background. Analysis of the data when work alone was compared to work plus knowledge of a high socio-economic background and a low socio-economic background, showed no statistical significance. Therefore, Sub-hypothesis 3. was not rejected.

Sex factor. Analysis of the data when work alone was compared to knowledge of the sex of the child, indicated that neither male teachers nor female teachers graded boys and girls any differently. Therefore, null Sub-hypothesis 4. was not rejected.

Chapter 1

INTRODUCTION

The practice of grading is used in school systems across our nation. (Education Index, 1965-1970) It is an activity expected by students, parents, and teachers. Students expect grades to the extent that they become ends in themselves. Parents often use grades of students to promote social status, compare siblings in the family, and as a basis for future plans. Teachers use grades to group students, retain or pass students, motivate, and threaten.(Austin, 1965) (Palmer, 1964)

Numerous studies have been done on various phases of grading including such areas as report cards, grading systems, motivation of students, social implications of failure, and predictions of success in college and business. (Education Index, 1965-1970) It seemed to the writer that one area neglected by researchers was the human factor of the teacher in grading. During nine years of experience by the writer, it appeared not uncommon when a child received a poor grade to say he didn't try, it was a bad day for him, and so on. Seldom was it stated that the reason for a low grade was some factor other than work itself that influenced a teacher to mysteriously arrive at a grade. Children often said, "The teacher doesn't like me." This type of statement was nearly always negated by the parent and the child was made to believe that it was not the teacher at fault, but some-

thing he, the student, was doing wrong. Perhaps children perceived more than they were given credit for. (Fink, 1962) (Gaier, 1966) (Pollack, 1968) (Yonge and Sassenrath, 1952)

STATEMENT OF THE PROBLEM

The problem of this study was to test whether or not teachers were influenced by external factors when they graded a piece of work by a student.

NEED OF THE STUDY

If, when grading a piece of work, teachers were being influenced by factors other than the quality of a student's work, and if they were unaware of this fact, or if they were aware of a bias and chose not to recognize it, this should be brought to their attention. In an informal survey conducted by the writer, it was noted that about fifty percent of the teachers interviewed commented that they definitely were not influenced by the external factor of physical appearance. However, observation of the teacher while responding to the interview led the writer to question such statements. If it could be shown that teachers did grade because of external factors, and they could be shown this tendency did exist, it would contribute to a more cognitive attitude on the part of the teacher due to that awareness.

NULL HYPOTHESIS

The major hypothesis tested was based on the assumption that elementary teachers working with students in the intermediate level of education were not influenced by factors other than the work itself. Therefore, grades assigned by teachers on the basis of work alone would not differ significantly from grades assigned when other factors were known by the teacher. To answer this, questions were asked concerning each external factor tested. The external factors investigated in this study were (a) physical appearance of the student, (b) an I. Q. score for each student, (c) socio-economic background ascribed to each student, and (d) the sex of the student.

GENERAL PROCEDURE

Teachers graded work samples of twenty-five students enrolled in the fifth grade. (See Appendix page 35.) The first experiment established a control group based on work alone. Subsequent experiments had only one of the previously mentioned variables introduced at a time.

The teacher was introduced to the study and was given specific information concerning the group of students. (See Appendix page 35.) They were then asked to grade the sample using a scale of one to ten. One was the lowest grade, and ten the highest. As soon as they had determined the grade for the sample, they verbally told the interviewer and it was recorded.

LIMITATIONS

The sample population was limited to fifth grade students in a Montana elementary school. The respondents were limited to teachers who had worked with pupils in grades four, five, and six within the last five years. The availability of teachers for interview was limited. Teachers in Helena, Bozeman, and Livingston were selected in addition to teachers attending Montana State University. Selection of work graded was not all taken from the same class assignment. Teachers responding to the experiments in the socio-economic group low and socio-economic group high, were told the students were from a deprived farming community and a wealthy ranching community, respectively. The terms, "farming community" and "ranching community", were chosen because they are familiar terms to Montana teachers. A specific instrument was not cited as the source for I. Q. scores prior to the interview. Had any respondents asked, "a standardized group test" would have been given as the source for the I. Q. score on each student. This was done in an attempt to control the effect that a specific test or an individual test might have on a teacher. A limited selection of research articles was available to the writer. Library research was obtained from Montana State University, Bozeman, Montana.

DEFINITION OF TERMS

For the purpose of this study, the following terms have been used:

Work: An original story created by each student which was usually one paragraph in length. (See Appendix page 38.)

External factors: Those factors that are anything other than the work as explained above.

Physical appearance: The physiognomy of the student as seen in a photograph of the student. (See Appendix page 38.)

I. Q. score: A numerical figure denoting intelligence potential as determined by a standardized intelligence test. (See Appendix page 39.)

Socio-economic background: This term was used in two experiments. In the first, the respondents were told the students came from a poor farming community, resulting in deprived living conditions. In the second experiment, respondents were told the students came from a rich ranching community.

Grading: For the purposes of this study, marking and grading have been used interchangeably.

The first step in this study was to review the related literature. This is discussed in Chapter 2.

Chapter 2

REVIEW OF RELATED LITERATURE

Professional journals in education have an abundance of articles concerning the experiments on marking. These articles cite authorities in the fields of education and social sciences to back up their statements, but seldom included the statistical data to prove their statements. Therefore, the validity of these studies appeared to be questionable.

Empirical studies conducted recently that are directly related to the problem seemed limited. Because of changing educational philosophy, studies prior to 1935 have not been included.

INFLUENCE OF EXTERNAL FACTORS

A review of literature showed a number of studies were concerned with a variety of external factors. Some of these factors and their influences will be discussed although they are not directly related to the experiments conducted in this study.

Grading is not an absolute. Gaier (1952:316) stated,

> Because of low reliability of grades both between and within the same instructors, institutions, and subject matter areas, plus the low predictive validity of grades, caution must be exercised in accepting grades as representatives of true ability.

In a later article, Gaier (1966:46) said, "It must be asked, what student was taking what course with what instructor in what school with what other students at that period of time."

Matlin and Mendelsohn (1965) concluded that personality adjustment measured by the California Test of Personality was more strongly associated with grades than with standard achievement test scores.

With grade school children, Davidson and Lang (1960:108) pointed out that the "more positive the children's perception of their teachers' feelings, the better was their academic performance, and the more desirable their classroom behavior as rated by the teachers." What might be reflected here is that personality factors become associated with scholastic evaluation in that teachers tend to assign grades on the basis of adjustment as well as accomplishment. When well-adjusted and poorly-adjusted students perform equally well, the better grades may be given to the students whose behaviors are approved by the teacher.

Friedenberg and Roth (1954) found that personality was involved in teacher-student relations. They concluded that students receiving high marks ascribed characteristics to themselves more like those selected by the teacher to describe themselves. Students receiving low grades did not have a high correlation of the teacher's characteristics.

Fink's study (1962) reported a relationship between the adequacy of the self-concept and the level of academic achievement, more so for boys than for girls.

Physical Appearance

Rosenthal (1966) conducted an experiment in which the experimenter

was to show a series of ten photographs of peoples faces. The subject was to rate the degree of success or failure shown in the face of each person pictured in the photo. Each face could be rated from a -10, meaning extreme failure, to +10 which means extreme success. Each experimenter was then given information about previous results from an identical test. Although all experimenters said exactly the same things and showed the pictures in identically the same way, the experimenters who thought that their photos had been rated as successful by prior experiment, had clear cut results showing their ratings as successful. The reverse was also true. If the experimenter thought that prior experimenters had rated the faces as unsuccessful, his results showed to be unsuccessful. It was noted that such clear-cut results were not common in behavioral research, so two replications were conducted. Both of the subsequent experiments gave the same results.

Jacobson, (1966) asked two groups of teachers to rank a set of unknown children's photographs on their American or Mexican appearance. "American" was not defined. The teachers agreed highly on their rankings. Then the same groups of teachers were asked to rank in the same manner photographs of Mexican children who were students in the school of one group of teachers but unknown to the other group of teachers. Here there was little agreement. The teachers at the school where the Mexican children attended school saw those with higher IQ's as looking more American. The significant correlation of IQ and appearance was

present only where the IQ scores were available. Apparently, teachers agreed in their perception of "Mexican-looking" until they knew how a child tested, and then perception was changed.

In a rather lengthy and complicated study, Rosenthal (1968:82) reported that "Mexican boys who looked more Mexican benefited more from teachers' favorable expectations than did the Mexican boys who looked less Mexican." Although Rosenthal stated there was no clear explanation for these findings, he speculated that the teachers' pre-experimental expectancies of the more Mexican-looking boys' intellectual performance was probably lowest of all. Because of a more favorable expectation in the minds of their teachers, those children may have had the most to gain.

I. Q. SCORE

In an investigation of 235 high school pupils taking algebra for the first time it was reported that although there were no statistically significant differences between boys and girls with respect to I. Q. , the teacher assigned part of the mark on the basis of intelligence. It was further concluded the grades assigned to boys by men teachers were not as greatly affected as grades assigned to girls by women teachers. "As to be expected, however, intelligence is a factor in the assignment of marks by both men and women teachers to both boys and girls." (Carter, 1952:218)

Studies by Crow (1964), Gordon and Durea (1948), Rosenthal (1964),

and Sacks (1952) cited by Rosenthal (1968) to show that an I. Q. score could be raised when an examiner behaved more warmly toward a subject. In his study, Sacks (1952) showed an increase of nearly ten I. Q. points for warmly treated nursery school children as opposed to more indifferently treated children.

Along this same line of thinking, Larrabee and Kleinsasser (1967) performed experiments that showed that when a child's examiner expected superior performance, a total I. Q. earned was 7.5 points higher on the average than when the child's examiner expected inferior performance.

Leonard Cahen (1966) investigated the extent that false information given to a teacher would influence the teachers' scoring of pupils' tests. To test the hypothesis, each teacher in a teachers-in-training class of 256, was asked to score a new test of "learning readiness". Each was told that children who scored higher on other I. Q. and reading tests, would score higher on this new test. On the front of each test booklet, the pupil's I. Q. and reading level were indicated. The results clearly showed that allegedly bright children were given a much greater benefit of the doubt than tests of children who were allegedly duller. A study by Pitt (1956) reported no effect of I. Q. scores on teachers. His sample was comprised of 165 fifth-grade boys with an I. Q. of 94 or higher. An intelligence test was administered to them and then they were divided into three groups. The I. Q. scores were reported accurately to one teacher, ten points were

arbitrarily added to the scores of a second group, and ten points were arbitrarily deducted from the scores of the third group. This was done at the beginning of the school year although teachers knew students nearly two months before obtaining the I. Q. score. At the end of that year, a comparison was made of all three groups as to school grades, achievement tests, teacher ratings, and pupil self-ratings. Results showed essentially no effects on the objective tests of achievement of the arbitrarily raised or lowered I. Q.'s. However the study further indicated that although the belief of the pupils' I. Q.'s did not affect pupils' academic performance, it did affect the pupils' views of themselves, of their teachers, and of the school.

Socio-economic Background

Attention has been focused on the disadvantaged child in our educational system. Rosenthal (1968:48) says,

> Numerous reports indicate that the I. Q. scores of disadvantaged children are lower than those of middle-class children, their reading is substandard, their attitudes are negative, and their behavior is annoying to teachers .... Disadvantaged children by definition come from lower socio-economic groups where low income is married to values alien to the school culture.

Sexton's (1961) study on the relation between income and educational opportunity pointed out that where the family income was lower than $7000, achievement was below grade level; and where the income exceeded $7000, the achievement was above grade level. Similarily, Kahl (1961) found that boys of high status performed at much higher levels

than boys of low status although the two groups had similar I. Q. scores.

Teachers in lower-class schools did not set standards as high as those in middle-class schools, nor were they as concerned with bringing their children up to grade level, was the statement made by Clark (1962). There appeared to be a difference in the way students of the various classes were treated. Carlson (1964:269) alluded to this when he noted,

> Preferential treatment may be given the clients that accommodate themselves to the purposes of the organization. Example: research has shown that preferential treatment is frequently given in the public schools to middle and upper-class children in such matters as discipline, punishment, and curriculums.

Similar findings were noted by Wilson (1963) when he stated that teachers held lower standards for children from the "poor part of town" than for the children from the "better part of town." Rosenthal (1968:107) concluded, "this lowering of standards or grading too high for a given performance level, may actually result in the lessened profit from education found among disadvantaged children."

Sex Factor

A number of articles in professional journals referred to the differences between the sexes in grading practices. (Day, 1938; Douglass, 1937; Kluger, 1968; Lobaugh, 1942; Newton, 1942; Skinner, 1944; Swenson, 1942) An article by Jack Harrison Pollack (1968) interested the writer and indicated a need for the study. This article in Today's Health entitled "Are Teachers Fair to Boys" cited a number of

authorities to substantiate his claims. Statistical data was not included in the article and thus reduced its effectiveness to opinions by authorities.

Evidence of sex differences in marks assigned by teachers was reported by Edminston (1943). He stated that in a high school in Indiana, the average grade for girls was 84.4 while boys received 80.0 as an average. Women teachers gave grades that averaged out 5.4 points above those given to boys while men teachers were less partial to girls, giving them an average of 3.4 points above those given to boys.

Although grades were not specifically mentioned, Pollack (1968) stated that girls received more considerations than boys. He further noted that psychologist, Phillip W. Jackson, University of Chicago, agreed that boys got into eight times more trouble than girls.

Rosenthal (1968) experimented with the degree of "expectancy advantage" of boys and girls placed in a three track system. He concluded that as a whole for the school, boys and girls showed the same degree of expectancy advantage. Differences showed up when the expectancy for the different tracks was calculated. For girls there were favorable expectations only in the top two tracks; for boys only in the bottom two tracks. The teachers of these students had been told the students were "late bloomers" and were about to show a spurt.

In summary, researchers seemed to agree that there was "a low reliability of grades both between and within the same instructors,

institutions, and subject matter areas, plus a low predictive validity of grades ...." (Gaier, 1952:316) Factors that had an influence on grading included such intangibles as personality adjustment, perception of a teachers' feelings, self-concept, and correlation of the teacher's and student's characteristics.

Literature directly related to the experiments used in this study was most limited in the area of physical appearance, but seemed to indicate that physical appearance was influencial. Several studies were conducted concerning the effect of an I. Q. score. In all but one study, results implied that knowledge of an I. Q. score influenced teachers as well as influencing the performance of the student. Available literature indicated a great deal of attention to socio-economic factors and implications to ethnic groups and races. Researchers seemed to agree that teachers set lower standards for students from deprived areas. The influence that the sex of a child had on the teacher received a great deal of attention in popular professional magazines. Unfortunately, statistical data was not included and the authors relied on the opinions of authorities in the field. The available literature usually pointed to a mixture of discipline problems, readiness, and grades as they were related to boys and girls.

The second step of the study was to collect data and analyze it. For this discussion, the reader is referred to Chapter 3.

Chapter 3

STATISTICAL DATA AND INTERPRETATIONS

The purpose of this experiment was to determine if a teacher's evaluation of a student's work, and consequently the grade assigned to that work, was influenced by factors other than the quality of the work itself. Such other factors included the physical appearance of the student, the student's I. Q. score, the student's sex, and the student's socio-economic background.

Knowledge of how these factors influenced the teacher's objectivity should help the teacher to grade on a more objective basis.

POPULATION DESCRIPTION

Data were gathered by personal interview from seventy teachers who had been actively teaching academic subjects within the last five years in the intermediate grades of Montana elementary schools. School systems in Helena, Bozeman, and Livingston, plus teachers attending Montana State University supplied the selection of teachers. Because teachers were interviewed only during free time, it was not practical to use a table of random numbers or a cluster sampling to pick subjects for interview. Therefore, any teacher who met the teaching qualifications and was available was asked to participate in the experiment. The method of selection of respondents was acceptable for the purposes of this study.

The empirical data consisted of the teacher's evaluation, on an

ascending scale of 1 to 10, of the work of each of twenty-five students. The students were all members of a heterogeneously grouped class of fifth graders attending public school in a Montana town of 2,000 population.

HYPOTHESES

The major hypothesis formulated to guide the experimental process, along with the sub-hypotheses, were as follows:

Major Hypothesis

A teacher's evaluation of a student's work is not influenced by factors other than the work itself. Therefore, grades assigned by teachers on the basis of work alone will not differ significantly from grades assigned when other factors are known by the teacher.

Sub-hypothesis 1. Teacher's evaluation of student's work is not influenced by the physical appearance of the student. Therefore, grades assigned by teachers on the basis of work alone will not be significantly different from grades assigned by teachers who also know the physical appearance of the student.

Sub-hypothesis 2. Teacher's evaluation of student's work is not influenced by knowledge of the student's I. Q. score. Therefore, grades assigned by teachers on the basis of work alone will not be significantly different from grades assigned by teachers who also know

the students I. Q. score.

Sub-hypothesis 3. Teacher's evaluation of student's work is not influenced by knowledge of the student's socio-economic background. Therefore, grades assigned by teachers on the basis of work alone will not be significantly different from grades assigned by teachers who also know the socio-economic background of the student.

Sub-hypothesis 4. Teacher's evaluation of student's work is not influenced by knowledge of the student's sex. Therefore, grades assigned by teachers on the basis of work alone will not be significantly different from grades assigned by teachers who also know the student's sex.

It was assumed that acceptance, on a statistical basis, of the major hypothesis must lead to acceptance of all four sub-hypotheses.

COLLECTION AND ORGANIZATION OF DATA

The first ten teachers were introduced to the study (See Appendix page 35) and were shown the work performed by the students. They were then to evaluate the work and assign a numerical grade for each student. They were given no information other than the students were in the fifth grade in a Montana town. (For a sample of the work, see the lower part of page 38 in the Appendix.) The grades were tabulated and each student was assigned the average of the ten individual teachers' grades. This

group was considered the Control Group and is labeled Work Only in Table 4. (page 41 in the Appendix).

A second group of teachers was introduced to the study. (See page 35 in the Appendix.) These teachers were shown the same samples of students' work. However, this time a picture accompanied each sample of work. (See page 38 in the Appendix.) These teachers were then asked to evaluate each student's work on the same scale of 1 to 10. The grades were tabulated and the average of the ten individual teachers' grades was assigned to each student. These grades are labeled Work Plus Appearance in Table 4, (page 41 in the Appendix.)

A third group of ten teachers was introduced to the study and shown the work and the I. Q. score of the student. (See page 39 in the Appendix.) The instrument used to measure the I. Q. score was not given unless requested. Although no teacher requested the source of the I. Q. score, had they asked, they would have been told it was derived from a standardized group intelligence test. This third group was asked to evaluate each student. The same procedure of averaging each student's grades as previously discussed was used and the results are shown in Table 4. (page 41 as Work Plus I. Q..)

A fourth group of ten teachers was introduced to the study and told the students were from a low socio-economic environment. (See page 36 in the Appendix for the text of the semi-structured interview.) The work was the same for all experiments. Any information given at the top

of the page was covered. (For an example of the work, see page 38 in the Appendix.) The fifth group of teachers was treated in the same manner differing only in that they were told the students were from a high socio-economic environment. For the purposes of this study, low socio-economic environment was referred to as a "poor farming community". High socio-economic environment was equated with a rich ranching community. The students' grades assigned by each group of teachers were averaged and the results appear in the two columns as Socio-economic Low and Socio-economic High in Table 4. (page 41 in the Appendix).

A final group of twenty teachers consisted of ten male and ten female teachers. These teachers were introduced to the study and asked to grade ten students who had been selected from the original twenty-five. The ten examples were selected on an inability to tell the sex of the student who had written the sample of work. Five of the male teachers were informed that the samples of work were written by boys. Five male teachers were told the samples were written by girls. The same procedure was used with women teachers by telling them the samples had been written by boys and then telling a second set of female teachers the samples were all done by girls. Each group of five teachers then assigned grades to the students' work. The grades were tabulated and the average of each group of five grades was then assigned to the students. The results of this tabulation are shown in Table 5. (page 43 in the Appendix.)

STATISTICAL MODEL AND ANALYSIS OF MAJOR HYPOTHESIS

The statistical model used was the "Randomized Blocks, One-Variable Classification Model" with the conclusions inferred from the Analysis of Variance of this model. The reason for using this model was to make comparisons between a set of treatments within blocks of relatively homogenous experimental material. As used in this experiment, variations in assigned grades given by the teachers constituted the treatment effect and variations among students comprised the block effect.

The analysis was carried out and the results are summarized in Table 1, the Analysis of Variance table. (Tabulation of data for Table 1. is located on page 42 in the Appendix.)

Table 1. Analysis of Variance

Source	SS	df	MS
Columns (between teachers)	13.675	4	3.41375
Rows (between students)	353.577	24	14.73237
Error	22.681	96	.23626
Total	389.933	124	----

The testing statistic is the ratio of the Mean Square Columns to Mean Square Error, distributed as the F-statistic with 4 and 96 degrees of freedom in the numerator and denominator respectively.

The calculation of this statistic is

$$F_{4,96} = \frac{MSC}{MSE} = \frac{3.41375}{.23626} = 14.449.$$

The tabulated F value at the .01 level of significance is

$$F_{4,96}\ (a=.01) = 3.52$$

Since the test F-statistic exceeded the tabulor F, the conclusion was that there was a significant difference between the columns. Consequently the major hypothesis was rejected. Rejection of the major hypothesis led to the testing of each sub-hypothesis.

TESTING THE SUB-HYPOTHESES

A Least significant Difference table was constructed to test Sub-hypotheses 1., 2., and 3. The test statistic is the absolute difference in treatment means (i. e. $/ \bar{x}.j - \bar{x}. k/$). The LSD was computed as

$$LSD_a = \sqrt{\frac{2}{r} (MSE) F_{1,96,a}}$$

The LSDs for a = .05 and a = .01 were:

$$LSD_{.05} = \sqrt{\frac{2}{25} (.23626) \quad 3.96} = .27325$$

$$LSD_{.01} = \sqrt{\frac{2}{25} (.23626) \quad 6.92} = .36166$$

Table 2. Least Significant Difference Table

		W.O. 4.744	W.A. 5.684	W.I.Q. 5.14	W.S.L 4.86	W.S.H. 4.948
W.O.	4.744	---				
W.A.	5.684	.940**	---			
W.I.Q.	5.14	.396**	.544**	---		
W.S.L.	4.86	.116	.824**	.280*	---	
W.S.H.	4.948	.204	.736**	.192	.088	---

W. O. is Work Only. W. A. is Work plus Appearance.
W.I.Q. is Work plus I. Q. Score. W.S.L. is Work plus Socio-Economic Low.
W.S.H. is Work plus Socio-Economic High
* Significant difference ** Highly significant difference

The values shown in the columns and rows at the top and side of Table 2. were obtained from Table 4. (page 41 in the Appendix. Table 4. is the average of grades of ten teachers given on a basis of work only and work plus other knowledge.) The values shown within Table 2. were the differences between each row and column. These values were compared with the values calculated for the Least Significant Difference at the .01 and the .05 level of significance shown on page 22.

The Least Significant Difference Table, Table 2., resulted in the following conclusions:

1. There was a significant difference between teachers' grades based on work alone and grades based on work plus physical appearance. Therefore, Sub-hypothesis 1. was rejected.

2. There was a significant difference between teachers' grades based on work alone and grades based on work plus knowledge of the student's I. Q. score. Therefore, Sub-hypothesis 2. was rejected.

3. There was no significant difference between grades based on work alone and grades based on work plus knowledge of the student's socio-economic environment. Therefore, Sub-hypothesis 3. was not rejected.

Sub-hypothesis 4. was tested using the same statistical model used to test the major hypothesis. That model was the Analysis of Variance as shown in Table 3.

Table 3. Analysis of Variance for Sub-hypothesis 4.

Source	SS	df	MS
Columns (between teachers)	2.003	3	.66767
Rows (between students)	105.881	9	11.76456
Error	7.627	27	.28248
Total	115.511	39	

The calculation of this statistic is

$$F_{3,27} = \frac{MSC}{MSE} = 2.3636$$

The tabulated F value for the significance level of $\alpha = .01$ is

$$F_{3,27,.05} = 2.96$$

The conclusion from this analysis was no significant difference existed between grades based on work alone and grades based on work plus knowledge of the student's sex. Therefore, Sub-hypotheses 4. was not rejected.

Table 5. (page 43 in the Appendix) contains the average of the five grades assigned by male and female teachers when they were told the sex of the student.

The final step was to draw conclusions from the analysis of the statistics. These, plus recommendations are in Chapter 4.

Chapter 4

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

SUMMARY

The central idea of this study was to examine the effect that external factors had on a teacher when grading a piece of work. This was not a new idea but there appeared to be limited empirical research and much opinion. Professional journals frequently had articles pertaining to various aspects of the problem, but failed to include the necessary statistical information. In the majority of cases, these articles relied on the experience of one person in a given case, or cited opinions by authorities in the field, omitting statistical data, thus reducing much of the literature to the "interesting reading" classification.

This study was confined to experiments using physical appearance, I. Q. score, socio-economic background, and the sex of the student as the external factors to be tested. The influence of external factors other than those just mentioned were discussed in Chapter 2. Researchers seemed to agree that there was "a low reliability of grades both between and within the same instructors, institutions, and subject matter areas, plus a low predictive validity of grades, ..." (Gaier, 1952:316) Factors that had an influence on grading included such intangibles as personality adjustment, perception of a teacher's feelings, self-concept,

and correlation of the teacher's and student's characteristics.

The physical appearance of the student as seen in a photograph was the first external factor tested. Research in this area appeared to be very limited. Rosenthal (1966) conducted an experiment based on photographs. Respondents were rating the degree of success or failure shown in the face of each person pictured in a photo. Results were clear cut. If the experimenter thought that prior experimenters had rated the faces as unsuccessful, his results showed to be unsuccessful. The reverse was also true.

Jacbson (1966) asked two groups of teachers to rank a set of unknown children's photographs on their American or Mexican appearance. The teachers agreed highly on their rankings.

The effect of an I.Q. score on a teacher had somewhat varied findings. Carter (1952) reported that teachers assigned part of a mark on the basis of intelligence. Other researchers noted that an I. Q. score could be raised in points when a teacher had a favorable expectancy of a child. This effect, however, was most dramatic with primary students. It appeared that teachers were not as influenced in upper elementary and advanced grades. This was pointed out in a study by Pitt (1956) that showed no changes occurred in evaluation of fifth-grade boys when points were arbitrarily added to and deducted from the actual I. Q. scores. There were some questions about the reliability of the results. Teachers knew the students for nearly two months before they were given

the I. Q. score.

Available literature indicated a great deal of attention to socio-economic factors and implications to ethnic groups and races. Researchers seemed to agree that teachers set lower standards for students from deprived areas. They further agreed that conditions in those areas tended to make it impossible for culturally deprived students to compete or even be compared with more advantaged children. However, Tulkin (1968) stated that with adequate control measures it was possible to demonstrate that racial groups are not significantly different on measure of intelligence and school achievement.

The influence that the sex of a child had on the teacher received a great deal of attention in popular professional magazines. Unfortunately, little empirical data was available to substantiate the claims of the writers. They did cite authorities in the field, but without statistics, their opinions were little better than interesting reading. The available literature usually pointed to a mixture of discipline problems, readiness, and grades as they were related to boys and girls. Seldom did any study discuss only the relationship of the sex of the child to the grade.

The purpose of this study was to determine whether or not external factors influenced a teacher when grading a piece of work. Not all external factors could be tested. Therefore, this study was limited to (a) physical appearance of the student, (b) an I. Q. score for each

student, (c) socio-economic background ascribed to each student, and (d) the sex of the student.

Data were collected from seventy teachers who had been actively teaching in the intermediate grades of elementary schools in Montana. The participating teachers graded samples of work of twenty-five students in the fifth grade.

CONCLUSIONS

Analysis of the statistical data showed there was a significant difference between the grades a teacher assigned to a piece of work and the same piece of work when other factors had been introduced. Because of this difference, the major null hypothesis was rejected, which resulted in the testing of each of the four sub-hypotheses.

Sub-hypothesis 1. stated that a teacher's evaluation of a student's work is not influenced by the physical appearance of the student. Analysis of the data indicated to the contrary. Therefore, Sub-hypothesis 1. was rejected.

It was not uncommon to hear people speak of "first impressions" whether they were good or bad. The writer was aware that this experiment was based on first impressions by a teacher. The teachers in the study had no prior knowledge of any of the students and therefore, had to base their judgements on information given them. It was the opinion of the writer that although such judgements existed, they would not exist for

an entire academic year. By this, it was meant that although a teacher might have an unfavorable or even favorable impression of a student at the beginning of the year, this impression would most likely change as other factors were introduced into the relationship. A situation where teachers, administrators, or leaders had to make snap judgements on first impressions would validate the conclusions of the experiment.

Sub-hypothesis 2. stated that a teacher's evaluation of a student's work was not influenced by knowledge of the student's I. Q. score. Analysis of the data showed there was a significant difference between grades assigned to work alone and work when an I. Q. score was known. Thus, Sub-hypothesis 2. was rejected.

The writer felt this was an unfortunate result because of the inaccuracy of intelligence measurement. Individual intelligence tests tended to give a more accurate picture of the child's potential, but few schools, if any, administered individual tests to each student. Even when an individual test had been used, most teachers had not had an opportunity to study the tests to the degree that they fully understood the implications of the results.

Sub-hypothesis 3. was concerned with two experiments that involved socio-economic status. One group of teachers was told the students were from a deprived farming community while a second group of teachers was told the students were from an affluent ranching community. The terms "farming community" and "ranching community" were used because

Montana teachers would be more familiar with students in a community of such a description.

Analysis of the data showed no significant difference in the grading of students when that information was known. Therefore, Sub-hypothesis 3. was not rejected.

The writer felt an explanation of the results could be that such terminology did not evoke a strong response in the minds of the experimental teachers. Had other terms been used such as "Indian reservation", "Negro ghetto", "metropolitan suburb", etc. perhaps a significant finding would have been obtained. These terms, on the other hand, might not fit Montana teachers' experiences.

Sub-hypothesis 4. stated that a teacher's evaluation of a student's work is not influenced by the sex of the student. In spite of some common beliefs, the statistical data of this study found no significant difference in the grades a female teacher gave to boys or girls nor a difference in the grades male teachers gave to boys or girls. Therefore, Sub-hypothesis 4. was not rejected.

The writer must then assume that teachers did not grade a student based solely on the sex of the student. Other factors must then be influencing the evaluation of a student if the reader is to assume that the popular belief, held by so many educators, has any merit.

RECOMMENDATIONS

The writer recommended that:

(1.) a repeat of this study be conducted in other states, changing the terms to fit the population and experiences of the teachers in that state.

(2.) a study be conducted to expand the information available concerning the effect of physical appearance on a teacher's evaluation.

(3.) a study be conducted to determine the effect of socio-economic background on the teacher. The writer suggested that care be taken in choosing the title for socio-economic groups being tested so as to elicit the greatest amount of response.

(4.) a study be conducted to experiment to a greater extent with the effect of the sex of the student in relation to grades.

APPENDIX

INTRODUCTION TO STUDY AND INSTRUCTIONS

WORK ONLY

I am conducting a survey of teachers to obtain statistics for a master's paper.

My paper is attempting to determine whether or not external factors influence a teacher's grading. By external factors, I mean such things as the sex of the student, socio-economic background, physical appearance, I. Q. score and so forth.

I would like for you to read these twenty-five examples and grade them on a basis from 1 to 10, one being lowest and ten being highest. You may go as quickly as you wish telling me only the number you have chosen for each example.

In deciding upon a grade, take into consideration the performance of the child and the content of the work. (If this needed more explanation, it was given.) These samples are not all from the same assignment.

The only specific information I can give you is that these children are all members of a fifth grade class in Montana.

PHYSICAL APPEARANCE

The teacher was introduced to the study as explained. The teacher's attention was drawn to the photograph at the top of the page and they were asked to evaluate the work in the same manner as previously described taking into consideration their feelings about the photograph. (For an example, see page 38.)

I. Q. SCORE

Introduction to the study, instructions on how to grade, and information were given to the teacher the same as described for "work only" on page 35. The only difference in this experiment is that an I. Q. score was printed at the top of the page. This information was drawn to the teacher's attention. (For an example see page 39.)

SOCIO-ECONOMIC BACKGROUND

The term socio-economic background was used in two experiments, with the explanation of a low socio-economic background being equated with a poor farming community and a high socio-economic background being equated with a rich ranching community. The teacher was introduced to the study as described in "work only" on page 35. The only thing visible on the page was the sample of work at the bottom, all other material had been covered. The teacher was given information

about the student depending upon which experiment they were responding to. (For an example, see page 38. However, the information at the top was obscured for this experiment.)

SEX FACTOR

Teachers were introduced to the study as previously described in the "work only" section on page 35. However, these teachers were given somewhat different information. Each teacher was told they were participating in a control group and would not be evaluating every student in the study, but would evaluate only the ones designated by the interviewer. The teacher was then told all of the samples were done by girls (or boys) and they were to evaluate them as previously described on page 32. Any material at the top of the page had been covered so the teacher could see only a sample of the work. (For an example, see page 39. The I. Q. score at the top of the page had been covered.)

Plate 1. Work Plus Physical Appearance



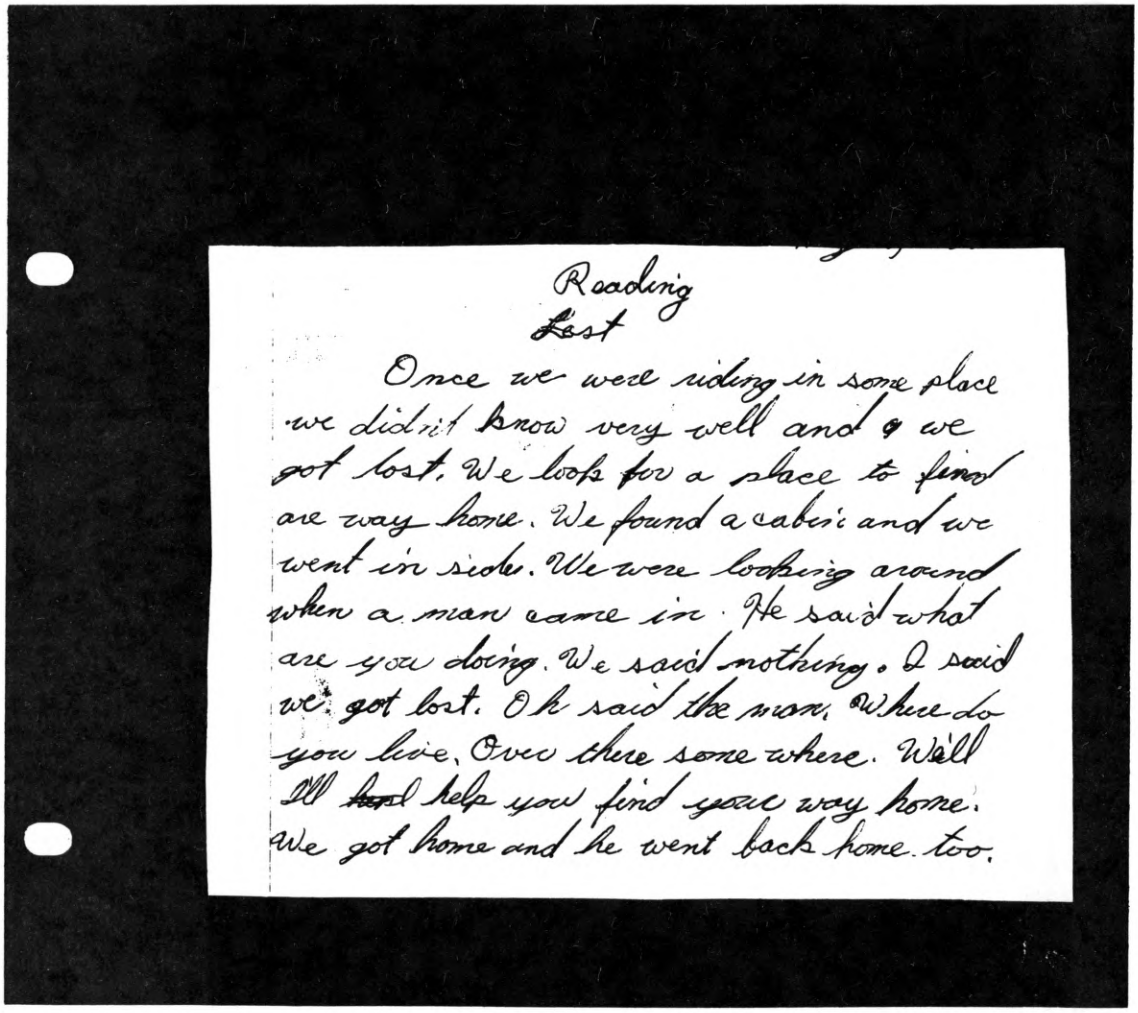
Reading
Lest

Once we were riding in some place we didn't know very well and we got lost. We look for a place to find are way home. We found a cabin and we went in side. We were looking around when a man came in. He said what are you doing. We said nothing. He said we got lost. Oh said the man. Where do you live. Over there some where. Well I'll help you find your way home. We got home and he went back home too.

18

Plate 2. Work Plus I. Q. Score

94



Reading
Lost

Once we were riding in some place we didn't know very well and we got lost. We look for a place to find are way home. We found a cabin and we went in side. We were looking around when a man came in. He said what are you doing. We said nothing. I said we got lost. Oh said the man. Where do you live. Over there some where. Well I'll help you find your way home. We got home and he went back home too.

STATISTICAL INFORMATION

The statistical model applicable to this type of experiment was the "Randomized Blocks, One-Variable Classification Model, with conclusions inferred from the Analysis of Variance of this model, which was used on the major hypothesis and Sub-hypothesis 4.

The linear form of this model is

$$x_{ij} = u + a_i + b_j + e_{ij} ,$$

where x_{ij} is the observation,

u is the population mean,

a_i is the i-th treatment effect,

b_j is the j-th block effect,

and e_{ij} is the random sampling error.

As the model was used in this experiment, variations in teachers' grades constituted the treatment effect and variation in students comprised the block effect.

Table 4, page 37, shows the average of grades of ten teachers given on a basis of work only and work plus other knowledge. Three experiments, plus the control group which is designated by "Work Only" are given in this table.

In Table 5, (page 38) ten students were graded by five male teachers and five female teachers who thought the students were boys. Then these same ten students were graded by five male teachers and five female teachers who thought they were girls.

Table 4. Average of Grades of Ten Teachers Given on Basis of Work Only and Work Plus other Knowledge

Student	Work Only	Work Plus: Appearance	I.Q.	Socio-economic Low	Socio-economic High	ΣX
1	5.5	7.6	4.3	4.8	4.5	26.7
2	3.4	4.9	3.9	5.2	4.5	21.9
3	3.1	4.7	3.2	3.5	3.3	17.8
4	5.8	6.3	7.2	7.3	7.1	33.7
5	5.4	6.3	5.9	6.0	5.8	29.4
6	6.9	7.4	7.7	7.0	6.8	35.8
7	6.4	7.4	7.0	6.5	6.4	33.7
8	2.3	4.3	3.6	2.5	2.5	15.2
9	2.5	4.3	3.0	2.4	3.2	15.4
10	7.9	7.5	7.8	7.9	7.5	38.6
11	5.0	5.5	5.2	5.5	5.0	26.2
12	3.5	4.9	3.8	3.7	3.5	19.4
13	3.8	5.4	4.3	4.3	4.2	22.0
14	2.3	3.5	2.1	1.7	2.2	11.8
15	5.8	5.4	6.1	5.0	5.9	28.2
16	3.0	4.3	3.0	2.8	3.1	16.2
17	4.9	6.4	6.6	6.1	6.5	30.5
18	5.0	6.1	5.7	5.4	5.8	28.0
19	5.9	6.0	6.2	5.8	6.0	29.9
20	3.2	4.4	4.2	2.8	3.4	18.0
21	6.3	6.1	6.9	6.1	6.3	31.7
22	9.5	9.3	9.4	8.8	9.0	46.0
23	3.3	4.3	3.1	3.0	2.6	16.3
24	3.1	4.4	2.9	2.9	3.4	16.7
25	4.8	5.4	5.4	4.5	5.2	25.3
ΣX	118.6	142.1	128.5	121.5	123.7	$\Sigma\Sigma X=634.4$
$\bar{X}i.$	4.744	5.684	5.14	4.86	4.948	$\Sigma\Sigma X^2=3609.64$
ΣX^2	643.90	853.55	747.55	675.61	689.03	$\bar{\bar{X}} = 5.0752$

Tabulation of data for Table 1.

$r = 25 \qquad c = 5$

Correction factor $= C = \frac{(\Sigma\Sigma X_{ij})}{rc} = 3219.707$

$SST = \Sigma\Sigma X^2_{ij} - C = 3609.64 - 3219.707 = 389.933$

$SSC = \frac{1}{r}\Sigma \bar{i}^2_{i.} - C = 3233.382 - 3219.707 = 13.675$

$SSR = \frac{1}{c}\Sigma \bar{i}^2_{.j} - C = 3573.284 - 3219.707 = 353.577$

$SSE = SST - SSR - SSC =$ 22.681

Table 5. Average of Grades Assigned by Five Male and Five Female Teachers When Sex of Student Was Known

Student	Male Teachers		Female Teachers		$\sum_j X_{ij}$	$\bar{X}_{i.}$
	Boys	Girls	Boys	Girls		
1	5.2	4.8	4.2	4.4	18.6	4.65
2	5.8	6.6	6.4	5.4	24.2	6.05
3	4.4	3.8	3.2	3.0	14.4	3.60
4	4.8	3.8	4.2	3.6	16.4	4.10
5	5.8	5.4	6.0	6.2	23.4	5.85
6	5.4	5.2	5.2	5.8	21.6	5.40
7	9.2	8.8	9.0	9.4	36.4	9.10
8	4.4	2.8	3.0	3.0	13.2	3.55
9	3.6	3.4	3.2	4.4	14.6	3.65
10	6.0	5.8	4.0	6.0	21.8	5.45
ΣX_{ij}	54.6	50.4	48.4	51.2	204.6	$=\Sigma\Sigma X_{ij}$
					1162.04	$=\Sigma\Sigma X^2$
$\bar{X}_{ij}$	5.46	5.04	4.84	5.12	5.115	$=\bar{\bar{X}}$

$\Sigma X^2_{.j}$ = 10,485.32 $\quad$ 10,485.32 ÷ 10 = 1,048.532

$\Sigma X^2_{i.}$ = 4,609.64 $\quad$ 4,609.64 ÷ 4 = 1,152.41

$(X..)^2$ = 41,861.16 $\quad$ 41,861.16 ÷ 40 = 1,046.529

SST = 1162.04 - 1,046.529 = 115.511

SSC = 1048.532- 1,046.529 = 2.003

SSR = 1152.41 - 1,046.529 = 105.881

SSE = 115.511 - 2.003 - 105.881 = 7.627

REFERENCES

Books

Cahen, Leonard S. "Experimental Manipulation of Bias in Teachers' Scoring of Subjective Tests." Paper read at American Psychological Association, New York, September, 1966. Cited by Robert Rosenthal and Lenore Jacobson. Pygmalion in the Classroom. New York: Holt, Rinehart, and Winston, 1968.

Carlson, Richard. Behavioral Science and Educational Administration. Sixty-third Yearbook of the National Society for the Study of Education. Chicago: University of Chicago Press, 1964.p. 269-273

Clark, B. R. Educating the Expert Society. San Francisco: Chandler Publishing Co., 1962.

Crow, Linda. "Public Attitudes and Expectations as a Disturbing Variable in Experimentation and Therapy." Unpublished paper, Harvard University, 1964. Cited by Robert Rosenthal and Lenore Jacobson. Pygmalion in the Classroom. New York: Holt, Rinehart, and Winston, 1968.

Education Index. New York: H. W. Wilson Co. 1927-1970.

Jacobson, Lenore. "Explorations of Variations in Educational Achievement among Mexican Children, Grades One to Six." Unpublished doctoral dissertation, University of California, Berkley, 1966. Cited by Robert Rosenthal and Lenore Jacobson. Pygmalion in the Classroom. New York: Holt, Rinehart, and Winston, 1968.

Kahl, J. A. "Common Man" Boys. Cited by A. H. Halsey, Jean Floud, and C. A. Anderson (eds.) Education, Economy, and Society. New York: Free Press, 1961. pp. 348-366.

Larabee, L. L. and L. D. Kleinsasser. "The Effect of Experimental Bias on WISC Performance." Unpublished paper. St. Louis, Missouri Psychological Associates, 1967. Cited by Robert Rosenthal and Lenore Jacobson. Pygmalion in the Classroom. New York: Holt, Rinehart, and Winston, 1968.

Pitt, C. C. V. "An Experimental Study of the Effects of Teachers' Knowledge or Incorrect Knowledge of Pupil IQ's on Teachers' Attitudes and Practices and Pupils' Attitudes and Achievement." Unpublished doctoral dissertation, Columbia University, 1961. Cited by Rosenthal and Jacobson. Pygmalion in the Classroom. New York: Holt, Rinehart, and Winston, 1968.

Rosenthal, Robert. "The Effect of the Experimenter on the Results of Psychological Research." Cited by B. A. Maher (ed.) Progress in Experimental Personality Research. Vol. 1. New York: Academic 1964. pp. 79-114.

_____1966. Experimenter Effects in Behavioral Research. New York: Appleton.

_____and Lenore Jacobson. Pygmalion in the Classroom. New York: Holt, Rinehart and Winston, 1968.

Wilson, A. B. "Social Stratification and Academic Achievement," cited by A. H. Passow (ed.) Education in Depressed Areas. New York: Bureau of Publications, Teachers College, Columbia University, 1963, pp. 217-235.

Periodicals

Austin, Mary C. "Report Cards and Parents," Reading Teacher 18: 660-663, May, 1965.

Carter, Robert S. "How Invalid are Marks Assigned by Teachers," Journal of Educational Psychology. 43: 218-228, 1952.

Davidson, Helen H. & Gerhard Lang,"Children's Perceptions of Their Teachers' Feelings Toward Them, Related to Self-Perception, School Achievement, and Behavior," Journal of Experimental Education, 29: 107-118, December, 1960.

Day, L. C. "Boys and Girls and Honor Ranks," School Review, 46: 282-288, 1937.

Deitz, S. M. and W. W. Purkey. "Teacher Expectations of Performance Based on the Race of the Student," Psychological Reports, 24: 691-694, June 1969.

Douglass, H. R. "Relation of High School Marks to Sex in Four Minnesota Senior High Schools," School Review, 45: 282-288, 1937.

Edmiston, R. W. "Do Teachers Show Partiality Toward Boys or Girls," Peabody Journal of Education. 20: 234-238, 1943.

Fink, M. B. "Self Concept as it Relates to Academic Achievement," California Journal of Educational Research, 13: 57-62, March, 1962.

Friedenberg, Edgar and Julius A. Roth, "Self-Perception in the University. A study of Successful and Unsuccessful Graduate Students," Supplementary Educational Monographs. University of Chicago Press, 80: 5, January, 1954.

Gaier, Eugene L. "The Criterion Problem in the Prediction of Medical School Success," Journal of Applied Psychology. 35: 316-322, October, 1952.

_____1966. "The Grade Society," National Elementary Principal. 45: 42-47, May, 1966.

Gordon, L. V. and M. A. Durea, "The Effect of Discouragement on the Revised Stanford Binet Scale." Journal of Genetic Psychology. 73: 201-207, 1948.

Kluger, Jules, Series of studies in National Elementary School Principal, 1968.

Lebaugh, Dean, "Girls, Grades, and I. Q." Nations School, 30: 42, 1942.

Matlin, Arnold H. and Frances A. Mendelsohn, "The Relationship Between Personality and Achievement Variables in the Elementary School," Journal of Educational Research, 58:457-459, August, 1965.

Newton, R. F. "Do Men Teachers Grade Higher Than Women Teachers," School and Society, 56:72, 1942.

Palmer, Orville, "What Report Cards Don't Tell," Parents Magazine, 39:50-51, 92-95, June, 1964.

Pollack, Jack Harrison, "Are Teachers Fair to Boys," Today's Health, 46:20-25, April, 1968.

Sacks, Eleanor L. "Intelligence Scores as a Function of Experimentally Established Relationships Between Child and Examiner," Journal of Abnormal and Social Psychology, 47:354-358, 1952.

Skinner, M. C. "Failure Ratio; Two Boys to One Girl," Clearing House, 18:264-270, 1944.

Swenson, Clifford, "Packing the Honor Society," Clearing House, 16:521-524, 1942.

Tulkin, Steven R. "Race, Class, Family, and Social Achievement," Journal of Personality and Social Psychology, 9:31-37, 1968.

Yonge, George D. and Julius M. Sassenrath, "Student Personality Correlates of Teacher Ratings," Journal of Educational Psychology, 1:44-52, 1968.