8-1975

The Relationships among Student Ratings of Teachers, Principal Ratings of Teachers, and Student Achievement

Alvin Brossette Jr.
Western Michigan University

Follow this and additional works at: https://scholarworks.wmich.edu/dissertations

Part of the Educational Administration and Supervision Commons

Recommended Citation
Brossette, Alvin Jr., "The Relationships among Student Ratings of Teachers, Principal Ratings of Teachers, and Student Achievement" (1975). Dissertations. 2818.
https://scholarworks.wmich.edu/dissertations/2818

This Dissertation-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Dissertations by an authorized administrator of ScholarWorks at WMU. For more information, please contact maira.bundza@wmich.edu.
THE RELATIONSHIPS AMONG STUDENT RATINGS OF TEACHERS, PRINCIPAL RATINGS OF TEACHERS, AND STUDENT ACHIEVEMENT

by

Alvin Brossette, Jr.

A Dissertation
Submitted to the
Faculty of The Graduate College
in partial fulfillment
of the
Degree of Doctor of Education

Western Michigan University
Kalamazoo, Michigan
August 1975
ACKNOWLEDGEMENTS

The writer wishes to express his appreciation and
gratitude to Drs. Charles Warfield, LeRoy Ray, Jr., and
Rodney Roth, for their advice and willing assistance while
serving as members of the writer's doctoral committee. A
special acknowledgement is extended to Dr. Roth, who pro­
vided the needed assistance while serving as chairman of
the committee.

I am deeply indebted to Drs. Dave Bartz and Charles E.
Townsend, for their support and guidance in the implementa­
tion and completion of this study.

My thanks go to Dr. Gary Vander Lugt, Dr. Wyatt Kirk,
and Jerald Fochtman for their time, effort, and cooperation
towards completion of the present study.

A word of appreciation and acknowledgement is made to
James Thomas, Jr., for years of guidance and encouragement
in my educational endeavors.

Acknowledgement is extended to Dorothy Harley, for her
skill and patience in typing this dissertation.

Gratitude is extended to my wife, Delores, for her
assistance, support, and encouragement. Appreciation is
extended to Derrie, Alicia, and Kathy, for their patience
and sacrifices.

Alvin Brossette, Jr.
INFORMATION TO USERS

This material was produced from a microfilm copy of the original document. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the original submitted.

The following explanation of techniques is provided to help you understand markings or patterns which may appear on this reproduction.

1. The sign or “target” for pages apparently lacking from the document photographed is “Missing Page(s)”. If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting thru an image and duplicating adjacent pages to insure you complete continuity.

2. When an image on the film is obliterated with a large round black mark, it is an indication that the photographer suspected that the copy may have moved during exposure and thus cause a blurred image. You will find a good image of the page in the adjacent frame.

3. When a map, drawing or chart, etc., was part of the material being photographed the photographer followed a definite method in “sectioning” the material. It is customary to begin photoing at the upper left hand corner of a large sheet and to continue photoing from left to right in equal sections with a small overlap. If necessary, sectioning is continued again — beginning below the first row and continuing on until complete.

4. The majority of users indicate that the textual content is of greatest value, however, a somewhat higher quality reproduction could be made from “photographs” if essential to the understanding of the dissertation. Silver prints of “photographs” may be ordered at additional charge by writing the Order Department, giving the catalog number, title, author and specific pages you wish reproduced.

5. PLEASE NOTE: Some pages may have indistinct print. Filmed as received.

Xerox University Microfilms
300 North Zeeb Road
Ann Arbor, Michigan 48106
BROSSETTE, Alvin, Jr., 1942-
THE RELATIONSHIPS AMONG STUDENT RATINGS OF TEACHERS, PRINCIPAL RATINGS OF TEACHERS, AND STUDENT ACHIEVEMENT.

Western Michigan University, Ed.D., 1975
Education, administration

Xerox University Microfilms, Ann Arbor, Michigan 48106

THIS DISSERTATION HAS BEEN MICROFILMED EXACTLY AS RECEIVED.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I STATEMENT OF THE PROBLEM</td>
<td>1</td>
</tr>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>The Problem</td>
<td>11</td>
</tr>
<tr>
<td>Hypotheses</td>
<td>12</td>
</tr>
<tr>
<td>Overview</td>
<td>13</td>
</tr>
<tr>
<td>Limitations of the Study</td>
<td>14</td>
</tr>
<tr>
<td>Definitions of Terms</td>
<td>15</td>
</tr>
<tr>
<td>Organization of the Study</td>
<td>16</td>
</tr>
<tr>
<td>II RATIONALE AND RELATED LITERATURE</td>
<td>18</td>
</tr>
<tr>
<td>Conceptual Framework</td>
<td>18</td>
</tr>
<tr>
<td>Related Studies of Teacher Effectiveness</td>
<td>24</td>
</tr>
<tr>
<td>Overview of Student Ratings of Teachers</td>
<td>36</td>
</tr>
<tr>
<td>Overview of Principal Ratings of Teachers</td>
<td>52</td>
</tr>
<tr>
<td>Summary</td>
<td>60</td>
</tr>
<tr>
<td>III DESIGN AND METHODOLOGY</td>
<td>62</td>
</tr>
<tr>
<td>The Sample</td>
<td>62</td>
</tr>
<tr>
<td>Variables and Instrumentation</td>
<td>64</td>
</tr>
<tr>
<td>Collection of Data</td>
<td>86</td>
</tr>
<tr>
<td>Data Analysis</td>
<td>87</td>
</tr>
<tr>
<td>IV RESULTS</td>
<td>90</td>
</tr>
<tr>
<td>Hypotheses 1-5</td>
<td>91</td>
</tr>
<tr>
<td>Hypotheses 6-13</td>
<td>96</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
<table>
<thead>
<tr>
<th>TABLE</th>
<th>Description of teacher population by number per grade level, sex, and race</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>63</td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>Percent of pupils by city size and geographic region in national population and metropolitan standardization samples</td>
<td>76</td>
</tr>
<tr>
<td>3.</td>
<td>Reliability coefficients and standard errors of measurement for Elementary subtests</td>
<td>77</td>
</tr>
<tr>
<td>4.</td>
<td>Reliability coefficients and standard errors of measurement for Intermediate subtests</td>
<td>78</td>
</tr>
<tr>
<td>5.</td>
<td>Extent of relationship between student ratings of teachers, principal ratings of teachers, and student achievement (reading and mathematics) in grades 4, 5, and 6</td>
<td>92</td>
</tr>
<tr>
<td>6.</td>
<td>Extent of relationship between student ratings of teachers, principal ratings of teachers, and student achievement (reading and mathematics) in grade 4</td>
<td>93</td>
</tr>
<tr>
<td>7.</td>
<td>Extent of relationship between student ratings of teachers, principal ratings of teachers, and student achievement (reading and mathematics) in grade 5</td>
<td>94</td>
</tr>
<tr>
<td>8.</td>
<td>Extent of relationship between student ratings of teachers, principal ratings of teachers, and student achievement (reading and mathematics) in grade 6</td>
<td>95</td>
</tr>
<tr>
<td>9.</td>
<td>One-way analysis of variance for determining the extent of relationship between student achievement (reading) and educational backgrounds of teachers in grades 4, 5, and 6</td>
<td>97</td>
</tr>
<tr>
<td>10.</td>
<td>One-way analysis of variance for determining the extent of relationship between student achievement (reading) and educational backgrounds of teachers in grade 4</td>
<td>97</td>
</tr>
</tbody>
</table>
11. One-way analysis of variance for determining the extent of relationship between student achievement (reading) and educational backgrounds of teachers in grade 5 ........... 98

12. One-way analysis of variance for determining the extent of relationship between student achievement (reading) and educational backgrounds of teachers in grade 6 ........... 98

13. One-way analysis of variance for determining the extent of relationship between student achievement (mathematics) and educational backgrounds of teachers in grades 4, 5, and 6 ............ 99

14. One-way analysis of variance for determining the extent of relationship between student achievement (mathematics) and educational backgrounds of teachers in grade 4 ........ 100

15. One-way analysis of variance for determining the extent of relationship between student achievement (mathematics) and educational backgrounds of teachers in grade 5 ........ 100

16. One-way analysis of variance for determining the extent of relationship between student achievement (mathematics) and educational backgrounds of teachers in grade 6 ........ 101

17. One-way analysis of variance for determining the extent of relationship between student ratings of teachers and educational backgrounds of teachers in grades 4, 5, and 6 101

18. One-way analysis of variance for determining the extent of relationship between student ratings of teachers and educational backgrounds of teachers in grade 4 ........... 102

19. One-way analysis of variance for determining the extent of relationship between student ratings of teachers and educational backgrounds of teachers in grade 5 ........... 103

20. One-way analysis of variance for determining the extent of relationship between student ratings of teachers and educational backgrounds of teachers in grade 6 ........... 103
21. One-way analysis of variance for determining the extent of relationship between principal ratings of teachers and educational backgrounds of teachers in grades 4, 5, and 6 104

22. One-way analysis of variance for determining the extent of relationship between principal ratings of teachers and educational backgrounds of teachers in grade 4 . . . . . . . 104

23. One-way analysis of variance for determining the extent of relationship between principal ratings of teachers and educational backgrounds of teachers in grade 5 . . . . . . . 105

24. One-way analysis of variance for determining the extent of relationship between principal ratings of teachers and educational backgrounds of teachers in grade 6 . . . . . . . 105

25. One-way analysis of variance for determining the extent of relationship between student achievement in reading in grades 4, 5, and 6 106

26. One-way analysis of variance for determining the extent of relationship between student achievement in mathematics in grades 4, 5, and 6 . . . . . . . . . . . . . 106

27. One-way analysis of variance for determining the extent of relationship between student ratings of teachers in grades 4, 5, and 6 107

28. One-way analysis of variance for determining the extent of relationship between principal ratings of teachers in grades 4, 5, and 6 108
<table>
<thead>
<tr>
<th>FIGURE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Student change model of an educational system</td>
</tr>
<tr>
<td>2.</td>
<td>A seven-variable class model for teacher effectiveness</td>
</tr>
<tr>
<td>3.</td>
<td>Nomothetic and idiographic dimensions of behavior</td>
</tr>
<tr>
<td>4.</td>
<td>Metropolitan Achievement Test breakdown</td>
</tr>
<tr>
<td>5.</td>
<td>Metropolitan Achievement Test levels, grades, and time of year administered</td>
</tr>
</tbody>
</table>
CHAPTER I

STATEMENT OF THE PROBLEM

Introduction

The 1970's has been referred to as the decade of accountability in education. According to Hottleman (1974), the accountability movement in public education first became visible in 1970, when President Nixon stated that "school administrators and school teachers are responsible for their performance, and it is in their interest as well as in the interests of their students that they be held accountable" (p. 17). The President was influenced by Assistant Commissioner of Education Lessinger, who openly declared his intention to make public education accountable. Hottleman (1974) stated that the doctrine of accountability was generated out of a theory espoused by Lessinger, that "Americans are becoming fed up with public education; it's too costly, too nebulous, many children have been failed, and nobody's accountable" (p. 17).

The concept of accountability has become a major controversial topic of discussion among legislators, school board members, taxpayers, administrators, teachers, and parents (Popham, 1971). Accountability means different things to different people, and it has been defined in a myriad of
ways. Barrilleaux (1972) spoke of accountability as including the tasks of establishing goals and examining results. Most descriptions of the accountability concept are variations of the following:

Accountability is the product of a process. At its most basic level, it means that an agent, public or private, entering into a contractual agreement to perform a service will be held answerable for performing according to agreed-upon terms, within an established time period, and with a stipulated use of resources and performance standards. (Lessinger, 1970b, p. 9)

Wax (1971) defined accountability in terms of making the schools responsible to the power structure of particular communities not at present represented in the institutions. Wax approached the problem by looking at the way in which children are presently "processed," contrasting it with the learning which arises out of peer association. Further, he saw accountability in terms of responsiveness to the student himself.

Perrone (Olson, Freeman, & Bowman, 1971) noted that in his contacts with educators locally and nationally there have been many arguments about "accountability"; it is all too frequently another defense mechanism, another way of building an exclusive condition, another way of further entrenching traditional values and traditional systems. Perrone further stated:

I see accountability being used as a way of maintaining a traditional set of educational assumptions, predetermined levels of performance, generally based on where we have been and not on the
possibilities of where we might go. And in the
name of accountability new programs are often
compared against systems that are no longer
viable. New programs ought to be allowed to
stand on their own and be judged apart from the
traditional norms and expected outcomes. (p. 55)

Hicks (Olson et al., 1971) gave his interpretation of
accountability when he emphasized:

In the South, the black power structure in the
public school system is on its way out. In my
own state [Louisiana], five or six years ago,
there were 150 black principals of elementary
and secondary schools. Today there are probably
fewer than twenty. The others were phased out.
There is a general tendency toward phasing out
black teachers in schools that have been inte­
grated; persons who have served school systems
for years have been found to be incompetent
because of school desegregation. There are,
then, teachers who were accountable in segre­
gated systems that are no longer accountable
because of desegregation. (p. 57)

According to Bartz (1974), the major purpose of the
Teacher Performance Profile, an accountability scheme used
in the Kalamazoo Public Schools to determine teacher effec­
tiveness, is to serve as a comprehensive teacher account­
ability scheme that (1) provides teachers with feedback from
various sources regarding their performance, and (2) supplies
management-level personnel with evaluative data that aid in
determining teacher effectiveness. The scheme is based on
the rationale that teacher evaluation utilizing multi­
components is better than using principal ratings only.
Bartz noted that "it is believed that the implementation of
the Teacher Performance Profile has resulted in the Kalamazoo
Public Schools being more accountable to students, parents,
and the general public" (p. 11). The Teacher Performance Profile measures performance based on six factors: (1) student ratings, (2) principal ratings, (3) peer ratings, (4) parent ratings, (5) self-ratings, and (6) student achievement data. Bartz (1974) pointed out that "individually these six factors may contain considerable bias, but collectively they constitute an objective, comprehensive, and detailed assessment of a teacher's effectiveness."

Rivers (1974) noted that accountability should be practiced at all levels of education. He further stated that accountability should serve both to improve and prove the quality of education. He believes that different conceptions of educational accountability need to be tested under field conditions, and that experimental efforts in accountability should be critically examined prior to widespread implementation.

In an article by Miller (1974), it was concluded that accountability can be defined as a means of holding groups or individuals responsible for a level of performance or accomplishment for a specific person. Miller stated that to reach the required level of sophistication, educators will have to develop greater skill in setting goals, diagnosing needs, and analyzing learning problems. An intensification of prescriptive teaching, individualization of instruction, and personalized evaluation will take place. Increased emphasis on improved communication and involvement

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
of pupils and parents will be necessary. This intensive participation should result in better understanding and support of the school program.

Lindman (1972) suggested that accountability also needs to consider a cost-analysis approach. Accountability as defined by Locke (1964) would really not imply anything different from an excellent evaluation program. Lindman's additional meaning would make accountability similar to the decision-theory approach and would include the concept of curriculum evaluation. To most users of the term, accountability includes, but goes beyond, establishing goals and examining the results in a cost-analysis framework.

Reasons for the popularity of the accountability concept are not difficult to find. The pressures of society demand that schools find answers to society's problems. In spite of the increased expenditures for our schools, citizens are not elated with their perceptions of the quality of education (Popham, 1971). Tyler (1973) pointed out that educational problems still exist. For example, schools are still having difficulty educating disadvantaged children, and parents of the disadvantaged are demanding that schools be held accountable for educating their children. Popham (1971) stated:

The distressing fact is that we haven't produced very impressive results for the nation's children. There are too many future voters who can't read satisfactorily, can't reason
respectably, don't care about learning in general, and are pretty well alienated from the larger adult society. (p. 107)

According to a Gallup survey (1971), the greatest problem facing the schools in 1971 involved finances. However, little is known about the effectiveness of the billions of dollars the nation spends each year for education. With schools facing shortages of money and with so little knowledge of how effectively educational institutions are presently spending educational dollars, it is easy to see why the cry of accountability has been heard throughout the land. Allen (1970) stated that "considering the undeniable logic of accountability and the obvious need for it, it is amazing that the current push has been so long coming" (p. 53).

According to Allen, we can no longer measure the quality of a school system by inputs such as educational expenditure per child, teacher salaries, and teacher-student ratio, but must be able to measure performance in terms of objectives. Allen further commented:

Within the school system, my hope is that negotiations between school boards and teachers will take into account the factors of accountability and incentive for performance. If we can tie together incentives with performance and say, in effect, to a school system that if it produces it will be rewarded—and do the same for teachers and even pupils—we can put new vitality into education. (p. 59)

Barro (1973) stated that in order to apply the concept of accountability in the field of education, the following
components should be present:

1. Measurable objectives in terms of output or student behavior, i.e., what the student should learn to do.

2. A learning program, designed to lead to the achievement of these measurable behavioral objectives.

3. The evaluation of the achievement of these objectives.

4. A determination of how successfully the objectives are realized.

5. An assessment of the cost of the program.

6. A systematic method of feedback to the ultimate decision makers, or those to whom the school personnel are accountable.

7. Appropriate revisions in the program to further improve student performance.

The innovative element of teacher effectiveness is the combining of all these tasks into a unified whole or into what may be called a "systems approach" to education.

In most formal organizations, including public schools and colleges, the superordinate has to evaluate the performance of subordinates. Getzels, Lipham, and Campbell (1968) stated that few issues in education are more explosive than the evaluation of teaching and teachers. Mitzel (Mathis, Cotton, & Sechrest, 1970) stated that classroom teaching would be easier to define if it could be conceptualized as either effective teaching or ineffective teaching. In a concise and perceptive review of approaches by Lucio (1973), it was pointed out that inadequate criteria, or standards,
used in evaluating teachers' behaviors are responsible for the lack of success in defining effective teaching behaviorally. Mitzel (Mathis et al., 1970) stated that teacher effectiveness has no meaning other than the criteria used to define successful teaching.

In most instances in the past, principal ratings have been the sole criterion for evaluating teachers. However, principal ratings constitute only one technique for evaluating teacher performance. Thus, it is also important to have student ratings of teachers plus student achievement as criteria for giving feedback to principals for making evaluative decisions pertaining to teacher performance. By using student ratings of teachers, principal ratings of teachers, and student achievement, a principal will have three components to evaluate teacher performance. Feedback will help make evaluation of a teacher's performance more comprehensible than in the past. The feedback from these components will serve as a management information system for principals to use as a basis for making evaluative decisions.

One might argue that educators have always been concerned with identifying and perpetuating good teaching. Nevertheless, the previous citations serve to stress the urgency of this task. There is indeed a need to understand better the components of effective teaching and how they may be assessed if educators are to respond to the concept
of accountability. Identification of these components has implications for two areas in education: (1) certifying, selecting, retaining, and promoting teachers; and (2) providing pertinent information for improving the education and in-service training of teachers.

The evaluation of teacher effectiveness is a difficult task. Hundreds of school districts have found the problem of rating an insurmountable task. Teacher organizations generally resist rating techniques, whereas many boards of education insist that performance of school personnel be evaluated and that performance be related to salaries. Burner (1958) stated that if an organization is to become effective, evaluation of present performances of personnel must be attempted, and, in this process, some examination of the teacher performance of its members seems inevitable.

At the 35th Annual National School Board Association Convention, Coats (1975) made the following comment about evaluation and accountability:

Evaluation in our school system (Kalamazoo Public Schools) is tied closely to the concept of accountability which is a label for indicating that we as a school system are assuming some responsibility for certain components of child growth. The responsibility which our Board has assumed is the development of basic skills and understandings such that graduates of our school system will be able to compete for higher jobs and for higher education. Basic to this goal is the successful teaching of reading, of how to express oneself in the English language and understanding of mathematical concepts and expressions. Beyond these basics we strive for academic excellence and career preparation. Comprehensive evaluation systems are essential if we are
to know and to improve the extent to which our school system is meeting this responsibility. (p. 2)

Research on predictive measures of teacher effectiveness has been pursued for many decades, yet the numerous efforts to identify the characteristics of an effective teacher or to describe and measure effective teaching performance generally led to inconclusive results (Lucio, 1973).

Ryans (1960) concluded that one major difficulty in predicting teacher effectiveness has been the lack of a solution to the criterion problem. Three sources (Barr, 1948; Domas & Fiedman, 1950; Marsh & Wilder, 1954) have summarized many of the studies that deal with characteristics of effective teachers. Barr (1948) published a summary of 153 studies concerned with the measurement and prediction of teacher effectiveness. Domas and Fiedman (1950) reprinted a bibliography of 1,006 publications dealing with the effectiveness of teachers. Marsh and Wilder (1954) reviewed 360 quantitative studies that dealt with the identification of effective instructors. The previous studies mentioned indicated that teacher effectiveness is multidimensional and, therefore, should be defined in accordance with supportable judgment. If satisfactory responses are to be made to the demands of accountability, there is a need to measure the performance of the instructional staff (teachers) in a multidimensional manner. Such information or feedback from various dimensions in which teachers perform should prove valuable to those
school systems using ratings in a comprehensive evaluation program of teacher effectiveness.

For many years, educators have assumed that teachers who were effective in one dimension were also effective in other dimensions. However, a teacher judged effective by one criterion may not necessarily demonstrate effectiveness using other criteria. Interpreting effectiveness as a unitary trait is an oversimplified and somewhat naive conceptualization of the problem. If success in teaching is multidimensional, it should be studied in terms of the interaction and interrelatedness of all its factors.

This study will utilize three components for measuring teacher performance: (1) student ratings of teachers, (2) principal ratings of teachers, and (3) student achievement.

The Problem

The purpose of the study is to determine the relationships among student ratings of teachers, principal ratings of teachers, and student achievement in reading and mathematics.

The specific objectives of the study are to answer the following questions:

1. What is the relationship between student achievement (reading) and student ratings of teachers?
2. What is the relationship between student achievement (mathematics) and student ratings of teachers?

3. What is the relationship between student achievement (reading) and principal ratings of teachers?

4. What is the relationship between student achievement (mathematics) and principal ratings of teachers?

5. What is the relationship between student achievement and educational backgrounds of teachers?

6. What is the relationship between student ratings of teachers and principal ratings of teachers?

Hypotheses

The hypotheses presented below comprise the 13 relationships investigated in the study:

$H_1$: There is a positive relationship between student achievement (reading) and student ratings of teachers.

$H_2$: There is a positive relationship between student achievement (mathematics) and student ratings of teachers.

$H_3$: There is a positive relationship between student achievement (reading) and principal ratings of teachers.

$H_4$: There is a positive relationship between student achievement (mathematics) and principal ratings of teachers.

$H_5$: There is a positive relationship between student ratings of teachers and principal ratings of teachers.

$H_6$: There is no significant difference in student achievement (reading) when comparing the educational backgrounds of teachers.
\( H_7: \) There is no significant difference in student achievement (mathematics) when comparing the educational backgrounds of teachers.

\( H_8: \) There is no significant difference in student ratings of teachers when comparing the educational backgrounds of teachers.

\( H_9: \) There is no significant difference in principal ratings of teachers when comparing the educational backgrounds of teachers.

\( H_{10}: \) There is no significant difference between student achievement (reading) in grades 4, 5, and 6.

\( H_{11}: \) There is no significant difference between student achievement (mathematics) in grades 4, 5, and 6.

\( H_{12}: \) There is no significant difference between student ratings of teachers in grades 4, 5, and 6.

\( H_{13}: \) There is no significant difference between principal ratings of teachers in grades 4, 5, and 6.

Overview

This investigation consists of examining three components used to evaluate teacher effectiveness: (1) student ratings of teachers, (2) principal ratings of teachers, and (3) student achievement.

The instruments used to gather data for this study were (1) the Elementary Teacher Image Questionnaire (ETIQ), to assess student ratings of teachers; (2) the Teacher Evaluation Instrument (TEI), to assess principal ratings.
of teachers; and (3) the Metropolitan Achievement Test (MAT), to assess students' actual gains in reading and mathematics. The instruments were administered to students enrolled in the Kalamazoo Public Schools during the 1973-74 school year. The sample population consisted of students in grades 4, 5, and 6. According to the studies by McMillan (1967) and Booth (1972), the choice of these grades would control to some degree the variability of students' responses because later elementary students (grades 4, 5, and 6) are less affected by like or dislike of the teacher.

Limitations of the Study

This study was limited to a school system having a biracial community, which is reflected in the schools' racial composition of approximately 20 percent minority students. At the present time, the Kalamazoo Public Schools are in the third year of a court-ordered two-way busing desegregation plan.

This study was also limited to a student population in grades 4, 5, and 6. Selection of these grades suggested limitations due to the closeness in range of student ages and grade levels. Selection of grades other than those used in this study would have imposed problems in controlling intervening variables such as curriculum and daily program of activities. The writer believed that the student population in grades 4, 5, and 6 would encounter
similar curriculum content and daily program experiences. The teacher population used in this study included all teachers in grades 4, 5, and 6. No students were interviewed since all data were obtained from records kept in the central office. These data covered the 1973-74 school session.

One major limitation was that data were gathered from a field study. There has been no experimental manipulation of variables. Such an investigation limits the possibility of making cause-and-effect inferences.

Definitions of Terms

The following definitions are given to clarify terms used and to specify the scope of the study:

1. Student ratings of a teacher
   a. Theoretical.—The students' perceptions of a teacher's overall performance in relation to selected characteristics.
   b. Operational.—The students' perceptions of a teacher's overall performance as measured by the ETIQ.

2. Principal ratings of a teacher
   a. Theoretical.—The principal's perceptions of a teacher's performance in relation to selected evaluation measures.
   b. Operational.—The principal's perceptions of a teacher's performance as measured by the TEI.
3. Student achievement
   a. Theoretical.—A measure of student learning of basic skills in the areas of reading and mathematics.
   b. Operational.—A measure of student learning of basic skills in reading and mathematics as measured by the MAT.

4. Teacher performance
   a. Theoretical.—The effectiveness of a teacher in a teacher/learning situation.
   b. Operational.—The effectiveness of a teacher in a teacher/learning situation as measured by three criteria: the ETIQ, the TEI, and the MAT.

Organization of the Study

Chapter I provided introduction to the study, basic considerations, limitations of the study, and definitions of terms pertinent to the study.

Chapter II contains a conceptual framework for the assessment of teacher effectiveness, identifying previous studies related to the assessment of teacher performance, student ratings of teachers, and principal ratings of teachers.

Chapter III consists of a description of the research design, procedures, instrumentation, sample and population, and method of data analysis.

Chapter IV reports the results of analyses related to the questions and hypotheses investigated in this study.
Chapter V involves a review and summary of the investigation and presents conclusions and recommendations for future research.
CHAPTER II

RATIONALE AND RELATED LITERATURE

Research and literature reviewed were pertinent to the following: (1) developing a conceptual framework for the assessment of teacher effectiveness, (2) identifying previous studies related to the assessment of teacher performance, (3) student ratings of teachers, and (4) principal ratings of teachers.

Conceptual Framework

Research on predictive measures of teacher effectiveness has been pursued for many decades, yet the numerous efforts to identify the characteristics of an effective teacher, or to describe and measure effective teaching performance, generally have led to inconclusive results (Lucio, 1973).

One major difficulty in predicting teacher effectiveness has been the lack of a solution to the criterion problem (Ryans, 1960). Traditionally, two general types of criteria have been employed: (1) process criteria and (2) teacher characteristics. Process criteria of teacher effectiveness are usually assessed against some standard performance or particular overt teaching acts generally assumed or inferred to relate to teaching. The assumption for teacher effectiveness is that if a teacher performs specified acts, pupil
behavior and teacher effectiveness can be predicted. Teacher characteristics of teacher effectiveness are defined as various personal characteristics such as intelligence, years and quality of schooling, personality traits, aptitudes, and other personal attributes of the teacher assumed to relate to, or predict, teaching performance. The assumption for teacher effectiveness is that if a teacher possesses particular personal traits, learner growth and teacher effectiveness can be predicted (Lucio & McNeil, 1969).

Dyer (1970) developed a model that portrays how administrators can measure the performance of educational systems for improving and maintaining the effectiveness of schools. Dyer (1973) further explained his model by stating:

An educational system is in essence a vast communication network in which every individual, whatever his role, is plugged in at some point to get the information he needs to do his own work properly and to supply others with the information they need to do theirs. The quantity of information that should somehow work its way through the system and into the world outside is larger and broader than most people, concerned with their own immediate tasks, generally realize. Its scope can be visualized by considering the informational components of what I have called the "Student Change Model of an Educational System." (p. 27)

This model (see Figure 1) is built around three major concepts. The first concept is that an educational system is a dynamic system consisting of a series of time segments through which students pass as they go through school. A given segment may be any length appropriate to the information needs of anyone inside or outside the system—a week,
Figure 1. Student change model of an educational system.

The second concept is that the central concern of the system is with the changes that take place in students as they go through any segment, however short or long. In what ways is a student different at the end of grade 6, for instance, from what he was at the beginning of grade 4? Or how does his command of arithmetic this week compare with what it was last week?

The third concept is that if the system is to operate effectively in the interest of student development, there must be available a wealth of information about (1) the changes in students that take place in each time period and (2) the many factors inside and outside of school that may
be affecting these changes for better or worse.

Such a conceptual framework was provided by Biddle and Ellena (1964), who suggested that the cause-and-effect relationships between classes of variables be studied in a social system (see Figure 2). Biddle and Ellena further explained their model by stating:

Each of these variables is postulated to affect one immediately behind it in sequence. In addition, the classroom situations (which are somewhat under teacher control) and school and community contexts (which are not) jointly constrain and interact with the linear effects of the five main sequence variables listed. (p. 14)

This model (Figure 2) is unique for its presentation of sequential cause-and-effect relationships of teacher characteristics and its recognition of the teaching context. The central problem in understanding teacher effectiveness is establishing relationships between teacher behaviors and teacher effects. However, it is certain that teacher competence involves a complex interaction between teacher properties and contextual factors in the community, school, and classroom.

A conceptual framework was provided by Getzels and Guba (1957) when they suggested that classroom situations be studied as a social system. The social system was conceived as involving two classes of phenomena that are conceptually independent and phenomenally interactive. One class constitutes the normative (or nomothetic) dimension of behavior, and the other, the personal (or idiographic) dimension of
Figure 2. A seven-variable class model for teacher effectiveness.
behavior (see Figure 3). If effective teaching is a part of a social process, then it is important to have interaction between the nomothetic and idiographic dimensions.

**Normative (Nomothetic) Dimension**

Social System

Social Behavior

Institution → Role → Expectation

Need-

Individual → Personality → Disposition

**Personal (Idiographic) Dimension**

Figure 3. Nomothetic and idiographic dimensions of behavior.

A practical means of assessing teacher effectiveness that allows for measuring different dimensions of a teacher's performance is based on the rationale that teaching is multidimensional and should be studied in terms of the interaction and interrelatedness of all its factors. While measuring student perception of teacher characteristics is important, teacher effectiveness cannot be based on this one criterion. Berelson and Steiner (1964) indicated that the perceiver organizes stimuli of his senses on the basis of his past experiences and motives. Students' perceptions of teacher characteristics are influenced by past experiences and need-dispositions. A composite of a class's perceptions of a teacher reflects such experiences and motives. Estvan and Estvan (1959) described the nature of social perception in the following manner:
The nature of social perception has regarded individuals as adapting or conforming to "social norms," the latter constituting a frame of reference which was brought into play in the perceptive act. In this sense, perception merges with cognition and it is regarded as the apprehension or understanding of the complex inter-play of social forces constituting the social situation in which the perceiver finds himself. (pp. 10-11)

Student ratings measure only one aspect of teacher effectiveness; when student ratings are combined with principal ratings and student achievement, however, teacher performance may be measured using three factors which individually may contain considerable bias but collectively constitute an objective, comprehensive, and detailed assessment of teacher effectiveness. In order to have a complete picture of teacher effectiveness, it is important to have measures in the cognitive and affective domains (Bartz, 1974). These criteria may be used to provide a total picture of effectiveness as they relate to teacher performance.

Related Studies of Teacher Effectiveness

During the period between 1900 and 1960, the focus of investigators was generally on various aspects of teacher behavior, and the study of the outcomes of instruction (pupil achievement) as a correlate of teacher effectiveness was neglected (Lucio, 1973). Investigators (Barr, Worcester, Abell, Beecher, Jensen, Peronto, Ringness, & Schmid, 1961; Eysenck, 1953; Howsam, 1960; Kleinman, 1966; Marsh & Wilder, 1954; Popham, 1971; Ryans, 1960) found that administrators
and supervisors most often judged teacher competence on the basis of (1) teaching ability, (2) disciplinary ability, (3) scholarship, and (4) personality, and demonstrated in their studies that commonly employed teacher-rating instruments and raters' assessments were unreliable.

According to Jenkins (1960):

In our society the authority to direct the learning activities of the student is given to the teacher. Both the teacher and the students expect the teacher to take charge, to initiate learning activities, and to contribute information as needed in the learning process. What the teacher does with his power makes a great deal of difference. (p. 164)

The earliest systematic studies of spontaneous student and teacher behavior were those of Anderson (1939), Brewer (1945), and Brewer and Reed (1946). These studies were based on the observation of "dominative" and "integrative" teacher behavior.

The qualitative differences that were determined between an integrative and a dominative social contact by these researchers established distinctions that have been followed in general ways by most of the research on teacher behavior since that done by Anderson (1939), who stated:

A preliminary study showed that it was possible to devise reliable measures of behavior of young children. Behavior was recorded as contacts and divided into categories. If a child snatched a toy, struck a playmate, or commanded him, or if he attempted to force him in some way, such contacts were included under the term "domination." By such behavior he ignored the rights of the companion; he tended to reduce the free interplay
of differences and to lead toward resistance or conformity in responding or adapting to another.

Other contacts were recorded which tended to increase the interplay of differences. Offering a companion a choice or soliciting an expression of his desires were gestures of flexibility and adaptation. These tended in the direction of discovering common purposes among differences. Such contacts were grouped under the term "socially integrative behavior." (p. 75)

The findings of Anderson (1939), Brewer (1945), and Brewer and Reed (1946), in the study of pre-, primary-, and elementary-school classrooms involving five different teachers and extending over several years, when taken together produced a series of significant findings:

1. The dominative and integrative contacts of the teacher set a pattern of behavior that spreads throughout the classroom; the behavior of the teacher, more than any other individual, sets the climate of the class. The conclusion is that when either type of contact predominates, domination stimulates further integration. It is the teacher's principle behavior pattern that spreads among pupils and is taken over by them even when the teacher is no longer in the room. Furthermore, the pattern a teacher develops in one year is likely to be continued by him the following year with different students.

2. When a teacher establishes a higher proportion of integrative contacts, students show more spontaneity and initiative, voluntary social contributions, and contributions to problem-solving.

3. When a teacher has a higher proportion of dominative contacts, the students are more easily distracted from school work and show greater compliance to, as well as rejection of, teacher domination.

Lippitt and White (1943) carried out laboratory...
experiments to analyze the effects of adult leaders' influence on boys' groups. The laboratory approach used had certain advantages in studying the effects of the adult leaders' behavior: (1) the contrasting patterns of leader behavior were clearly defined in advance and were made more consistent as a result of training and role-playing; (2) differences in underlying personality and appearance of the adult leaders were controlled through role rotation; and (3) the effect of the pattern of leader behavior was intensified (when compared with a classroom), since there were only five boys to a group.

Flanders (1951) created laboratory situations in which one student at a time was exposed to contrasting patterns of teacher behavior. A sustained dominative pattern was consistently disliked by students, reduced their ability to recall the material studied, and produced disruptive anxiety as indicated by galvanic skin response and changes in heartbeat rates. The opposite trends were noted in student reactions to integrative contacts.

Perkins (1951), using Withall's technique, studied groups of teachers organized to discuss the topic of child growth and development. Perkins found that greater learning about child growth and development occurred when group discussion was free to focus on that topic; groups with an integrative type of leader were able to do this more frequently than were groups led by a dominative type of leader.

The study by Mitzel and Rabinowitz (1953) supplies
evidence that teachers are flexible in their use of integrative and dominative contacts. Four teachers were studied using Withall's technique, and their data were organized to permit analysis of variation among teachers, visits, and observers. The median length of an observer's visit was 20 minutes; statistically significant, wide variability of the integrative-dominative balance among visits for the same teacher suggests that teachers adapt their influence to the immediate situation.

Since about 1960, following the lead of industrial organizations which have made use of performance objectives as the basis for judging personnel effectiveness (Adams, 1959; Flanagan & Burns, 1955; Meyers, Kay, & French, 1965), various educational researchers, in attempting to seek solutions to the criterion problem, have shifted from studying primarily what the teacher does (means of instruction) to examining changes in learner behavior as a result of instruction (outcomes of instruction). Change in student behavior thereby becomes the criterion for evaluating teacher performance (Hastings, 1952; Marsh, Burgess, & Smith, 1958; McNeil, 1966; Popham, 1971).

In answer to arguments against the merit of using performance criterion measures in teacher evaluation (referring to studies before the 1960's), Justiz (1968) designed a study to test the validity of nine identified criticisms. He interpreted his results as having invalidated most of
the criticisms and as presenting evidence that, in general, teaching ability of teachers could be measured reliably in terms of student achievement. His study also found a relationship between student teacher attitude (as measured by the Minnesota Teacher Attitude Inventory) and student achievement.

McNeil (1967) conducted an experimental study with 72 secondary-school student teachers which, based on the concept of supervision by objectives, revealed that a large number of student teachers supervised by objectives were perceived by their supervisors as achieving greater success as measured by student performance than those supervised by a rating scheme. The student teachers who contracted with their supervisors to achieve explicit changes in student performance also were perceived as being more successful in applying the principles of learning.

A second investigation by McNeil (1967), using 44 elementary student teachers as subjects, revealed that students of student teachers whose success was measured by their students' attainment of agreed-upon objectives achieved significantly better in punctuation skills than students whose teachers were assessed not by student achievement criteria, but by the usual rating scale. These same student teachers completed a questionnaire to measure their perceptions of the supervisor and the methods of evaluation, and were almost unanimous in their preference for the student progress
criterion as the measure for evaluation of their teaching. Both groups tended to report the same amount of preparation time given to the punctuation lessons, felt free to select their own teaching procedures, and found their supervisors' suggestions helpful.

Moffett (1966) conducted a study in which he compared the performance of student teachers evaluated on the basis of attaining agreed-upon instructional objectives with that of teachers evaluated by means of rating instruments. Data were obtained on the extent of students' attainment of instructional objectives in geography skills and their attitudes toward subject matter, and on teachers' attitudes toward supervisory help, satisfaction with midterm grades, and preferences for types of performance rating. Secondary school students in grades 7-12, selected by random sampling technique, served as subjects. After pretesting students in geography skills, 36 student teachers were randomly assigned to either an experimental or a control group. Teachers in the experimental group executed preinstructional contacts with their supervisors based on the instructional objectives to be achieved as revealed by the pretests, and their teaching performance was evaluated in terms of student achievement. The control group teachers, while informed about students' pretest deficiencies and the need to correct them, did not enter into an instructional contract; their performance was evaluated by means of rating scale measures.
Findings were reported by Moffett (1966) as follows: (1) the students in the experimental group performed significantly better on a posttest of geography skills than did students of teachers in the control group; (2) teachers in the experimental group expressed more confidence in supervisory help and reported satisfaction with their midterm grades significantly more than the control group; and (3) of all teachers, regardless of whether they were in the experimental group or control group, 94 percent expressed preference for having their teaching performance evaluated on student achievement as a result of instruction rather than on a rating scale measure.

Smithman (1970), in a study employing the strategy of supervision by objectives, attempted to (1) determine the extent to which agreement on instructional objectives stated in behavioral terms by supervisor and teacher prior to teaching would increase student performance, and (2) discover whether the process of supervision by objectives resulted in evaluations of teaching more germane to instructional performance. The study was predicated on two assumptions suggested by Lucio and McNeil (1969): (1) learning is evidenced by a change in behavior, and (2) teaching is successful only when the instructor's predetermined and intentional changes sought in the learner actually occur.

Through use of a stratified random sampling technique in 20 classroom units in 9 Canadian schools, 20 certified
teachers and 558 students in their sixth year of school were selected as the sample. Together with their principals, the 10 teachers assigned to the experimental group selected the criteria which were to be accepted as evidence that learners had reached the desired behavior, and they were evaluated accordingly. The control group of 10 teachers, employing the same instructional objectives as the teachers in the experimental group but without executing an instructional agreement with their principals, were evaluated by means of the school district rating scale. Students in both the experimental and control groups were administered pretests and posttests of mathematical skills.

Based on results of the posttest, the experimental group was favored; it was concluded that students whose teachers were evaluated by objectives outperformed those students whose teachers were evaluated on a rating scale. On five measures of teacher attitude toward evaluation as measured by a posttest attitude questionnaire, significant differences were not found between the experimental and control groups. The amount of individual instruction provided the two groups was approximately equal.

Popham (1971) developed a teaching performance test. It was an attempt to isolate a readily usable indicator to assess a teacher's instructional skills by measuring students' attainment of instructional objectives. Popham's approach to assessing instructional performance was predicated
on the assumption that the main reason for a teacher's existence is to promote beneficial changes in the learner. He constructed his tests in the fields of social science, electronics, and auto mechanics and used them to compare the performances of credentialed and experienced teachers with those of persons neither credentialed nor experienced. Each of the three teaching performance tests consisted of a set of specific instructional objectives measured by a posttest. The number of objectives and posttest items varied with the particular subject fields. All three tests contained sets of resource materials which could be used by the teacher in planning the instructional sequence to accomplish the objectives.

Popham (1971) wrote that "there appear to be no readily available [methodological] loopholes by which we can explain away the non-significant outcomes. A more straightforward explanation is available. Experienced teachers are not particularly skilled at bringing about specified behavior changes in the learner" (p. 601).

Rosenshine (1970) compiled the results of several studies which used methods like those of the Popham-McNeil-Millman Method, of the short-term stability of teacher effect on change in student behavior: 20 stability coefficients from 5 separate studies were reported. The 20 correlation coefficients, correlating teacher effect on 2 separate (but closely related) topics on 2 distinct (but randomly equiva-
lent) student groups, ranged from a low of .45 to a high of .87; only 2 of the coefficients are significantly different from 0 at the .05 level of confidence (two-tailed test).

Rosenshine (1970) concluded that when teachers taught different topics to different students, the direction of the correlations was erratic, and few correlations were significant.

In a study of teaching behaviors, Barr (1929) concluded that instability of teacher behavior from one lesson to another is the dominant source of unreliability in the effect of teachers' actions on student learning. The contemporary studies reviewed above prove Barr's conclusion beyond reasonable doubt.

Over all, the findings of early investigators provided relatively little evidence to demonstrate that particular teaching acts or teacher characteristics, assumed to relate to teacher effectiveness, were associated consistently with learner achievement. More recently, Cohen and Brawer (1969), McNeil (1967, 1971), and Popham (1971) have shown that by specifying desired changes in learners, arranging instructional events to produce the changes, and assessing the learners' attainment of instructional objectives, selective indexes of teacher performance (based on student achievement) can be derived.

According to Lucio (1973), systematic efforts in the direction of analyzing teacher performance as a correlate
of predicted changes in learners might be expected:

1. Establishing appropriate criteria for assessment of teacher performance.
2. Improving teachers' skills in defining and achieving instructional objectives.
4. Defining better the degrees of accountability for school personnel in accomplishing the goals of schooling.
5. Providing evidence for the public that schools are achieving stated objectives.

Kleinman's (1966) discussion of Ellena's (1961) summary of teacher effectiveness yielded a number of important results:

1. There is only a low correlation between measures of on-the-job performance of teachers and earlier scholarship.
2. There is no evidence that married teachers are in any way inferior to unmarried teachers.
3. There is some evidence that more professional knowledge (National Teacher Examination scores) tends to be associated with more effective teaching.
4. No particular differences in effectiveness between men and women teachers have been found.
5. Teachers' rated effectiveness at first increases rather rapidly with experience and then levels off at five or beyond.

In a comprehensive review of the literature, Yamamoto (1964) concluded: (1) there is no single, simple pattern of characteristics of the "successful" teacher; (2) supervisor ratings of teacher effectiveness quite often do not
agree with student or colleague ratings; and (3) there is some indication that those teachers who are well adjusted do not know these principles mentioned above.

Teacher competence has been a matter of concern in all efforts to improve education. Before the turn of the century, studies were being conducted in this country in an attempt to isolate the factors which contributed significantly to teacher effectiveness. No one criterion is complete, and a preference for one as opposed to another involves value judgment on the part of the assessor. In the studies reviewed, the criteria used for judging teacher effectiveness differed depending on their relevance to the research problem.

Overview of Student Ratings of Teachers

Student ratings, as bases for differentiating teacher effectiveness, have been used increasingly for the past 35 years; it was as early as 1896 that researchers capitalized on student reactions to evaluate teachers and their performance. Ratings have been made by students from intermediate grade levels through secondary school, college, and graduate school levels. Early skepticism has given way to widespread confidence in the ability of young consumers of education to make reliable and valid judgments of teaching performance.

Probably the most common technique used in research on teacher effectiveness is the rating form (Biddle & Ellena,
In a typical application, one is asked to rate the teacher's ability, the teacher's performance, or student response to the teacher. Ratings have been used as criteria for evaluating teacher competence or predicting to the "outcome" variables in the classroom—which then have been measured objectively. Multitudes of persons have been used as raters in competence research.

Remmers (1963), in a review of student ratings, stated:

In a sense the study of raters and the rating process is equivalent to the study of accuracy of social perception, or interpersonal perception, or of empathy, or of social sensitivity, as it has been variously termed. (p. 333)

As noted by Walberg and Anderson (1968) and Anderson, Walberg, and Welch (1969), Remmers (1934) stated that students' responses on questionnaires and inventories have been used as a measure of perception of environment from the inhabitants of typical environments. Such responses have been useful for research on teacher performance. Perhaps to the extent that the affective domain is viewed as important to the teaching-learning process, student ratings would appear to have value.

Paraschevopoulos' (1968) rationale for student ratings stated:

Student ratings allow us to see how students perceive and interpret the behavior of teachers. This subjective perception, more than the independently and objectively assessed behaviors by trained observers, supervisors, and other "outsiders," determines essentially the interpersonal relationships in the classroom and colors
its social and emotional climate. The atmosphere of interpersonal relationships is crucially important in its effects on the child's learning and adjustment. (p. 25)

Barr et al. (1961) stated that the nature of human abilities cannot be adequately defined without some reference to the mind-body relationship. They assumed that mind and body are one and discussed teaching effectiveness, chiefly in terms of (1) operations, processes, and behaviors; (2) conditions (internal and external) considered essential to easy, smooth, and efficient operations; and (3) end products (outcomes and results) that follow when forcefulness, forthrightness, or cooperation are used—they will, in general, be used to describe behavior rather than innate qualities or traits. The conditions, both external and internal, under which teaching takes place will be treated as far as possible on a factual basis and in a manner so as to include both psychophysical processes and socio-physical environment factors and their many interrelationships.

Flesher (1952) suggested that student ratings, within the limits of their reliability, are valid measures of student opinion of instructors. In evaluating reliability of student ratings of teachers, Flesher reported that in three studies (Remmers & Brandenburg, 1927; Root, 1931; Smeltzer & Harter, 1934) the reliability coefficients showed the consistency with which the same students rated a particular instructor, using either the same or different rating devices.
Remmers (1960), in extensive research with the Purdue Rating Scale for Instructors and its revision, lent support to Flesher's (1952) findings. Remmers stated that the reliability of ratings of teachers by students is a function of the number of raters, in accordance with the Spearman-Brown formula. If 25 or more student ratings are averaged, they are as reliable as the better educational and mental tests presently available.

Goldberg (1968) expressed that the validity of using student observation as an approach for determining differential student reaction to teacher and classroom activities stems from the fact that students observe more of a teacher's typical behavior than is usually available to the outside observer.

Ryan (1966) studied the "halo effect" in the using of student ratings. Ryan noted that a reliable instrument that rates teachers' skills on a scale would restrict the influence of any "halo effect." Ryan felt that Bryan's Student Opinion Questionnaire, which was developed over a 30-year period, was the most reliable and useful instrument that existed for measuring students' opinions of their teachers.

Major objections voiced relative to the use of such ratings are included in a list compiled by Amatora (1954):

1. Students are too immature to give valid judgments.

2. The halo effect offsets any possible validity.
3. Student responses are influenced by grades, teacher attitudes, and the like.

4. Teacher morale is lowered when students are permitted to do the ratings.

Remmers (1963) identified the following properties as necessary for rating scales to be used in research on teaching: (1) objectivity, (2) reliability, (3) validity, (4) sensitivity, and (5) utility. Remmers reported that correlation studies have shown little relationship between such factors as student grades, student interest in the subject, general attitude of the student toward school, amount of work required by the teacher, and student ratings of teacher performance.

Wilson (1971) reported that Bryan's Student Opinion Questionnaire has undergone constant modification, but as far back as 1939 it was shown to have reliability approximating .85 for a class of 15 to 20 high school students. Amatora's Diagnostic Teacher-Rating Scale and the Purdue Rating Scale possess reliability ratings for upper elementary and college-age students comparable to those obtained on the Student Opinion Questionnaire.

A recent study by the Educator Feedback Center (1969) at Western Michigan University supports the notion that students do honestly and reliably respond to student rating devices. Personnel from the center conducted in-depth interviews with 75 students to determine the extent of agreement between the students' written rating of the teacher on the
Teacher Image Questionnaire (TIQ) and the students' oral appraisal of the teacher. Questions on the TIQ served as the structure for the oral interview. The sample was comprised of 5 students randomly selected from each of the 15 classrooms in the study.

Analysis of correlations for each item showed written and oral responses of the same individuals to be markedly similar. Item 4, "classroom control," obtained the lowest reliability coefficient at the .62 level. Reliability for item 10, "sense of humor," was the highest at the .92 level. The correlation between individual written and oral responses for an average of all items was .95. These findings would appear to add support to the validity of student rating instruments.

Marsh and Wilder (1954) pointed out that fairly good reliability could be obtained regarding supervisor judgment, but that validity could not be demonstrated. Validity of such instruments seems to be inherent in their design. Significance of validity is not with the teacher characteristics actually possessed by the teacher, but with the student's perception of their presence or absence. To the degree that the student honestly and reliably responds to items, such instruments are also valid.

Coats (1970) found a mean intercorrelation of .57 for the Teacher Image Questionnaire (TIQ). In a paper presented on teacher effectiveness, he gave a plausible explanation of the halo effect:
Students do not respond directly to specific questions regarding teacher effectiveness. Rather, a kind of halo effect based on teacher charisma or popularity determines to a large extent how students react to questions about their teacher. This is not to say that student ratings of teachers are not important or meaningful. Teacher charisma is probably a function of teacher effectiveness. Furthermore, as indicated above, at least 40% of the variance in student ratings of teachers is independent of the charismatic factor and probably represents fairly objective student judgments. (p. 8)

The study by Ryans (1960), on characteristics of teachers, is one of the more extensive studies of teachers. Three major areas were investigated in this study: (1) the identification and analysis of patterns of classroom behavior, attitudes, viewpoints, and intellectual and emotional qualities that characterize teachers; (2) the development of a paper-and-pencil instrument suitable for assessing patterns of classroom behavior and personal qualities of teachers; and (3) comparisons of various groups of teachers with respect to points 1 and 2. During the study, which took place over a 6-year period, 100 separate research projects were undertaken involving more than 6,000 teachers in 1,700 schools and about 450 school systems. In the study, trained investigators observed student-teacher behavior in classrooms in an effort to discover patterns of teacher behavior and concurrent student behavior. The participating teachers were surveyed concerning their activities, preferences, and attitudes. The following findings are pertinent to this study:

1. Three distinct patterns of teacher behavior that were measured by the Teacher Characteris-
tic Schedule were (a) warm versus aloof teacher behavior, (b) responsible versus slipshod teacher behavior, and (c) stimulating versus dull teacher behavior. These and lesser factors not mentioned were related in different degrees with student behavior, level of instruction, and scores on other standardized personality inventories such as the Minnesota Multiphasic Personality Inventory and the Thurstone Temperament Schedule.

2. Certain dimensions of teacher attitudes, verbal understanding, and educational viewpoints differed significantly for teachers who were stratified in the study according to age, experience, sex, marital status, and college achievement.

3. When comparing scores on the Teacher Characteristics Schedules, teachers with high (positive) scores differed in behavior from those with low (negative) scores as follows:

a. Teachers with high scores tended to:

1) be extremely generous in appraisal of the behavior and motives of other persons.

2) possess strong interests in reading and literary affairs.

3) be interested in music, painting, and arts in general.

4) participate in social groups.

5) enjoy student relationship.

6) prefer non-directed, or permissive, classroom procedures.

7) manifest superior verbal intelligence and evidence superior emotional adjustment.

b. Teachers with low scores tended to:

1) be restrictive and critical in their appraisals of other persons.
2) prefer activities that did not involve close personal contacts.

3) express less favorable opinions of students.

4) manifest less high verbal intelligence.

5) show less satisfactory emotional adjustment.

6) represent older age groups.

In a recent study, Wilson (1971) compared certain teachers, classrooms, and community characteristics with ratings given teachers by their students. The selected characteristics included (1) teacher variables of sex, age, experience, college degree, college major, marital status, and attitude toward the class; (2) classroom variables of subject matter, grade level, and class size; and (3) socioeconomic level. The subjects consisted of 1,180 teachers and 51,966 secondary-school students in 2,101 classes serviced during 1968-1970 by the Educator Feedback Center at Western Michigan University. Findings pertinent to this study are:

1. A positive relationship was found to exist between teachers' perceptions of their classes and students' ratings of teacher performance.

2. Teachers from suburban communities were rated more favorably by their students than teachers from rural and large urban communities. The analysis of socioeconomic status of the community as related to student ratings revealed similar results in that teachers who judged their community to be middle class were rated more favorably than teachers who judged their communities to be of low socioeconomic status.
3. Students in classes with 36 or more members rated their teachers less favorably than students in classes of smaller size.

4. Married teachers were rated more favorably than single teachers.

5. Teachers with Master's degrees were rated more favorably by their students than those teachers with Bachelor's degrees.

6. Teachers with graduate majors in sciences and mathematics were rated more favorably by their students than teachers with graduate majors in the fine arts, counseling, and guidance.

7. Teachers with 10-14 years of experience were rated more favorably than teachers who had less than 10 or more than 14 years of teaching experience.

8. Teachers in the 36- to 45-year age bracket were rated more favorably than teachers who were younger or older.

9. The variables of grade level taught, sex of teacher, and subject in which students were enrolled were found to be unrelated to student ratings.

Smith (1974) investigated classroom environmental variables related to students' perceptions of teachers and the effects of perception on academic growth. Racial composition of the classroom and grade level were factors treated to determine possible relationships to students' perceptions of teachers.

The sample consisted of 699 students, representing 30 classrooms of grades 4, 5, and 6. The investigation took place in a midwestern school system which satisfied the racial classroom composition criterion: 20 percent or more black students within each classroom. The criteria used to
select classrooms for the study were as follows: (1) classroom groupings were composed of a racial mixture—black students and white students—with significant percentage differences between classrooms; and (2) grades 4, 5, and 6 were represented for each of the classrooms. Grades 4, 5, and 6 were matched with each of the racial classroom groupings; 10 fourth-, 10 fifth-, and 10 sixth-grade groups were studied. Two classrooms from each grade level were representative of the five racial groupings: 90-100 percent black, 70-85 percent black, 50-60 percent black, 40-49 percent black, and 20-30 percent black. The classrooms were selected from schools possessing the above racial groupings.

The variables studied were students' perceptions of teachers as related to racial classroom composition and grade level (the major independent variables) and academic growth (the dependent variable). To obtain students' perceptions of teachers, the Elementary Student Opinion Questionnaire (ESOQ) was used. Student academic growth was ascertained by the following achievement tests: the Wide Range Achievement Test (WRAT), the Standard Achievement Test (SAT), and the Michigan Education Assessment Program (MEAP). The ESOQ was administered to the 30 classrooms by the writer during the first week in May, 1974. The achievement pre- and posttests were administered by the school system. The following findings are related to this study:

1. Students' perceptions of teachers correlated significantly with mathematics growth, but
not with reading growth when the 30 classrooms were analyzed.

2. Classrooms having student populations with approximately equal numbers of black and white students reported significantly higher mean scores for perception.

3. A significant difference was found between the mean scores for grades 4, 5, and 6 when students' perceptions of teachers were analyzed.

4. There were no significant differences between classroom mean scores for grades 4, 5, and 6 when students' perceptions of teachers were analyzed.

5. Racial classroom composition was a factor influencing students' perceptions of teachers.

6. Grade level and racial classroom composition had no significant interaction effect on students' perceptions of teachers. (Smith, 1974)

A study was conducted by Townsend (1972), to determine what relationships might exist between certain characteristics of junior-high and high-school science teachers and student ratings of these teachers.

The subjects for the study consisted of all general science, chemistry, physics, and biology teachers who voluntarily requested the services of the Student Reaction Center and Educator Feedback Center at Western Michigan University during 1961-1963 and 1968-1970. A total of 142 science teachers were rated by 8,633 students in 392 classes during 1961-1963. A total of 249 science teachers were rated by 9,105 students in 379 classes during 1968-1970. The non-science subjects for the years 1968-1970 included 934 teachers.
They were rated by 41,329 students from 1,680 classes. The teachers included in Townsend's (1972) study mainly were from the North Central states.

Student ratings of the teachers in this study were measured by the Student Opinion Questionnaire and the Teacher Image Questionnaire, developed for use in the Educator Feedback Center at Western Michigan University. The major dependent variables consisted of an average of the ratings of the teacher characteristics common to various forms of the above questionnaires. The data compiled from teacher responses to inquiries on the Class ID Form and the Teacher ID Form were the independent variables in the study.

The primary statistical treatment used in this study was the one-way analysis of variance. Coefficients of correlation and t tests were computed where deemed appropriate. The probability level, strength of association, and strength of determination were reported for interpretation of significance for each comparison. Results of the statistical analyses collected for this study are summarized as follows:

1. Science students from the 1961-1963 period rated the teacher characteristic of interest significantly higher than those from the 1968-1970 period. No significant differences were detected between student ratings of the two periods for the 5 other characteristics of knowledge of subject matter, clarity of presentation, fairness, control, and attitude toward students.

2. Of 16 teacher characteristics that were investigated, science teachers were rated significantly higher than non-science teachers.
on fairness, attitude toward student ideas, sense of humor, and self-control. Science teachers were also rated higher on 10 of the remaining 12 characteristics, although these differences were not considered significant according to the criterion for significance established in this study (p \leq .05 and E^2 \geq .015).

3. No significant differences were detected among the mean ratings of teacher characteristics of general science, biology, chemistry, and physics teachers.

4. Significant differences were detected between student ratings of science teachers and the teacher characteristics of age, college degree, teaching experience, teachers' perceptions of classes, and size of class taught. The specific natures of the relationships detected were these:

a. Science teachers from 26 to 45 years of age were perceived by their students to be more effective than younger or older teachers.

b. Science teachers with Master's degrees were rated more favorably by their students than those teachers whose highest degree was a Bachelor's degree.

c. Science teachers with more than 2 years of teaching experience were rated more favorably than teachers who had only 2 years, or less, of teaching experience.

d. A positive relationship was found to exist between teachers' perceptions of their classes and students' ratings of teacher effectiveness.

e. Students in classes with enrollments of 26 or more rated their teachers less favorably than students in smaller classes.

5. No significant differences were detected between student ratings of science teachers and the characteristics of teacher sex, marital status, type and socioeconomic status of the community in which the teacher taught, and the student sex ratio of science classes. (Townsend, 1972)
Proctor's (1972) study examined the manner in which role stress varies with student ratings, race of teacher, class racial balance, sex of teacher, socioeconomic attendance district of elementary school, and teaching grade level. The composition of the sample included 31 second-grade classrooms, 32 fourth-grade classrooms, 20 seventh-grade classrooms, 14 tenth-grade classrooms, and was drawn from 34 of the 36 public schools in Kalamazoo, Michigan. The subjects responding from this sample included 77 classrooms (or 79 percent of the classrooms), which involved 77 teachers and 2,023 students.

The major instruments used in this study were the Teacher Opinion Questionnaire (TOQ), to assess teacher role stress; and the Student Opinion Questionnaire (SOQ), to assess student perceptions of teachers. The teachers described what they thought were reasonable expectations and the actual situation of the job, the building principal, and students. The difference between the scores on these two dimensions comprised the role stress score. The major dependent variable, "role stress," was assessed on three distinct components: teaching as a job, the building principal, and students. The sum of the scores for the three components was labeled "overall stress." The major independent variable, "student ratings of teacher," was a composite measure of teacher effectiveness based on student responses. The probability of observing differences by
chance and an estimate of the strengths of association were reported. The following findings were related to this study:

1. Student ratings of teachers were found not to be related significantly to overall teacher role stress and teacher role stress on the job and principal components.
   a. There did seem to be a relationship between teacher role stress on the student component and student ratings of teachers.
   b. Teachers with low student stress scores were perceived more favorably by students than teachers with average and high role stress scores on the student component.

2. Racial balance of class was related to teacher role stress. Teachers with majority white classes indicated more overall, job, and student stress than teachers with majority black classes.
   a. Teachers with racially mixed classes were found to indicate more role stress than teachers with all white classes.

3. Sex of teacher was related to teacher role stress and teacher role stress on the job and student components.
   a. Male teachers indicated more role stress than female teachers.

4. Teaching grade level was related to teacher role stress on the student component.
   a. Teachers of grade 10 reported more stress and teachers of grade 2 less stress when compared with the combined groups of teachers of grades 4 and 7 on teacher role stress on the student component. (Proctor, 1972)
Overview of Principal Ratings of Teachers

Rating of a teacher has become more than just an evaluation of the teaching and learning process. Ratings have been used to place and dismiss teachers and to set salary schedules as well as to provide quality education. Merit rating is now based on teacher evaluation. Crossley (1957) spelled out the fact that rating forms have been devised to measure the teacher's relationship with the community, with the other teachers, and with the parents of pupils; his cooperation with his colleagues; his desire to grow in the profession; his appearance; his personality; and many other factors of concern to the principal when rating. An accurate rating has demanded visitation and observations.

Coats (1974) expounded on a method used in Kalamazoo, Michigan, to judge staff on performance. His explanation for the model used in Kalamazoo was that "extra effort by school administrators puts extra bucks in their pockets." At the beginning of each school year, administrators establish, with help and guidance from the assistant superintendents, performance objectives for the coming year. At intervals during the year, each administrator's performance and area of responsibility are evaluated and studied by the superintendent and his assistants, who have established umbrella-like performance goals for the entire school system. Specific objectives for administrators can be changed;
sometimes, if extenuating circumstances arise, they are relaxed or strengthened in midyear.

Stoops and Johnson (1967) have indicated that classroom visitation, for many years, was about the only supervisory practice. The responsibility of rating (or evaluation) was added to supervision later. Teacher evaluation, erroneously, has become synonymous with supervision in the minds of many teachers. The reason for this belief is that the greatest significance of the teacher is his work in the classroom. The value of the teacher could, therefore, be determined by observing his classroom work.

Stoops and Johnson (1967) further stated that some teachers hold the opinion that rating and rating cards destroy what is needed most between the principal and teacher, which is rapport. This viewpoint maintains that most of the items used for rating purposes can be judged with very little accuracy. It has been pointed out in studies by Crossley (1957) and Stoops and Rafferty (1961) that there is very little correlation between ratings on certain topics, which means that there is very little reliability in most rating forms. This view holds, for instance, that the principal cannot demand professional growth, cooperation, or loyalty; he must earn these personal qualities.

Brighton and Rose (1965) explained the principal's role as a chief administrative officer of a school: the principal wears several "hats." One which he tends to put on with
trepidation, unlike most laymen, is that of evaluator of the teaching performance of his professional staff. Principals often regard passing judgment on their professional colleagues—on their competence, traits, and effectiveness—and being expected to defend these judgments as tasks with many unpleasant aspects. However, as long as the principal is held accountable to his superintendent and to the board of education for the quality of the program in his school and for the administration of his staff, the job of evaluating teachers is one that cannot be avoided.

Biondolillo (1972) explained the principal's role for helping teachers improve themselves when he said:

> Education today places primary emphasis on the growth and development of the individual child. School leadership should place an emphasis on helping each teacher move toward instructional growth and improvement. Encouraging good instructional practice is a very subtle thing. A teacher's creativity and initiative can easily be hindered as a direct result of overassistance and supervision. On the other hand, inadequate help and a lack of encouragement may lead to discouragement and failure. Somehow, the principal must maintain a school environment which allows for personal initiative and at the same time fosters the development of a professional attitude toward and commitment to the improvement of teaching methods. Commitment is one of the essential elements of a teacher improvement program. It is a pledge everyone should make when entering the profession—to work for the increasing betterment of educational opportunities for all students. (p. 48)

It was further believed by Biondolillo that an effective and well-balanced teacher improvement program could be initiated and nurtured within a school system that respects the
individual differences of its teachers and students. Innovation will flourish within a school environment that provides opportunities for teachers to plan, evaluate, and change through a cooperative and democratic process.

Travers (1973) alluded to the fact that principals and supervisors sometimes use rating scales as tools for measuring teacher effectiveness. According to Travers, it is not uncommon to find such vaguely worded items on these scales as the following, under the heading "teaching techniques": planning and organizing appropriately methods and instructional skills, classroom control, awareness of individual needs, etc. Halo, lack of operational definitions, failure to control for sampling of teacher behavior, effect of observer on teacher performance—all such limitations make rating scales of doubtful worth in the hands of administrators and principals.

Principals visiting classrooms should not only rate the teacher, but should also point out aids that can be used to enhance teaching, pinpoint possible trouble areas, and strengthen the confidence of the teacher. The principal, as an infrequent classroom visitor, views the teaching situation from a perspective different from the teachers. From his vantage point, he can more easily spot areas that need improvement than might the teacher, who is too close to the problem. The ultimate outcome of visitation is the improvement of the learning situation (Ragan, 1966).
The purposes of classroom visitation were explained by Stoops and Johnson (1967) when they said the more the principal visits classrooms, the more he will observe and learn. The more he uses outside consultants to help teachers, the more prepared and informed he becomes. Supervision is a learning process on the part of the principal. The more the principal learns about teaching methods and educational techniques, the more he has to share with, and help, other teachers. Thus, supervision is a two-way process whereby the principal learns from teachers and shares his knowledge with teachers. The purposes are:

1. To learn and evaluate practices of teachers.
2. To discover good and promising qualities of teachers, special abilities.
3. To stimulate teachers to do their best. Teachers who are not observed feel neglected.
4. To discover the needs of teachers.
5. To gain information as a basis for improving the supervisory program of the school.
6. To determine the needs of pupils.
7. To gain firsthand information on how pupils in the school are responding to teachers.
8. To learn what administrative changes will help teaching procedures.
9. To determine the amount of preparation a teacher makes for a lesson.
10. To observe the curriculum in action. (Stoops & Johnson, 1967)

Research has shown that the average nonteaching principal visits each teacher's class three times during the year.
Crossley further stated that the principal should determine the length of each visit, which will vary with the purpose; the timing of the visit is also important, and the principal should determine the kind of visit—whether it is to be unannounced, announced, or invitational.

According to Otto (1954), the techniques of the observational visit may vary, but the pattern should meet the needs of the principal and the personnel of the school. The following may be adhered to:

1. The principal should have an awareness of the physical surroundings, the classroom situation, and the role of the teacher in the particular situation. Particular emphasis should be placed on the curriculum and the way it is presented.

2. Rapport should be established with the teacher and with the class. This is not accomplished through interruptions, but through the vitality of the interest that is shown.

3. As little disturbance as possible should be created in entering and leaving the classroom.

4. Notes that are to be used in the follow-up conference should be jotted down after leaving the classroom.

5. The reactions of the students should be observed.

6. The handling of routines should be analyzed.

7. The principal should stay long enough to determine the total picture of the situation.

(After Otto, 1954)

A checklist should be developed by the principal so that when he gets back to his office he can record his observa-
In the evaluation of a city school system, there is a great diversity in the priorities, procedures, and instruments used (Misner, Schneider, & Keith, 1963). The plan of evaluation in a specific large-city school system and a suburban school system should prove valuable as examples. Although there are many differences in the means of evaluation among school systems, there are also many common elements that may serve as sources of ideas and suggestions for planning programs.

One example is the Vega City School District, which was used in an evaluation program in 1960 by the American Association of School Administrators and National School Board Association (Misner et al., 1963). Vega City is the fictitious name for an actual city with a school system enrolling 123,429 students in kindergarten through grade 14. It is a rapidly growing community, and the school administration realizes that evaluation processes are factors which will determine whether inevitable changes are orderly, planned, and intelligent, or whether they will be frantic and aimless.

The Vega City School District depends upon a strong research department for the coordination of its evaluation program. The research staff includes a director of research, an administrative analyst, and a systems and procedures analyst, all drawn from varied fields. The pattern established for the district's evaluative procedure includes the
following steps: (1) defining the problem or center of concern, (2) detailing information needed, (3) reviewing the background, (4) analyzing possible problems involved, (5) developing questions to be answered, (6) collecting information, (7) analyzing information, (8) determining strengths and weaknesses, and (9) reporting findings.

The evaluation program at Vega City includes several different research approaches: (1) comparative studies, (2) studies of trends and projections, (3) problem-solving, (4) quality control studies, (5) opinion surveys, and (6) experimentation and analysis.

The pattern of evaluation is a clean one. Coordination of the program is attained through a well-staffed research department. The model used in Vega City is very similar to the model used by the Kalamazoo Public Schools. Experimentation is going on continuously, and each new experimental project has built into it the means for evaluating what will be done.

Also in 1960, the American Association of School Administrators and National School Board Association conducted an evaluation of the Apollo school system: a small-city/suburban school system (Misner et al., 1963). Apollo is the fictitious name for a colonial-brick-and-stone suburban residential community of 55,872 persons who are largely professional and business people. There are 10,000 students in the Apollo Township School District. Apollo has a definite
pattern of evaluation. Not all of the elements in the pattern could be used in many school systems, but certain aspects should be included in all district evaluation programs. Some of the central features of the Apollo evaluation program are: (1) the parent council, (2) the educational program meetings, (3) evaluation by outside agencies, (4) unique board meetings, and (5) in-service programs.

Misner et al. (1963) stated that evaluation of the educational enterprise in a school system is the responsibility of the board of education, the central office staff, the principal, the teachers, and the school patrons. If all these people are to share in the responsibilities of evaluation, there must be a well-defined and well-planned program.

They also noted that the school principal is once more in a strategic position of leadership. The principal helps in the basic planning for a program that extends throughout the school system and into each school. He works with the teachers, parents, and pupils to implement the program in the school he administers. The strengths and weaknesses of the educational program will first be discovered in each separate school. The principal is close to the problems that the teachers seek to isolate and solve.

Summary

The review of the literature and related studies presented in this chapter concerned teacher behavior and student
and principal ratings of teachers as they relate to teacher performance. Many factors help to determine teacher effectiveness in a school system. These studies attempted to extend the usefulness of student ratings and provide guidelines for accountability of classroom teachers. The practice of obtaining student ratings of teacher effectiveness appears to be increasing.

As a preliminary step to focus on student ratings of teachers, principal ratings of teachers, and student achievement with respect to teacher performance, Chapter II reported on a variety of studies which have dealt with these factors in one way or another. Since teaching is multidimensional, there is a need to evaluate teacher effectiveness in a multidimensional manner.

Chapter III presents the research design, procedures, sample and population, and data analysis for the study.
CHAPTER III

DESIGN AND METHODOLOGY

The major task of the study was to examine the relationships among student ratings of teachers, principal ratings of teachers, and student achievement. Student ratings, principal ratings, and student achievement were the independent variables. The purpose of this chapter is to describe the selection and characteristics of the population, variables and instrumentation, and procedures for collection and analysis of data.

The Sample

The nature of the overall procedure imposed several requirements in the selection of the teacher sample for this study. The sample consisted of teachers in grades 4, 5, and 6 who had been rated by their students and principals, and whose students' test scores were available in reading and mathematics. Student ratings were collected by using the Elementary Teacher Image Questionnaire (ETIQ). A composite score from the student ratings served as the correlation of the class perception of the teacher. The principal ratings consisted of the total score obtained on the Teacher Evaluation Instrument (TEI).
achievement data for reading and mathematics consisted of the gains measured by administering the Metropolitan Achievement Test (MAT) as a pretest and as a posttest.

There were incomplete data for 50 of the 152 elementary teachers in grades 4, 5, and 6. Therefore, the 102 elementary teachers for whom complete data were available were used for the study. By grade level, the numbers used in the study included 34 fourth-, 31 fifth-, and 37 sixth-grade teachers (Table 1). By sex, there were 27 male and 75 female teachers in the three grades (26.5 percent and 73.5 percent, respectively). There were 91 white teachers and 11 black teachers used in the study.

TABLE 1.—Description of teacher population by number per grade level, sex, and race

<table>
<thead>
<tr>
<th>Teacher Description</th>
<th>Number&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Grade level</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>34</td>
<td>.</td>
</tr>
<tr>
<td>5</td>
<td>31</td>
<td>.</td>
</tr>
<tr>
<td>6</td>
<td>37</td>
<td>.</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>27</td>
<td>26.5</td>
</tr>
<tr>
<td>Female</td>
<td>75</td>
<td>73.5</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>91</td>
<td>89.2</td>
</tr>
<tr>
<td>Black</td>
<td>11</td>
<td>10.8</td>
</tr>
</tbody>
</table>

<sup>a</sup>N = 102 teachers used in the study.
There were approximately 2,856 later elementary students rating teachers in the study, using the ETIQ. The approximation of 2,856 was reached by multiplying the number of teachers in the study (N = 102) times the average pupil ratio used by the Kalamazoo Public Schools (N = 28): 102 x 28 = 2,856.

Principal ratings were received from 10 later elementary principals, 9 of whom were white and 1 was black.

Variables and Instrumentation

The variables studied were student perceptions of teachers (using a composite score from the ETIQ student ratings of each teacher), principal ratings of teachers (using the total score from the TEI), and student achievement (using gains in reading and mathematics as measured by administering the MAT as a pretest and posttest). The ETIQ is a modified form of the Student Opinion Questionnaire, developed by Bryan in the early 1950's. The TEI was developed by the Department of Research and Development of the Kalamazoo Public Schools, and is used by principals in the evaluation of teachers. The MAT is published by Harcourt Brace Jovanovich, Inc.

Elementary Teacher Image Questionnaire (ETIQ)

When designing an instrument for obtaining student
ratings of elementary-school teachers, Coats (1970) directed the development of the Elementary Student Opinion Questionnaire (ESOQ) from the Teacher Image Questionnaire (TIQ). In the 20 years since its development, the TIQ has undergone several modifications. The major modification has consisted of some rewording to make it suitable for elementary-school students, plus some changes in the type and number of teacher characteristics for which items were included.

Bryan (1968) pioneered the development of an instrument designed to acquire student ratings of secondary-school teachers. In the 1950's, his questionnaire and an accompanying feedback service were made available to, and made use of for years by, teachers in the Midwest under the direction of Western Michigan University. The questionnaire has been continuously improved and modified and is currently being used by the Educator Feedback Center at Western Michigan University to assist secondary-school teachers in obtaining student perceptions of their performance. Characteristics measured by this instrument include such variables as (1) knowledge of subject, (2) ability to stimulate interest, (3) fairness, (4) control, (5) sense of humor, and (6) attitude toward student opinion. In its present form, the instrument has been renamed the Elementary Teacher Image Questionnaire (ETIQ) and consists of 18 items (Appendix A).
Validity and reliability

The validity and reliability of the ESOQ items drawn from the TIQ have been utilized by the Educator Feedback Center, which has used the questions in hundreds of classroom analyses (Coats, 1970). These analyses conducted by the center on the TIQ have indicated that the image averages are stable and can be changed only with concerted and well-directed effort (Bryan, 1968).

Bryan reported impressive reliability coefficients with 50 randomly selected classes in grades 7-12. In this study, 24-32 students per class reacted to the TIQ. The scores were converted to those reported for whole classes by means of the Spearman-Brown formula for computing test reliability. Reliability coefficients obtained for the first 12 items were as follows:

1. Knowledge of subject .87
2. Clarity of explanation .82
3. Fairness .84
4. Control .95
5. Attitude toward students .88
6. Ability to stimulate interest .87
7. Attitude toward subject .90
8. Attitude toward student opinion .86
9. Variety of teaching procedures .91
10. Encouragement of student participation .77
11. Sense of humor .91
12. Planning and preparation .90

Reliability of the 12 items was determined by correlating average student responses for chance-halves of 50 randomly selected classes. The correlation coefficients are based on
responses of 24-32 students per class and were converted to those reported for whole classes by means of the Spearman-Brown formula for computing test reliability.

According to the study by Coats (1970), it was concluded that one basic factor, a kind of halo effect based on "charisma" or "popularity," accounted for 61 percent of the variance in student reactions to teachers. Of the remaining 39 percent variance, approximately 16 percent was attributed to two factors less clearly defined. Coats alluded to the fact that at least 39 percent of the variance in student ratings of teachers is independent of the charismatic factor and probably represents fairly objective student judgment (Proctor, 1972).

Teacher Evaluation Instrument (TEI)

Principal ratings were made on the Teacher Evaluation Instrument (TEI) for teachers in grades 4, 5, and 6. The TEI was developed by the Kalamazoo Public Schools Department of Research and Development and has been in use for four years by principals in the school system for evaluating teacher performance (Appendix B). Teacher evaluations are made by principals and turned in to the Employee Relations Division. Teacher evaluation data were available for each of the teachers in the study. A 5-point rating scale is used on the TEI. The scale is as follows:
1 = Poor (unacceptable performance)
2 = Fair (needs improvement—does not meet performance requirements)
3 = Satisfactory (meets performance requirements)
4 = Good (exceeds performance requirements)
5 = Excellent (far exceeds performance requirements)

Based on the principal's rating of a teacher's performance, the principal checks the appropriate response. If the principal believes a certain criterion does not apply, he does not check a response category. For the areas rated below satisfactory, the principal has to provide explanations. There is also space for the principal to give specific suggestions for improvements.

A conference regarding the appraisal must be held with the teacher. Teachers are asked to sign the TEI, with the principal signing also. The signature does not necessarily indicate agreement with the appraisal, but indicates that the teacher has seen it. The teacher may attach a memo explaining his or her disagreement regarding any aspect of the evaluation.

Validity and reliability

Because the TEI items have been found to be reliable and valid, the instrument has been used by principals in hundreds of classroom analyses since 1973. The analyses conducted by principals in the Kalamazoo Public Schools
have indicated that the instrument averages are stable. The validity of an instrument is the indicator of how well the instrument measures the subject-matter content and behaviors under consideration (Gronlund, 1971).

Reliability refers to the consistency of measurement. In short, reliability merely provides the consistency which makes validity possible (Gronlund, 1971). Information gathered by the Department of Research and Development yielded relatively precise results for each principal, and a teacher would receive about the same ratings on the instrument if the teacher were rated more than once under the same conditions. In this study, the split-half method was used to determine internal consistency (reliability). The TEI was administered once, and the reliability (split-half) was measured by using the Spearman-Brown formula.

In 1975, the Department of Research and Development, with complete data for 481 randomly selected instruments on teachers in grades K-12, reported an impressive composite reliability coefficient of .7745.

**Metropolitan Achievement Test (MAT)**

The Metropolitan Achievement Test (MAT) was developed and prepared by Harcourt Brace Jovanovich, Inc. Various levels (Primer, Primary I, Primary II, Elementary, Intermediate, and Advanced) are administered in the Kalamazoo
Public Schools each school year as pretests and posttests. For this study, however, data were used from the Elementary and Intermediate levels in reading and mathematics. These two subject areas were selected because of the similar content and level of difficulty.

There are various subdivisions of the MAT. In grades 1-6, five different levels or batteries (mentioned above) are used. Each level has three different forms (F, G, and H). The various levels have been deemed appropriate for administration at specific grade levels by the publisher, Harcourt Brace Jovanovich, Inc. Figure 4 shows a breakdown of the MAT.

![Figure 4. Metropolitan Achievement Test breakdown.](image)

In the Kalamazoo Public Schools, a fourth-grade student takes the Elementary level of the MAT in the fall (September) and in the spring (May) of a particular school year. Both
fifth- and sixth-grade students take the Intermediate level of the MAT in the fall (September) and spring (May) each school year. Figure 5 shows when during the school year the various levels of the MAT are administered to the appropriate grades as pretests (September) and/or posttests (May).

<table>
<thead>
<tr>
<th>Level</th>
<th>Grade</th>
<th>Pretest</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primer</td>
<td>1</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Primary I</td>
<td>1</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Primary I</td>
<td>2</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Primary II</td>
<td>2</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Primary II</td>
<td>3</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Elementary</td>
<td>3</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Elementary</td>
<td>4</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Intermediate</td>
<td>5</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Intermediate</td>
<td>6</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Advanced</td>
<td>7</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Advanced</td>
<td>8</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Advanced</td>
<td>9</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

Figure 5. Metropolitan Achievement Test levels, grades, and time of year administered.

Item analysis

According to Durost, Bixler, Wrightstone, Prescott, and Balow (1971), items which appear on the MAT were examined through a process known as item analysis. In an item analysis program, each potential test item is examined. Test authors examine individual test items in order to eliminate undesirable items and to assure that items are appropriate for the level at which they are intended to be administered.
In the case of the MAT, item analysis was conducted by administering an experimental test form to a sample of students.

The Metropolitan Item Analysis (IA) Program was conducted in April 1968. Approximately 45,000 students in grades 1 through 8, in 22 school systems in 14 states, were tested. All students in the sample population took the Otis-Lennon Mental Ability Test and an experimental form of the MAT. Five forms of the experimental MAT were used at the following levels: Primary I, Primary II, Elementary, Intermediate, and Advanced. The use of five forms at each level (battery) allowed about 12,000 new MAT test items to be examined or "tried out." Each battery was tried out at one grade level above the one for which it was intended.

Schools participating in the Metropolitan Item Analysis Program were selected to provide representative samples of various socioeconomic levels, sizes of cities, and geographic regions found in the United States. The socioeconomic levels in the sample were considered of utmost importance because of the high relationship between socioeconomic levels and test performance.

In determining socioeconomic levels of the sample population, consideration was given to median family income and median years of schooling of people in the community over 24 years of age. Representation in the IA sample was obtained from metropolitan areas, rural areas, and "medium-sized"
cities and towns. The percentage of the U.S. population residing in cities and towns of various sizes was used to determine the percentage of students from various sizes of cities to be included in the IA sample. In selecting a representative geographic sample, an attempt was made to match the percentage of the U.S. population living in the northeast, north central, southern, and western parts of the United States with the percentage of students living in those areas who took the experimental forms of the MAT.

Each school system that participated in the IA program provided information regarding, among other things, characteristics of the schools in the system, the population served by the school system, a list of textbooks used, and the ethnic breakdown of the students. Approximately 73.2 percent of the students in the MAT IA group were native-born white, 12 percent were black, 11.8 percent were either Puerto Rican or Mexican, 1.5 percent were Oriental, 1.1 percent were foreign-born white, and .3 percent were classified as "other."

Additional information was also provided by school systems participating in the IA program. The schools in the IA program indicated that about 25 percent of their students came from rural areas. About 63 percent of the first-grade students had attended kindergarten. The ungraded plan for grades 1-3 was used by 3 of the systems in the IA group, and 4 other systems were to start using the ungraded plan within
3 years. Team teaching was found in 9 systems, and teacher aides were found in 9 systems. Homogeneous grouping was used in 16 school systems.

Information was also provided by IA participating school systems regarding teachers' starting salaries, teachers' average salaries, per-pupil expenditures, pupil-teacher ratios, and the average years of teaching experience for IA teachers. In each of these characteristics, the IA sample approximated the characteristic for the national population.

The IA school systems were asked questions regarding their mathematics curriculum. The majority of school systems said that their mathematics curriculum could be classified as "modern" (as opposed to being "traditional" or "transitional"). None of the school systems indicated that their mathematics program was "traditional."

IA school systems were questioned about their language programs. One school system indicated that they had adopted linguistically oriented textbooks, and eight systems in the IA group said they would adopt them within the next 2 years.

Also, data were obtained from each of the 22 IA school systems regarding the amount of instruction time in the average school day devoted to various subjects in each grade.

Finally, for each of the 12,000 items used in the IA program, information was obtained regarding (1) the percentage of students choosing each option, (2) the percentage of students in the top 27 percent and the bottom 27 percent of
the total test score distribution choosing each option, and (3) those not marking an answer for an item. In turn, item cards were prepared showing all data gathered for each item, for each grade level at which it was used with the IA sample.

MAT standardization

Durost et al. (1971) stated that approximately 250,000 students and their teachers were involved in the 3 years of research during which the MAT was developed.

During the first stage of the test development, the test authors and assistants analyzed and summarized current curricular materials including leading textbook series, syllabuses, and state guidelines. Test "blueprints" were then prepared which indicated the proportion of test items needed on various topics in order to give balanced coverage of the curriculum. Actual test items were then written to cover each subtopic in the blueprints.

After the test items were edited, they were subjected to extensive classroom tryout. A sample of 45,000 students were tested with a total of 12,000 test items for the various levels and forms of the MAT. This IA program provided data on the difficulty level of each item as well as various statistical measures for the item and the subtest. Since only about 6,000 items were needed for the final edition of the test, those items which proved to be ambiguous, too easy, too difficult, or otherwise unsuitable could be eliminated.
Teachers who participated in the IA research provided valuable information which was used in formulating the final test content, directions, and time limits.

The national standardization for the MAT was conducted to obtain final norms and supporting data for the test series. The sample of pupils tested was selected to provide a set of norms which accurately reflected national levels of achievement. Factors such as geographic region, size of city, socioeconomic status, and public versus nonpublic schools were all considered to insure that the standardization sample was representative of the national population (Table 2).

**TABLE 2.--Percent of pupils by city size and geographic region in national population and metropolitan standardization samples**

<table>
<thead>
<tr>
<th>Factor</th>
<th>National Population</th>
<th>Metropolitan Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>City size</td>
<td></td>
<td></td>
</tr>
<tr>
<td>250,000 +</td>
<td>22%</td>
<td>21%</td>
</tr>
<tr>
<td>25,000 - 249,000</td>
<td>20</td>
<td>21</td>
</tr>
<tr>
<td>2,500 - 24,000</td>
<td>28</td>
<td>29</td>
</tr>
<tr>
<td>Rural</td>
<td>30</td>
<td>28</td>
</tr>
<tr>
<td>Total</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Geographic region</th>
<th>National Population</th>
<th>Metropolitan Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Northeast</td>
<td>27%</td>
<td>23%</td>
</tr>
<tr>
<td>Midwest</td>
<td>29</td>
<td>28</td>
</tr>
<tr>
<td>Southeast</td>
<td>22</td>
<td>27</td>
</tr>
<tr>
<td>West</td>
<td>22</td>
<td>22</td>
</tr>
<tr>
<td>Total</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Note. Metropolitan data are averages for fall and spring standardization groups combined.
Validity and reliability

The validity of an achievement test is defined primarily in terms of content validity. A test has content validity if the test items adequately cover the curricular areas that the test is supposed to evaluate. Since each school has its own curriculum, the content validity of the MAT must be evaluated by each school. It cannot be claimed that the tests are universally valid. To assist schools in judging the content validity of the test, Harcourt Brace Jovanovich (Durost et al., 1971) prepared content outlines and described the procedures used in developing test content.

Grade 4.--Table 3 presents reliability data for the

<table>
<thead>
<tr>
<th>Test</th>
<th>$r_{KE}$</th>
<th>$r_{TT}$</th>
<th>Standard Error of Measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>RS</td>
</tr>
<tr>
<td>Word knowledge</td>
<td>.94</td>
<td>.95</td>
<td>2.5</td>
</tr>
<tr>
<td>Reading</td>
<td>.92</td>
<td>.93</td>
<td>2.5</td>
</tr>
<tr>
<td>Total reading</td>
<td>.96</td>
<td>.97</td>
<td>3.5</td>
</tr>
<tr>
<td>Language</td>
<td>.93</td>
<td>.93</td>
<td>2.8</td>
</tr>
<tr>
<td>Spelling</td>
<td>.96</td>
<td>.97</td>
<td>2.0</td>
</tr>
<tr>
<td>Math computation</td>
<td>.88</td>
<td>.91</td>
<td>2.2</td>
</tr>
<tr>
<td>Math concepts</td>
<td>.90</td>
<td>.91</td>
<td>2.4</td>
</tr>
<tr>
<td>Math problem-solving</td>
<td>.91</td>
<td>.93</td>
<td>2.0</td>
</tr>
<tr>
<td>Total mathematics</td>
<td>.96</td>
<td>.97</td>
<td>4.0</td>
</tr>
</tbody>
</table>

Elementary battery tests administered at the beginning of Grade 4. Split-half coefficients corrected by the Spearman-Brown formula are given, as well as Saupe's estimate of

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Kuder-Richardson Formula 20 reliability, and standard errors of measurement in terms of raw score, standard score, and grade equivalent. Data are based on all students tested in the fall standardization with Form G at grade 4. Standard errors of measurement are based on use of the split-half coefficient.

**Grades 5 and 6.**—Table 4 presents reliability data for the Intermediate battery tests administered at the beginning of grades 5 and 6. Split-half coefficients corrected by the Spearman-Brown formula are given, as well as Saupe's estimate of Kuder-Richardson Formula 20 reliability, and standard errors of measurement in terms of raw score, standard score, and grade equivalent. Data are based on all students tested in the fall standardization with Form G at grades 5 and 6.

<table>
<thead>
<tr>
<th>Test</th>
<th>( \frac{r_{KE}}{r_{LT}} )</th>
<th>RS</th>
<th>SS</th>
<th>GE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word knowledge</td>
<td>.92 .93</td>
<td>2.7</td>
<td>3.4</td>
<td>.5</td>
</tr>
<tr>
<td>Reading</td>
<td>.93 .93</td>
<td>2.6</td>
<td>3.8</td>
<td>.5</td>
</tr>
<tr>
<td>Total reading</td>
<td>.96 .96</td>
<td>3.7</td>
<td>2.7</td>
<td>.4</td>
</tr>
<tr>
<td>Language</td>
<td>.95 .95</td>
<td>4.1</td>
<td>2.6</td>
<td>.4</td>
</tr>
<tr>
<td>Spelling</td>
<td>.90 .91</td>
<td>2.7</td>
<td>3.4</td>
<td>.5</td>
</tr>
<tr>
<td>Math computation</td>
<td>.84 .88</td>
<td>2.3</td>
<td>3.5</td>
<td>.4</td>
</tr>
<tr>
<td>Math concepts</td>
<td>.88 .89</td>
<td>2.5</td>
<td>3.7</td>
<td>.5</td>
</tr>
<tr>
<td>Math problem-solving</td>
<td>.89 .89</td>
<td>2.3</td>
<td>4.1</td>
<td>.5</td>
</tr>
<tr>
<td>Total mathematics</td>
<td>.95 .96</td>
<td>4.1</td>
<td>2.2</td>
<td>.3</td>
</tr>
<tr>
<td>Science</td>
<td>.94 .95</td>
<td>3.5</td>
<td>2.6</td>
<td>.4</td>
</tr>
<tr>
<td>Social studies</td>
<td>.95 .95</td>
<td>3.8</td>
<td>2.5</td>
<td>.4</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Standard errors of measurement are based on use of the split-half coefficient.

Use of standardized instruments with urban and minority-group students

Fitzgibbon (1971) set forth the position of the Harcourt Brace Jovanovich Test Department on the use of standardized instruments with urban and minority-group students, and described the specific steps which have been taken and "the concern which will continue to be shown by the Test Department in our effort to improve the effectiveness and useability of our test instruments in the evaluation of urban and minority-group pupils":

There is little question that testing today is in a position of visibility and public concern such as it has not experienced throughout most of its prior history on the American educational and social scene. The issues that have engaged us most seriously on the domestic front for the past several years have been those centering around the emergence of minority groups, the impact of the types of education afforded these groups, on their aspirations and advancement in American society, and the social consequences of various types of governmental resources. In all of these contexts, test data have a focus of concern or a major line of evidence in support of one or another type of governmental action. Because the issues are difficult and complicated ones, beset with political and emotional overtones, there has been a predisposition to criticize the test results reported in support of one or another course of action, and an inclination to question or even to repudiate them when they have seemed to run counter to deeply held attitudes or prejudices. It is difficult to recall a period in which test results have been so surrounded by controversy, or when they have generated such intense reaction, either
positive or negative, from large segments of the citizenry. One could cite many instances which bespeak the growing conviction on the part of minority groups that tests constitute for them an unfair obstacle to advancement in our society. Their responses range from demands for development of more appropriate instruments to modifications in the ways in which measures of performance are interpreted and used in selection decisions.

Lennon (1969) expressed his concern about the issue when he said that: "In the face of this concern and controversy about testing in relation to various social issues, we are conscious of the special responsibilities which fall upon makers and distributors of tests. As test publishers, we are giving renewed and intensified attention to certain aspects of the traditional test-development enterprise in order to ensure the appropriateness of the final instruments for the uses to which they will be put. We believe it is fair to characterize our response not as 'business as usual,' but as a more sensitive and more sharply focused conduct of our unusual business—that of making good tests and facilitating their proper use."

The above statement by Lennon (1969) of HBJ [Harcourt Brace Jovanovich] is the corporate philosophy in regard to our responsibilities as test publishers in the matter of minority concerns. (pp. 1-2)

Raspberry (1971) made the following statement about testing minorities:

You know that poor blacks and other disadvantaged minorities don't do as well on standardized tests as middle-class whites.

You also know that one of the reasons frequently given for this shortcoming is "cultural bias"—that is, the tests either don't ask questions based on the knowledge that the minorities have accumulated, or they ask them in such a way that the minorities show less knowledge than they have.

But if the standardized tests are biased against certain minorities on these counts, it
certainly ought to be possible to construct tests that are biased toward these same minorities: tests that are based on the peculiar life-styles of the big-city ghetto, for instance, and administered in ways that will show ghetto test-takers to advantage. (p. 19)

Williams, professor of psychology and head of the black studies program at Washington University, St. Louis, has a modest example: the Black Intelligence Test of Cultural Homogeneity (BITCH).

According to Raspberry (1971), Williams stated that his research has shown that nearly all blacks who take his test do better than nearly all whites who take it. Williams sees that outcome as proof that cultural bias is a manipulable thing that can be made to favor any group the test-makers want it to.

Raspberry further explained that the tests that are the despair of disadvantaged blacks--the tests that keep them out of the good tracks in school, the good colleges, and the good jobs--are those that purport to measure skills, aptitudes, achievement, and reasoning ability. Williams' test measures knowledge of a specialized vocabulary.

Most critics charge that many of the tests are inadequate measures of what they purport to measure and that some of them--notably, the IQ type--may be positively harmful, at least in some of their usages. Raspberry's (1971) solution to the problem is the point that many critics of standardized tests keep missing: since there are going to be tests for
as long as there are more applicants than places, the solution is not to throw out the tests but to insist on making the tests do what they allege to do, and to give minorities the wherewithal to pass them—by teaching them how to pass tests, if necessary.

Green (1974) reported that a school's reliance on intelligence test scores as measures of learning capacity helps stunt development of many black youths. Green stated that educational testing is "big business." Last year, the standardized testing industry reported an income of more than $100 million. The industry makes this money by selling its tests to schools and convincing school administrators that test results are valid indicators of learning ability and future educational success.

Testing blacks in order to ascertain their potential, Green (1974) contends, is often no more scientific than the Gypsy practice of reading tea leaves. Gypsies read great truths from tea leaves; often, educators read great truths from test scores. Neither reading is necessarily valid, but many people believe them. Green alludes to the fact that studies have shown that aptitude and intelligence tests frequently do not accurately predict a child's academic future, especially when children have been the victims of racial and other forms of discrimination.

For example, 13-year-old Bill, from San Diego, repairs all the broken lamps, television sets, toasters, and vacuum
cleaners in his neighborhood. His neighbors say he's alert, friendly, and always ready with a quick, funny reply. Bill moved to San Diego when he was 8 years old. During the first week of school there, he took an intelligence test, scored poorly, and was placed in a special education class. He's been there ever since. Bill hates school because he is taught only to play games and color. He feels very frustrated because he has never learned to read or write.

According to Green (1974), Bill is a victim of testing superstition, the belief that an intelligence test can actually measure natural intelligence. This belief is one that psychologists and educators have not been able to prove. Undergirding this superstition are two general assumptions: (1) everyone is born with a specific identifiable level of intelligence, and (2) that level remains fairly constant throughout life. When a kindergarten child achieves an average score on an intelligence test, his teachers predict that he will do average work. If a child scores below average, then one more janitor; above average, one more lawyer. Thus, a low score allows educators to make specific decisions about a child's future. For many black children, these decisions often limit their educational future and career potential.

Standardized tests are tools which Green (1974) refers to when talking about measuring the learning capacity of youths. Obviously there is no magic way to judge a child's
learning capabilities. Yet, there are three types of "intelligence" tests used in elementary schools to judge a child's learning capabilities. With intelligence tests, many school administrators try to measure a child's general ability to reason. With aptitude tests, they try to measure the ability to learn specific skills such as art, music, or mathematics. Achievement tests, while designed to measure acquired skills in areas such as vocabulary or mathematics, are often believed to measure ability. None of these tests can in fact measure a child's learning capabilities. Educators should realize that tests are tools which can be used wisely or harmfully, to help or to hinder the educational growth of children.

Williams (1974) contended that the standardized test "is the hired gun in the war of scientific racism. . . . Since the American society is pluralistic on the one hand and racist on the other, it would be virtually impossible to conceptualize an instrument which would be fair to all people: Asians, Blacks, Caucasians, Chicanos, Indians, and Puerto Ricans" (p. 77). Williams stated that although the search for culturally fair tests has been intelligently criticized, an equally strong objection can be raised against norm-referenced and other conventional tests. In light of the methodological and theoretical difficulties involved in developing culturally fair and culturally free tests, it is necessary, therefore, to examine several alternative consid-
erations in test construction. Based on this belief, Williams (1974) developed the BITCH. He cited that "in spite of the many efforts made to develop culturally fair and culturally free tests, none has been developed" (p. 78).

Previously, Williams (1970) had suggested constructing a test based on items drawn exclusively from the black culture. Part of Williams' rationale was based on the fact that culture-specific tests have the advantage of dealing with content material that is familiar to the black child. This means that he already has "stowed away" mental images of the material, so he does not have to deal with the foreign or unfamiliar aspects of these materials. Thus, a combination of dialect- and culture-specific tests would certainly enhance the possibility of measuring accurately what is inside the black child's head. This, then, was the basic rationale for the BITCH-100.

Williams' (1974) study for the use of the BITCH was conducted in St. Louis, and included 100 white subjects and 100 black subjects. All subjects were high school students ranging in age from 16 to 18 years. Half of the subjects were from low socioeconomic levels, whereas the other half came from the middle levels.

Experience with the tryouts during standardization, as reported by Williams (1974), indicated that virtually all black subjects became intensely interested in the test. Comments were made such as: "Man, this is a bad test. . . .
This is really hip . . . it's outta sight." Black subjects frequently came across items which were humorous and quite familiar to them. White subjects seemed to be quite challenged by the test and appeared tense. Many sighed and showed other signs of discomfort. A few questioned the validity of the instrument; others stated that if the test was valid, then they had little knowledge of the black experience.

The black group showed a clear superiority of 36 mean points over the white group, a difference that is significant at the .001 level of confidence. The interpretation of the results was considered two-fold: (1) a culture-specific test clearly shows the abilities of the group for which the test was intended; and (2) a culture-specific test does not accurately reflect the abilities for a nonrepresentative group.

Collection of Data

All classroom teachers in grades 4, 5, and 6 included in the present study were rated on the ETIQ by the students in their classrooms. Student rating materials were supplied to building principals by the Kalamazoo Public Schools Department of Research and Development. The completed student ratings were returned to the Department of Research and Development by March 15, 1974.

Teachers in grades 4, 5, and 6 were rated by principals using the TEI. While there was not a maximum length of time
required of the principals for classroom observation of a teacher's classroom performance, the observation should have been sufficiently frequent and long enough to allow for an objective analysis of the teacher's performance. The minimum length of time required for observation was 30 minutes. Principals returned the completed evaluation instruments to the Employee Relations Division by March 15, 1974.

In grades 4, 5, and 6, the MAT was administered in September (pretest) and May (posttest) during the 1973-74 school year. In grade 4, the Elementary MAT was administered by the homeroom teacher. In grades 5 and 6, the Intermediate MAT was administered by the homeroom teacher. Results were mailed to the Department of Research and Development, where information was obtained for the study. Data were analyzed in reading and mathematics, using the actual gain scores from the reading and mathematics pre- and posttest results.

Data Analysis

Two statistical treatments were used in this study to determine the extent of relationships between the independent and dependent variables. Answers to the various questions were sorted by using a one-way analysis of variance and coefficients of correlation.

The coefficients of correlation were used for Hypotheses 1-5 for grades 4, 5, and 6. These hypotheses are:
\[ H_1: \text{There is a positive relationship between student achievement (reading) and student ratings of teachers.} \]

\[ H_2: \text{There is a positive relationship between student achievement (mathematics) and student ratings of teachers.} \]

\[ H_3: \text{There is a positive relationship between student achievement (reading) and principal ratings of teachers.} \]

\[ H_4: \text{There is a positive relationship between student achievement (mathematics) and principal ratings of teachers.} \]

\[ H_5: \text{There is a positive relationship between student ratings of teachers and principal ratings of teachers.} \]

A one-way analysis of variance was used for Hypotheses 6-13 for grades 4, 5, and 6. These hypotheses are:

\[ H_6: \text{There is no significant difference in student achievement (reading) when comparing the educational backgrounds of teachers.} \]

\[ H_7: \text{There is no significant difference in student achievement (mathematics) when comparing the educational backgrounds of teachers.} \]

\[ H_8: \text{There is no significant difference in student ratings of teachers when comparing the educational backgrounds of teachers.} \]

\[ H_9: \text{There is no significant difference in principal ratings of teachers when comparing the educational backgrounds of teachers.} \]

\[ H_{10}: \text{There is no significant difference between student achievement (reading) in grades 4, 5, and 6.} \]

\[ H_{11}: \text{There is no significant difference between student achievement (mathematics) in grades 4, 5, and 6.} \]
There is no significant difference between student ratings of teachers in grades 4, 5, and 6.

There is no significant difference between principal ratings of teachers in grades 4, 5, and 6.

Correlation matrices were used to determine the extent of relationships among student ratings, principal ratings, and student achievement. F values were reported for differences among and between means. Probabilities were reported at the .05 level of significance.
CHAPTER IV

RESULTS

Results of analyses related to the hypotheses stated in Chapter III are presented in this chapter. Statistical treatment of the data involved the one-way analysis of variance and coefficients of correlation. For the purpose of this study, the relationship between two variables was considered to be significant if the probability level was < .05.

Six major questions were posed for the present study:

1. What is the relationship between student achievement (reading) and student ratings of teachers?

2. What is the relationship between student achievement (mathematics) and student ratings of teachers?

3. What is the relationship between student achievement (reading) and principal ratings of teachers?

4. What is the relationship between student achievement (mathematics) and principal ratings of teachers?

5. What is the relationship between student achievement and educational backgrounds of teachers?

6. What is the relationship between student ratings of teachers and principal ratings of teachers?

Presentation of data analysis is consistent with the order in which the main questions and related hypotheses
were posed in Chapter I. Each hypothesis was tested for all grades (4, 5, and 6) and for each grade separately.

Hypotheses 1-5

Coefficients of correlation were computed to test Hypotheses 1-5 for grades 4, 5, and 6 combined, and for each grade separately. The hypotheses are:

\( H_1 \): There is a positive relationship between student achievement (reading) and student ratings of teachers.

\( H_2 \): There is a positive relationship between student achievement (mathematics) and student ratings of teachers.

\( H_3 \): There is a positive relationship between student achievement (reading) and principal ratings of teachers.

\( H_4 \): There is a positive relationship between student achievement (mathematics) and principal ratings of teachers.

\( H_5 \): There is a positive relationship between student ratings of teachers and principal ratings of teachers.

A summary of data obtained from statistical analyses of these pair-wise comparisons is presented in Table 5. The table contains coefficients of correlation for the relationships relevant to the first five hypotheses for all grades (4, 5, and 6) included in the study. The total number of teachers was 102 \((N = 102)\). Based on analyses reported in Table 5, Hypotheses 1, 2, 3, 4, and 5 for all grades were not found to be significant. Results indicated a correlation of .02 between student achievement in reading
and student ratings of teachers, .10 between student achievement in mathematics and student ratings of teachers, .14 between student achievement in reading and principal ratings of teachers, .12 between student achievement in mathematics and principal ratings of teachers, and .14 between student ratings of teachers and principal ratings of teachers. None of the coefficients of correlation was found to be significant at the .05 level.

TABLE 5.—Extent of relationship between student ratings of teachers, principal ratings of teachers, and student achievement (reading and mathematics) in grades 4, 5, and 6

<table>
<thead>
<tr>
<th>Matched Pair</th>
<th>N</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student achievement (reading) x Student ratings</td>
<td>102</td>
<td>.0226</td>
</tr>
<tr>
<td>Student achievement (mathematics) x Student ratings</td>
<td>102</td>
<td>.1044</td>
</tr>
<tr>
<td>Student achievement (reading) x Principal ratings</td>
<td>102</td>
<td>.1395</td>
</tr>
<tr>
<td>Student achievement (mathematics) x Principal ratings</td>
<td>102</td>
<td>.1164</td>
</tr>
<tr>
<td>Student ratings x Principal ratings</td>
<td>102</td>
<td>.1384</td>
</tr>
</tbody>
</table>

Correlations were used to analyze the relationships between student ratings of teachers, principal ratings of teachers, and student achievement in reading and mathematics for grades 4, 5, and 6 separately. For grade 4 (Table 6),
TABLE 6.—Extent of relationship between student ratings of teachers, principal ratings of teachers, and student achievement (reading and mathematics) in grade 4

<table>
<thead>
<tr>
<th>Matched Pair</th>
<th>N</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student achievement (reading) x Student ratings</td>
<td>34</td>
<td>.1197</td>
</tr>
<tr>
<td>Student achievement (mathematics) x Student ratings</td>
<td>34</td>
<td>.0038</td>
</tr>
<tr>
<td>Student achievement (reading) x Principal ratings</td>
<td>34</td>
<td>.4538*</td>
</tr>
<tr>
<td>Student achievement (mathematics) x Principal ratings</td>
<td>34</td>
<td>.2998*</td>
</tr>
<tr>
<td>Student ratings x Principal ratings</td>
<td>34</td>
<td>.0983</td>
</tr>
</tbody>
</table>

*p < .05

The correlation (.12) between student achievement in reading and student ratings of teachers was not found to be significant. The correlation (.00) between student achievement in mathematics and student ratings of teachers was not found to be significant. Therefore, Hypotheses 1 and 2 could not be accepted at the .05 level of significance. There were significant correlations (.45 and .30, respectively) between student achievement in reading and principal ratings of teachers, and between student achievement in mathematics and principal ratings of teachers. Therefore, Hypotheses 3 and 4 were accepted at the .05 level of significance. The
correlation (.10) between student ratings of teachers and principal ratings of teachers was not found to be significant. Therefore, Hypothesis 5 could not be accepted at the .05 level of significance.

For grade 5 (Table 7), the correlation (.04) between student achievement in reading and student ratings of teachers was not found to be significant. The correlation (.12) between student achievement in mathematics and student ratings of teachers was not found to be significant. Therefore, Hypotheses 1 and 2 could not be accepted at the .05 level of significance. There was a significant correlation (.35)

**TABLE 7.--Extent of relationship between student ratings of teachers, principal ratings of teachers, and student achievement (reading and mathematics) in grade 5**

<table>
<thead>
<tr>
<th>Matched Pair</th>
<th>N</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student achievement (reading)</td>
<td>x</td>
<td>Student ratings</td>
</tr>
<tr>
<td>Student achievement (mathematics)</td>
<td>x</td>
<td>Student ratings</td>
</tr>
<tr>
<td>Student achievement (reading)</td>
<td>x</td>
<td>Principal ratings</td>
</tr>
<tr>
<td>Student achievement (mathematics)</td>
<td>x</td>
<td>Principal ratings</td>
</tr>
<tr>
<td>Student ratings</td>
<td>x</td>
<td>Principal ratings</td>
</tr>
</tbody>
</table>

*p < .05

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
between student achievement in reading and principal ratings of teachers. Hypothesis 3 was accepted at the .05 level of significance. The correlation (.01) between student achievement in mathematics and principal ratings of teachers was not found to be significant. The correlation (.04) between student ratings of teachers and principal ratings of teachers was not found to be significant. Therefore, Hypotheses 4 and 5 could not be accepted at the .05 level of significance.

For grade 6 (Table 8), the correlation (.06) between student achievement in reading and student ratings of teachers was not found to be significant. The correlation (.10)

<table>
<thead>
<tr>
<th>Matched Pair</th>
<th>N</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student achievement (reading) x Student ratings</td>
<td>37</td>
<td>-.0648</td>
</tr>
<tr>
<td>Student achievement (mathematics) x Student ratings</td>
<td>37</td>
<td>.1044</td>
</tr>
<tr>
<td>Student achievement (reading) x Principal ratings</td>
<td>37</td>
<td>-.3401*</td>
</tr>
<tr>
<td>Student achievement (mathematics) x Principal ratings</td>
<td>37</td>
<td>.1438</td>
</tr>
<tr>
<td>Student ratings x Principal ratings</td>
<td>37</td>
<td>.2566*</td>
</tr>
</tbody>
</table>

*p < .05
between student achievement in mathematics and student ratings of teachers was not found to be significant. Therefore, Hypotheses 1 and 2 could not be accepted at the .05 level of significance. There was a significant correlation (.34) between student achievement in reading and principal ratings of teachers. Therefore, Hypothesis 3 was accepted at the .05 level of significance. The correlation (.14) between student achievement in mathematics and principal ratings of teachers was not found to be significant. Therefore Hypothesis 4 could not be accepted at the .05 level of significance. There was a significant correlation (.26) between student ratings of teachers and principal ratings of teachers. Therefore, Hypothesis 5 was accepted at the .05 level of significance.

Hypotheses 6-13

A one-way analysis of variance was used to test Hypotheses 6-9 for grades 4, 5, and 6 and for each grade separately, and to test Hypotheses 10-13 for all grades combined.

H₆: There is no significant difference in student achievement (reading) when comparing the educational backgrounds of teachers.

Results shown in Table 9 indicated that a significant difference was not found between student achievement in reading in grades 4, 5, and 6 and the educational backgrounds of the participating teachers.
TABLE 9.—One-way analysis of variance for determining the extent of relationship between student achievement (reading) and educational backgrounds of teachers in grades 4, 5, and 6

<table>
<thead>
<tr>
<th>Degree</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bachelor's</td>
<td>58</td>
<td>9.02</td>
<td>2.9465</td>
</tr>
<tr>
<td>Master's</td>
<td>43</td>
<td>8.52</td>
<td>2.1183</td>
</tr>
<tr>
<td>Master's + 30 hours</td>
<td>1</td>
<td>8.00</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between groups</td>
<td>5.38</td>
<td>2</td>
<td>2.69</td>
<td>.39</td>
<td>.68</td>
</tr>
<tr>
<td>Within groups</td>
<td>688.82</td>
<td>99</td>
<td>6.90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>694.20</td>
<td>101</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Results shown in Table 10 failed to indicate a significant difference between student achievement in reading in grade 4 and the educational backgrounds of the teachers.

TABLE 10.—One-way analysis of variance for determining the extent of relationship between student achievement (reading) and educational backgrounds of teachers in grade 4

<table>
<thead>
<tr>
<th>Degree</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bachelor's</td>
<td>19</td>
<td>11.00</td>
<td>3.4480</td>
</tr>
<tr>
<td>Master's</td>
<td>15</td>
<td>10.00</td>
<td>2.0354</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between groups</td>
<td>8.38</td>
<td>2</td>
<td>4.19</td>
<td>.48</td>
<td>.62</td>
</tr>
<tr>
<td>Within groups</td>
<td>272.00</td>
<td>31</td>
<td>8.77</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>280.38</td>
<td>33</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Results shown in Table 11 failed to indicate a significant difference between student achievement in reading in
TABLE 11.—One-way analysis of variance for determining the extent of relationship between student achievement (reading) and educational backgrounds of teachers in grade 5

<table>
<thead>
<tr>
<th>Degree</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bachelor's</td>
<td>17</td>
<td>8.53</td>
<td>1.5049</td>
</tr>
<tr>
<td>Master's</td>
<td>14</td>
<td>8.26</td>
<td>1.7289</td>
</tr>
</tbody>
</table>

Source

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between groups</td>
<td>0.46</td>
<td>2</td>
<td>0.23</td>
<td>.09</td>
<td>.92</td>
</tr>
<tr>
<td>Within groups</td>
<td>75.09</td>
<td>28</td>
<td>2.68</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>75.55</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

grade 5 and the educational backgrounds of the teachers.

Results shown in Table 12 failed to indicate a significant difference between student achievement in reading in grade 6 and the educational backgrounds of the teachers.

TABLE 12.—One-way analysis of variance for determining the extent of relationship between student achievement (reading) and educational backgrounds of teachers in grade 6

<table>
<thead>
<tr>
<th>Degree</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bachelor's</td>
<td>22</td>
<td>7.68</td>
<td>2.4570</td>
</tr>
<tr>
<td>Master's</td>
<td>14</td>
<td>7.36</td>
<td>1.7368</td>
</tr>
<tr>
<td>Master's + 30 hours</td>
<td>1</td>
<td>8.00</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Source

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between groups</td>
<td>1.09</td>
<td>2</td>
<td>0.54</td>
<td>.11</td>
<td>.89</td>
</tr>
<tr>
<td>Within groups</td>
<td>165.98</td>
<td>34</td>
<td>4.88</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>.67.07</td>
<td>36</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
For Hypothesis 7, results were reported for all grades (4, 5, and 6) and for each grade separately.

H₇: There is no significant difference in student achievement (mathematics) when comparing the educational backgrounds of teachers.

Results shown in Table 13 failed to indicate a significant difference between student achievement in mathematics in grades 4, 5, and 6 and the educational backgrounds of the teachers.

| TABLE 13.—One-way analysis of variance for determining the extent of relationship between student achievement (mathematics) and educational backgrounds of teachers in grades 4, 5, and 6 |
|---|---|---|---|
| Degree | N | Mean | SD |
| Bachelor's | 58 | 10.97 | 3.5736 |
| Master's | 43 | 9.95 | 3.2947 |
| Master's + 30 hours | 1 | 8.00 | 0.0000 |
| Source | SS | df | MS | F | p |
| Between groups | 31.65 | 2 | 15.83 | 1.32 | .27 |
| Within groups | 1183.84 | 99 | 11.96 | | |
| Total | 1215.49 | 101 | | | |

Results shown in Table 14 failed to indicate a significant difference between student achievement in mathematics in grade 4 and the educational backgrounds of the teachers.

Results shown in Table 15 failed to indicate a significant difference between student achievement in mathematics in grade 5 and the educational backgrounds of the teachers.
TABLE 14.—One-way analysis of variance for determining the extent of relationship between student achievement (mathematics) and educational backgrounds of teachers in grade 4

<table>
<thead>
<tr>
<th>Degree</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bachelor's</td>
<td>19</td>
<td>13.11</td>
<td>3.0349</td>
</tr>
<tr>
<td>Master's</td>
<td>15</td>
<td>11.13</td>
<td>3.0675</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between groups</td>
<td>32.59</td>
<td>2</td>
<td>16.30</td>
<td>1.70</td>
<td>.20</td>
</tr>
<tr>
<td>Within groups</td>
<td>297.52</td>
<td>31</td>
<td>9.60</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Total 330.11 33

TABLE 15.—One-way analysis of variance for determining the extent of relationship between student achievement (mathematics) and educational backgrounds of teachers in grade 5

<table>
<thead>
<tr>
<th>Degree</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bachelor's</td>
<td>17</td>
<td>9.35</td>
<td>2.7826</td>
</tr>
<tr>
<td>Master's</td>
<td>14</td>
<td>10.14</td>
<td>2.9835</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between groups</td>
<td>4.79</td>
<td>2</td>
<td>2.40</td>
<td>.28</td>
<td>.76</td>
</tr>
<tr>
<td>Within groups</td>
<td>239.60</td>
<td>28</td>
<td>8.56</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Total 244.39 30

Results shown in Table 16 failed to indicate a significant difference between student achievement in mathematics in grade 6 and the educational backgrounds of the teachers.

For Hypothesis 8, results were reported for all grades (4, 5, and 6) and for each grade separately.
TABLE 16.—One-way analysis of variance for determining the extent of relationship between student achievement (mathematics) and educational backgrounds of teachers in grade 6

<table>
<thead>
<tr>
<th>Degree</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bachelor's</td>
<td>22</td>
<td>10.36</td>
<td>3.7739</td>
</tr>
<tr>
<td>Master's</td>
<td>14</td>
<td>8.50</td>
<td>3.4807</td>
</tr>
<tr>
<td>Master's + 30 hours</td>
<td>1</td>
<td>8.00</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between groups</td>
<td>32.33</td>
<td>2</td>
<td>16.16</td>
<td>1.20</td>
<td>.31</td>
</tr>
<tr>
<td>Within groups</td>
<td>456.59</td>
<td>34</td>
<td>13.43</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>488.92</td>
<td>36</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Hₐ: There is no significant difference in student ratings of teachers when comparing the educational backgrounds of teachers.

Results shown in Table 17 failed to indicate a significant difference between student ratings of teachers in grades 4, 5, and 6 and the educational backgrounds of the teachers.

TABLE 17.—One-way analysis of variance for determining the extent of relationship between student ratings of teachers and educational backgrounds of teachers in grades 4, 5, and 6

<table>
<thead>
<tr>
<th>Degree</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bachelor's</td>
<td>58</td>
<td>40.67</td>
<td>3.1755</td>
</tr>
<tr>
<td>Master's</td>
<td>43</td>
<td>40.09</td>
<td>2.8935</td>
</tr>
<tr>
<td>Master's + 30 hours</td>
<td>1</td>
<td>44.00</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between groups</td>
<td>20.94</td>
<td>2</td>
<td>10.47</td>
<td>1.12</td>
<td>.33</td>
</tr>
<tr>
<td>Within groups</td>
<td>926.40</td>
<td>99</td>
<td>9.36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>947.34</td>
<td>101</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Thus, the null hypothesis was not rejected at the .05 level of significance.

Results shown in Table 18 did not indicate a significant difference between student ratings of teachers in grade 4 and the educational backgrounds of the teachers.

<table>
<thead>
<tr>
<th>TABLE 18.—One-way analysis of variance for determining the extent of relationship between student ratings of teachers and educational backgrounds of teachers in grade 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Degree</td>
</tr>
<tr>
<td>Bachelor's</td>
</tr>
<tr>
<td>Master's</td>
</tr>
<tr>
<td>Source</td>
</tr>
<tr>
<td>Between groups</td>
</tr>
<tr>
<td>Within groups</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

Results shown in Table 19 did not indicate a significant difference between student ratings of teachers in grade 5 and the educational backgrounds of the teachers.

Results shown in Table 20 failed to indicate a significant difference between student ratings of teachers in grade 6 and the educational backgrounds of the teachers.

For Hypothesis 9, results were reported for all grades (4, 5, and 6) and for each grade separately.

H$_9$: There is no significant difference in principal ratings of teachers when comparing the educational backgrounds of teachers.
TABLE 19.—One-way analysis of variance for determining the extent of relationship between student ratings of teachers and educational backgrounds of teachers in grade 5

<table>
<thead>
<tr>
<th>Degree</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bachelor's</td>
<td>17</td>
<td>41.35</td>
<td>3.6045</td>
</tr>
<tr>
<td>Master's</td>
<td>14</td>
<td>41.29</td>
<td>3.0237</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between groups</td>
<td>0.35</td>
<td>2</td>
<td>0.02</td>
<td>.02</td>
<td>1.00</td>
</tr>
<tr>
<td>Within groups</td>
<td>326.74</td>
<td>28</td>
<td>11.67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>327.09</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TABLE 20.—One-way analysis of variance for determining the extent of relationship between student ratings of teachers and educational backgrounds of teachers in grade 6

<table>
<thead>
<tr>
<th>Degree</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bachelor's</td>
<td>22</td>
<td>39.91</td>
<td>3.3652</td>
</tr>
<tr>
<td>Master's</td>
<td>14</td>
<td>38.43</td>
<td>3.0055</td>
</tr>
<tr>
<td>Master's + 30 hours</td>
<td>1</td>
<td>44.00</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between groups</td>
<td>39.94</td>
<td>2</td>
<td>19.97</td>
<td>1.91</td>
<td>.16</td>
</tr>
<tr>
<td>Within groups</td>
<td>355.25</td>
<td>34</td>
<td>10.45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>395.19</td>
<td>36</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Results shown in Table 21 failed to indicate a significant difference between principal ratings of teachers in grades 4, 5, and 6 and the educational backgrounds of the teachers.

Results shown in Table 22 failed to indicate a significant difference between principal ratings of teachers in
TABLE 21.—One-way analysis of variance for determining the extent of relationship between principal ratings of teachers and educational backgrounds of teachers in grades 4, 5, and 6

<table>
<thead>
<tr>
<th>Degree</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bachelor's</td>
<td>58</td>
<td>44.21</td>
<td>4.5141</td>
</tr>
<tr>
<td>Master's</td>
<td>43</td>
<td>44.47</td>
<td>5.0112</td>
</tr>
<tr>
<td>Master's + 30 hours</td>
<td>1</td>
<td>43.00</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between groups</td>
<td>3.36</td>
<td>2</td>
<td>1.68</td>
<td>.08</td>
<td>.93</td>
</tr>
<tr>
<td>Within groups</td>
<td>2216.22</td>
<td>99</td>
<td>22.39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>2219.58</td>
<td>101</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TABLE 22.—One-way analysis of variance for determining the extent of relationship between principal ratings of teachers and educational backgrounds of teachers in grade 4

<table>
<thead>
<tr>
<th>Degree</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bachelor's</td>
<td>19</td>
<td>43.89</td>
<td>5.7532</td>
</tr>
<tr>
<td>Master's</td>
<td>15</td>
<td>43.67</td>
<td>4.5930</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between groups</td>
<td>0.44</td>
<td>2</td>
<td>0.22</td>
<td>.76</td>
<td>.99</td>
</tr>
<tr>
<td>Within groups</td>
<td>891.12</td>
<td>31</td>
<td>28.75</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>891.56</td>
<td>33</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

grade 4 and the educational backgrounds of the teachers.

Results shown in Table 23 failed to indicate a significant difference between principal ratings of teachers in grade 5 and the educational backgrounds of the teachers.

Results shown in Table 24 did not indicate a significant difference between principal ratings of teachers in grade 6
TABLE 23.—One-way analysis of variance for determining the extent of relationship between principal ratings of teachers and educational backgrounds of teachers in grade 5

<table>
<thead>
<tr>
<th>Degree</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bachelor's</td>
<td>17</td>
<td>44.41</td>
<td>3.7924</td>
</tr>
<tr>
<td>Master's</td>
<td>14</td>
<td>45.93</td>
<td>6.0569</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between groups</td>
<td>17.66</td>
<td>2</td>
<td>8.84</td>
<td>.35</td>
<td>.71</td>
</tr>
<tr>
<td>Within groups</td>
<td>707.05</td>
<td>28</td>
<td>25.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>724.71</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TABLE 24.—One-way analysis of variance for determining the extent of relationship between principal ratings of teachers and educational backgrounds of teachers in grade 6

<table>
<thead>
<tr>
<th>Degree</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bachelor's</td>
<td>22</td>
<td>44.32</td>
<td>3.9807</td>
</tr>
<tr>
<td>Master's</td>
<td>14</td>
<td>43.86</td>
<td>4.2762</td>
</tr>
<tr>
<td>Master's + 30 hours</td>
<td>1</td>
<td>43.00</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between groups</td>
<td>3.08</td>
<td>2</td>
<td>1.54</td>
<td>.01</td>
<td>.91</td>
</tr>
<tr>
<td>Within groups</td>
<td>570.49</td>
<td>34</td>
<td>16.78</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>573.57</td>
<td>36</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

and the educational backgrounds of the teachers.

H₀: There is no significant difference between student achievement (reading) in grades 4, 5, and 6.

Results shown in Table 25 failed to indicate a significant difference between student achievement in reading in grades 4, 5, and 6, the F value being 15.69 and the
TABLE 25.—One-way analysis of variance for determining the extent of relationship between student achievement in reading in grades 4, 5, and 6

<table>
<thead>
<tr>
<th>Grade</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>34</td>
<td>10.56</td>
<td>2.9149</td>
</tr>
<tr>
<td>5</td>
<td>31</td>
<td>8.42</td>
<td>1.5869</td>
</tr>
<tr>
<td>6</td>
<td>37</td>
<td>7.57</td>
<td>2.1543</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between groups</td>
<td>165.81</td>
<td>2</td>
<td>82.91</td>
<td>15.69</td>
<td>.00</td>
</tr>
<tr>
<td>Within groups</td>
<td>523.01</td>
<td>99</td>
<td>5.28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>688.82</td>
<td>101</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Probability being .00.

H₁₁: There is no significant difference between student achievement (mathematics) in grades 4, 5, and 6.

Results shown in Table 26 failed to indicate a significant difference between student achievement in mathematics.

TABLE 26.—One-way analysis of variance for determining the extent of relationship between student achievement in mathematics in grades 4, 5, and 6

<table>
<thead>
<tr>
<th>Grade</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>34</td>
<td>12.24</td>
<td>3.1628</td>
</tr>
<tr>
<td>5</td>
<td>31</td>
<td>9.71</td>
<td>2.8542</td>
</tr>
<tr>
<td>6</td>
<td>37</td>
<td>9.60</td>
<td>3.6853</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between groups</td>
<td>152.07</td>
<td>2</td>
<td>76.03</td>
<td>7.08</td>
<td>.00</td>
</tr>
<tr>
<td>Within groups</td>
<td>1063.42</td>
<td>99</td>
<td>10.74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1215.49</td>
<td>101</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
in grades 4, 5, and 6, the F value being 7.08 and the probability being .00.

H₁₂: There is no significant difference between student ratings of teachers in grades 4, 5, and 6.

Results shown in Table 27 failed to indicate a significant difference between student ratings of teachers in grades 4, 5, and 6, the F value being 3.54 and the probability being .03. Therefore, Hypothesis 12 could not be accepted at the .05 level of significance.

**TABLE 27.—One-way analysis of variance for determining the extent of relationship between student ratings of teachers in grades 4, 5, and 6**

<table>
<thead>
<tr>
<th>Grade</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>34</td>
<td>40.76</td>
<td>2.2165</td>
</tr>
<tr>
<td>5</td>
<td>31</td>
<td>41.32</td>
<td>3.3004</td>
</tr>
<tr>
<td>6</td>
<td>37</td>
<td>39.45</td>
<td>3.3132</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between groups</td>
<td>63.26</td>
<td>2</td>
<td>31.63</td>
<td>3.54</td>
<td>.03</td>
</tr>
<tr>
<td>Within groups</td>
<td>884.08</td>
<td>99</td>
<td>8.93</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>947.32</td>
<td>101</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

H₁₃: There is no significant difference between principal ratings of teachers in grades 4, 5, and 6.

Results shown in Table 28 failed to indicate a significant difference between principal ratings of teachers in grades 4, 5, and 6. Therefore, Hypothesis 13 could not be rejected at the .05 level of significance.
TABLE 28.—One-way analysis of variance for determining the extent of relationship between principal ratings of teachers in grades 4, 5, and 6

<table>
<thead>
<tr>
<th>Grade</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>34</td>
<td>43.79</td>
<td>5.1978</td>
</tr>
<tr>
<td>5</td>
<td>31</td>
<td>45.10</td>
<td>4.9150</td>
</tr>
<tr>
<td>6</td>
<td>37</td>
<td>44.11</td>
<td>3.9915</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between groups</td>
<td>29.74</td>
<td>2</td>
<td>14.87</td>
<td>.67</td>
<td>.51</td>
</tr>
<tr>
<td>Within groups</td>
<td>2189.83</td>
<td>99</td>
<td>22.12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>2219.57</td>
<td>101</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Chapter V contains a summary of the study, conclusions, implications indicated by the results of analyses reported in this chapter, and recommendations for further research.
CHAPTER V

SUMMARY, CONCLUSIONS, IMPLICATIONS, AND RECOMMENDATIONS

In this chapter, the study has been summarized and the results discussed under four major headings: (1) Summary of the Study, (2) Conclusions, (3) Implications, and (4) Recommendations.

Summary of the Study

The purpose of the present study was to investigate relationships between student ratings of teachers, principal ratings of teachers, and student achievement in reading and mathematics. Student achievement, student ratings of teachers, and principal ratings were factors treated to determine possible relationships with teacher effectiveness. The study was undertaken in the Kalamazoo school system.

The student population consisted of approximately 2,856 students in grade levels 4, 5, and 6. The teacher population included 102 teachers, 91 of whom were white and 11 were black. In the three grades studied, there were 34 fourth-, 31 fifth-, and 37 sixth-grade teachers. Also included in the study were 10 elementary principals, 9 of whom were white and 1 was black.

The school system had a racial composition of about
20 percent black and about 80 percent white. The following criteria were used to select teachers for the study: (1) elementary teachers in grades 4, 5, and 6; (2) elementary teachers with student ratings; (3) elementary teachers with principal ratings; and (4) elementary teachers with student achievement data in reading and mathematics. Data about teachers that met these criteria were obtained from the Research and Development Department of the school system. Student ratings were collected by using the Elementary Teacher Image Questionnaire (ETIQ). The ratings for each class were combined into a composite score that was used as a criterion of a class's perception of the teacher. The principal rating for each teacher consisted of a composite synthesized from the ratings for each of the Teacher Evaluation Instrument (TEI) items. Student achievement data were obtained from administration of the Metropolitan Achievement Test (MAT) on a pre- and posttest basis. The pretests and posttests were administered in September and May of the school year. All information relevant to data collection was received from the Research and Development Department of the school system.

Coefficients of correlation and a one-way analysis of variance were used for determining the nature and extent of relationships. These statistical analyses pertained to the following questions:
1. What is the relationship between student achievement (reading) and student ratings of teachers?

2. What is the relationship between student achievement (mathematics) and student ratings of teachers?

3. What is the relationship between student achievement (reading) and principal ratings of teachers?

4. What is the relationship between student achievement (mathematics) and principal ratings of teachers?

5. What is the relationship between student achievement and educational backgrounds of teachers?

6. What is the relationship between student ratings of teachers and principal ratings of teachers?

The specific hypotheses tested by the present study were discussed in Chapter IV.

The major finding was that statistically significant correlations were not found between student ratings of teachers, principal ratings of teachers, and student achievement in reading and mathematics in grades 4, 5, and 6.

The statements that follow summarize some of the other findings:

1. Student achievement in mathematics and reading in grade 4 was significantly related to principal ratings of teachers in grade 4.

2. There was a significant relationship between student achievement in reading growth and principal ratings of teachers in grade 5.
3. In grade 6, a negative correlation was found between student achievement in reading growth and principal ratings of teachers.

4. There was also a statistically significant correlation between student ratings of teachers and principal ratings of teachers in grade 6.

5. A significant difference was not found between student achievement in reading and mathematics growth in grades 4, 5, and 6 and educational backgrounds of teachers.

6. A significant difference was not found between student ratings of teachers and educational backgrounds of teachers in grades 4, 5, and 6.

7. A significant difference was not found between principal ratings of teachers and educational backgrounds of teachers in grades 4, 5, and 6.

8. The gains made in achievement in reading and mathematics differed significantly between grades 4 and 5 and grades 5 and 6. The greatest gains were made in grades 4 and 5 and least in grade 6.

9. There was a significant difference among student ratings of teachers in grades 4, 5, and 6. Student ratings were highest in grades 4 and 5 and lowest in grade 6.

10. Significant differences were not found among principal ratings of teachers in grades 4, 5, and 6.

Conclusions

In the present study, significant relationships were not found among student ratings of teachers, principal ratings of teachers, and student achievement or growth in reading and mathematics when data for 102 elementary
teachers in grades 4, 5, and 6 were analyzed collectively. However, there were statistically significant correlations reported in grade 4 between principal ratings of teachers and both reading and mathematics growth. In grade 5, there was a statistically significant relationship between reading growth and principal ratings of teachers. In grade 6, a significant negative relationship was reported between student growth in reading and principal ratings of teachers. This finding may be viewed from two dimensions: (1) high reading growth and low principal ratings or (2) high principal ratings and low reading growth. The opposite effect probably is due to the principal's lack of knowledge as to what teacher performances are necessary to foster high reading achievement. For high reading growth, teachers must have knowledge as to what should be taught, whereas the principal may not be familiar with instruction at this level. In grade 6, there was a significant positive relationship between student ratings of teachers and principal ratings of teachers. Student growth in reading and mathematics was greatest at grades 4 and 5 and lowest at grade 6. Based on a review of the literature, the researcher anticipated a significant correlation between principal ratings of teachers with both academic measures of growth at each grade level.

Educational backgrounds of teachers were not found to be significantly related to gains in reading and mathematics.
in grades 4, 5, and 6. Thus, the degrees held by the teachers did not seem to be an influencing factor in student achievement in these areas. This tended to support the literature that suggested that teachers should be rated on observed performance rather than on the degrees they hold.

In concluding, the findings of this study suggest that teaching is multidimensional and should be evaluated from different dimensions. Consequently, teacher assessment models should be designed to measure dimensions in which teachers perform.

Implications

Considerable research has been directed toward identifying and measuring teacher performance. The method most commonly used to assess the teacher-learning process is principal rating.

The major finding of this study is that statistically significant correlations were not found among student ratings of teachers, principal ratings of teachers, and student achievement for 102 teachers in grades 4, 5, and 6. Another finding in this study is that principal ratings of teachers were related to gains in reading and mathematics in grade 4. Student ratings of teachers were significantly related to principal ratings of teachers in grade 6. Since inverse relationships were found between student gains in reading and principal ratings of teachers, the researcher believes
that principals need further training in the teaching of reading and mathematics at the later elementary level in order to better assess teacher performance.

There are implications to suggest that principals should give more attention to evaluating teacher performance in reading and mathematics instruction at all grade levels, become familiar with content or curriculum areas needed to promote high academic growth, and develop leadership abilities for evaluating teacher performance.

Consistent with the findings of other studies, students in the lower grades rated teachers higher than in the upper grade. The ratings in grades 4 and 5 were similar, but higher than those in grade 6. This study did not reveal the reasons for the difference.

Based on findings in the study, principals need to be qualified in identifying and measuring teacher performance. Elementary principals should have educational leadership training to become more efficient building leaders. Teachers should have more practical graduate training in the teaching of reading and mathematics together with training in educational leadership in order to produce efficient, capable students in both disciplines at the later elementary level.
Recommendations

Further research related to the present study might include:

(1) Studies to examine principal ratings by race and sex to determine if white principals rate teachers differently than do black principals.

(2) Studies to examine instructional evaluation instruments to determine if characteristics used are appropriate for measuring reading and mathematics achievement.

(3) Studies to examine principal leadership training and expertise in evaluating teacher performance in reading and mathematics both at the university level and in the county school system.

(4) Studies to determine the nature of the relationships between student ratings of teachers and principal ratings of teachers.

(5) Studies to examine the disparities in achievement in reading and mathematics among grades 4, 5, and 6.

Based on the findings of this study, the researcher proposes the following model to be used by public schools for improving student achievement and for improving teacher and principal performance at this level:

The Model

School systems may establish an Appraisal Commission—racially balanced and representing selected interest groups—for the purpose of determining:

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
1. Instructional performance criteria (standards for teachers and principals)

2. Instructional performance objectives (job targets for teachers and students)
   a. Principals and teachers
   b. Students

3. Performance activities
   a. Deciding who should be involved in the evaluation process
   b. Selecting the criteria for instructional evaluation of teacher performance
   c. Using instructional evaluation instruments for teacher performance

4. Utilization of evaluation results
   a. School, home, and community
   b. Professional community and publications

5. Professional growth
   a. Principal and teacher needs
      1) Graduate training in the instruction of teaching reading and mathematics
      2) Graduate training in educational leadership, with emphasis on evaluation
   b. Student and community needs
APPENDIX A

Elementary Teacher Image Questionnaire
Elementary Teacher Image Questionnaire

Instructions for the Student

Please do not begin until you are told to do so by the person in charge. Write your teacher's name on the line at the upper right-hand corner of this page. Do not put your name or any other information on this page. After you read each question and decide on your answer, blacken the appropriate number with a No. 2 pencil. Use the following scale when answering the questions:

1 = Never (N)
2 = Almost Never (AN)
3 = Sometimes (S)
4 = Almost Always (AA)
5 = Always (A)

EXAMPLE

A. My teacher is on time . . . 1 2 3 4 5

If your response to this question was "Almost Always," you would blacken number 4, as shown.

Be sure to blacken only one number for each question.
<table>
<thead>
<tr>
<th></th>
<th>WHAT ARE YOUR OPINIONS CONCERNING THIS TEACHER:</th>
<th>N</th>
<th>AN</th>
<th>S</th>
<th>AA</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>My teacher explains things so that I understand them</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>My teacher treats everyone fairly</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>The kids in my class behave</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>My teacher is friendly toward all students</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>My teacher makes school seem fun and interesting</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>My teacher laughs and enjoys jokes in the classroom</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>My teacher enjoys teaching school</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>8</td>
<td>My teacher lets students tell about their ideas in class</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>9</td>
<td>I do interesting things in my teacher's class</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>10</td>
<td>My teacher looks neat and dresses nicely</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>11</td>
<td>My teacher controls his/her anger</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>12</td>
<td>My teacher is able to answer questions about school work</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>13</td>
<td>My teacher thinks what I say is important</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>14</td>
<td>My teacher likes me</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>15</td>
<td>My teacher wants me to ask questions and give my ideas in class</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>16</td>
<td>I feel free to give my ideas in class</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>17</td>
<td>I feel like I learn a lot in class</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>18</td>
<td>I think my teacher is a good teacher</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>
APPENDIX B

Teacher Evaluation Instrument
Teacher Evaluation Instrument

1st appraisal ___________
2nd appraisal ___________
3rd appraisal ___________
4th appraisal ___________

Teacher_________________ Subject or grade_______________
School_________________ Principal_____________________
Date__________________

Rating scale defined

1 = POOR (unacceptable performance)
2 = FAIR (needs improvement—does not meet minimum performance requirements)
3 = SATISFACTORY (meets performance requirements)
4 = GOOD (exceeds performance requirements)
5 = EXCELLENT (far exceeds performance requirements)

Based on your (principal's) appraisal of this teacher's performance, check the appropriate response. (If you believe a certain criterion does not apply, do not check a response category.)

A. PROFESSIONAL PERFORMANCE

<table>
<thead>
<tr>
<th>Criterion</th>
<th>P</th>
<th>F</th>
<th>S</th>
<th>G</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Shows cooperation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Displays a positive relationship with faculty</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Demonstrates effective daily preparation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Reacts rationally to constructive criticism</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Demonstrates initiative and resourcefulness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Exhibits flexibility in meeting changes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Displays promptness and dependability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. BEHAVIOR TOWARD PUPILS</td>
<td>P</td>
<td>F</td>
<td>S</td>
<td>G</td>
<td>E</td>
</tr>
<tr>
<td>--------------------------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>1. Shows interest in pupils</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Solicits pupils' point of view</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Secures cooperation of pupils</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| C. SCHOLARSHIP | |
|----------------|---|---|---|---|---|
| 1. Demonstrates knowledge of subject matter |   |   |   |   |   |
| 2. Demonstrates skills in applying knowledge |   |   |   |   |   |

| D. CLASSROOM PERFORMANCE | |
|---------------------------|---|---|---|---|---|
| 1. Exhibits effective classroom control and management |   |   |   |   |   |
| 2. Handles problems of discipline effectively |   |   |   |   |   |
| 3. Adapts materials to needs and interests of students |   |   |   |   |   |
| 4. Gives definite, reasonable assignments |   |   |   |   |   |
| 5. Demonstrates skill in directing supervised study |   |   |   |   |   |
| 6. Provides individual assistance to students |   |   |   |   |   |
| 7. Uses a variety of instructional materials |   |   |   |   |   |
| 8. Uses a variety of learning experiences |   |   |   |   |   |
| 9. Displays an awareness of the emotional atmosphere of the classroom |   |   |   |   |   |
| 10. Identifies learning opportunities |   |   |   |   |   |

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
11. Involves pupils in planning
12. Provides for individual differences
13. Diagnoses pupil needs accurately
14. Evaluates pupil progress effectively

E. PERSONAL BEHAVIOR
1. Shows enthusiasm and interest
2. Demonstrates emotional stability
3. Exhibits poise and self-confidence
4. Displays a willingness to share ideas and materials
5. Exhibits a willingness to seek help when needed

F. EFFECTIVENESS
1. What is your overall evaluation of this teacher's effectiveness? (This is not necessarily based on the average of the preceding criteria.)

G. EXPLAIN THOSE AREAS RATED BELOW SATISFACTORY (be specific—if necessary, attach another page):

H. SPECIFIC SUGGESTIONS FOR IMPROVEMENT
I. TEACHING IN

_____ Major Certified Area
_____ Minor Certified Area
_____ Other (explain below)

J. Approximate number of students in class _______

K. Is the physical setting and room size adequate? ______

L. Number of different teacher preparations _______

Disposition recommended:

____ a. Reemploy
____ b. Terminate
____ c. Extend probation one year
____ d. Do not place on tenure
____ e. Place on tenure

A conference regarding this appraisal must be held with the teacher. The signature does not necessarily indicate agreement with above appraisal, but indicates he/she has seen it. The teacher may attach a memo explaining his or her disagreement regarding any aspect of this evaluation.

__________________________  ___________________________
Teacher's Signature        Principal's Signature

**FINAL EVALUATION SHALL BE COMPLETED AND SIGNED BY MAY 1.

For Employee Relations
Division use:
Date tenure granted ________
REFERENCES


Biondolillo, T. C. Principal's role. Instructor, 1972, 81, 39.

Booth, J. Kalamazoo desegregation study--Phase II. Kalamazoo, Mich.: Western Michigan University, 1972.


Misner, P. J., Schneider, F. W., & Keith, L. G. Elementary School Administration. Columbus, O.: Charles E. Merrill, 1963.


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.


Smeltzer, C. H., & Harter, R. S. Comparison of anonymous and signed ratings of teachers. Educational Outlook, 1934, 8, 76-84.


Wax, M. L. How should schools be held accountable? In P. A. Olson, L. Freeman, & J. Bowman (Eds.), Education for 1984 and after. Chicago: Directorate of the Study Commission on Undergraduate Education and the Education of Teachers, 1971.

Williams, R. L. Abuses and misuses in testing black chil-
dren. Ebony, August 1974, pp. 77-90.

Wilson, D. T. A study of factors related to student ratings
of teachers. Unpublished doctoral dissertation, Western
Michigan University, 1971.

Yamamoto, T. K. Creative thinking and teacher effectiveness:
A review of research. Journal of School Psychology,
1964, 2, 70-71.