DOCUMENT RESUME

ED 108 105 CS 001 813

AUTHOR Willard, Richard W.
TITLE Success and Failure in Reading Performance among Massachusetts City Schools: A Study for the Massachusetts Advisory Council on Education.
INSTITUTION Educational Research Corp., Watertown, Mass.
SPONS AGENCY Massachusetts Advisory Council on Education, Boston.
PUB DATE Feb 75
NOTE 52p.

EDRS PRICE MF-$0.76 HC-$3.32 PLUS POSTAGE
DESCRIPTORS *Effective Teaching; Elementary Grades; Inner City; Reading; *Reading Achievement; Reading Development; *Reading Programs; *Reading Research; Reading Skills; *Urban Schools
IDENTIFIERS *Massachusetts

ABSTRACT

The purpose of this study was to document the workings of successful inner-city schools in teaching children to read. The study schools were identified by gathering poverty data on the proportion of the attending students who received free lunch or free milk and the proportion of children in the attendance areas who were from low income families. Each measure of poverty was used to rank elementary schools in ten Massachusetts cities, and the average of the two rank values was used as the final ranking from high to low poverty. Average test scores in reading were added to the ranked list of schools in order to identify those poverty schools whose students read at or above grade level. Successful schools were also identified and matched with another school on the two poverty measures. The project staff drew upon eight factors identified by George Weber to provide labels for the following factors: leadership, coordination, additional reading personnel, atmosphere, individualization, evaluation, high expectations, strong emphasis on reading, use of phonics, staff training and experience, and quality of teaching. The analysis of data failed to uncover consistent differences between the ratings of the two groups of schools. (WR)

SUCCESS AND FAILURE IN READING PERFORMANCE AMONG MASSACHUSETTS CITY SCHOOLS

A Study for the Massachusetts Advisory Council on Education

Richard W. Willard

Educational Research Corporation
85 Main Street
Watertown, Massachusetts 02172

February 1975

CONTENTS

FOREWORD

When the Massachusetts Advisory Council on Education contracted with Educational Research Corporation for a study aimed at documenting the workings of "successful" inner-city schools in teaching children to read, the complications of conducting such a study were only dimly comprehended. The Council was aware that there would be definition problems with such words and phrases as "inner-city" and "successful" schools and practical problems with gaining access to "unsuccessful" schools to see how they contrasted with their "successful" counterparts. However, as this technical report shows, these problems were only the tip of the proverbial iceberg!

Hidden from view, when the study was launched over two years ago, were such problems as:

- -developing valid measures of poverty (so as to identify inner-city schools) and locating reliable data to support these measures.
- -making different measures of "success" comparable and deciding the basis for selecting "unsuccessful" (or contrast) schools.
- -selecting and developing operational definitions for the factors (presumably having a bearing on the "successful" teaching of reading) to investigate.
- -designing appropriate procedures and instruments to elicit valid and reliable data on the study schools.

It is to the credit of the study team, led by Dr. Richard Willard, that they not only faced these problems frankly, honestly and patiently but also demonstrated a willingness to change and revise procedures and plans as the need arose. As a result, the reader will find the study's methodology and instrumentation clearly spelled out so as to allow judgments on the study's research adequacy to be made. To many of us associated with the study, the approach and procedures used are superior to any studies of this type yet undertaken in the country.

One cautionary note. For readers to use the study instruments employed in this study without undertaking the changes and preprations advised by the study team can only lead to invalid results and wasted effort. Nevertheless, the Council hopes readers will find in the report (and the report summary) enough ideas and suggestions on how schools might proceed to warrant moving in the directions recommended by the study team. MACE would certainly want to hear from readers who decide to move in such directions.

Allan S. Hartman
Associate Director
Advisory Council on Education

Looking to research for guidance in how to teach children to read has proven fruitless for many who are concerned about reading. Some research, in fact, suggests that there is hardly anything that schools can do, since home influences are so dominant in learning to read. The study directed by James Coleman for the U.S. Office of Education concluded that schools have little influence separable from the backgrounds of their students. A similar study conducted by the International Association for the Evaluation of Educational Achievement (IEA) found very little evidence of the impact of schools on reading. Many researchers have concluded therefore that schools do not make a difference.

There are other researchers, though, who conclude that schools do make a difference. Some have used survey techniques to identify school factors associated with reading performance.

Guthrie, in providing a summary of several such studies, noted that there is little doubt that schools are important, but the survey studies still do not suggest what schools must do to succeed. A different approach from the use of a survey was taken by George Weber and described in his monograph Inner-City Children Can be Taught to Read: Four Successful Schools. Weber searched for and found four city elementary schools whose students were at the national norms in reading. After visiting each school Weber identified eight factors that the schools shared and that seemed to explain their success: strong leadership, high expectations, good atmosphere, strong emphasis on reading, additional reading personnel, use of phonics, individualization, and careful evaluation of pupil progress.

The approach used by Weber to identify successful schools and to look at them was recognized by the Massachusetts Advisory Council on Education (MACE) to be an appropriate approach to be used in the Commonwealth to identify what factors are associated with reading success in Massachusetts city schools. MACE commissioned Educational Research Corporation (ERC) to search for successful schools in Massachusetts cities and to identify what these schools did to become successful.

Identifying the Study Schools

By using the 1973 allocations of funds from the Federal ESEA Title I program the study staff identified the ten Massachusetts cities with the largest allocations and thereby the largest numbers of poverty students. Ultimately nine of these cities agreed to participate in the study. (Note 1)

ERC gathered poverty data, (which consisted of (1) the proportion of children in an attendance area who according to Title I applications were from low income families and (2) the proportion of the attending students who were recipients of free lunch or free milk) on each elementary school in the nine cities. These two measures of poverty are of course fallible. It is the case, for example, that cities differ in the ways they compute the proportion of children from low income families, some using census data, some using welfare rolls, and some a combination of both, and that the identification of students eligible for free lunch or free milk programs is frequently subject to misreporting of income by parents and the judgment by the principals of those in need but otherwise not identified. The combination of these two measures, however, is still better than either is alone, and so serves as a generally valid

measure of poverty in the schools. Each measure of poverty was used to rank the elementary schools in each city, and the average of the two rank values was used as the final ranking of the schools from high to low poverty. This ranking was designed to identify poverty schools by means of data that are consistent across cities, but evidence collected later in the study showed that the poverty classifications were not as uniform as they appeared to be at first and that the identification of poverty is itself a complex matter.

Average test scores in reading were added to the ranked list of schools in order to identify those poverty schools whose students were reading at or above grade level according to standardized tests. These test scores were from existing records of city-wide testing programs rather than a special administration of tests since the Study Review Committee strongly urged that poverty children not be subjected to any more tests. They are already tested extensively, and one more test would prove both burdensome and redundant.

The test results were derived from the scores of sixth graders because sixth grade test results draw upon the cumulative learning of reading throughout the elementary school years, and a student scoring at grade level in his last grade in elementary school has probably been progressing successfully through most of his earlier years in school. Students may experience different sequences of reading instruction or different approaches to reading that might result in some early deviation from the norm, but if they converge later upon the norm in the sixth grade the school should still be judged to be successful.

ERC focused the inspection of sixth grade scores among the poverty schools where the general performance was low. In the nine cities there were, nevertheless, ten poverty schools that had students performing at the national norms or better, and the schools therefore stood out from the rest. These ten schools represented a cross-section of city schools. They were generally situated close to the center of the cities, which were of somewhat different sizes, with school populations varying from approximately 10,000 to 100,000 students. Most of these successful schools were in old, traditional buildings with self-contained classrooms, but some were in newer buildings with open spaces. Most were neighborhood schools, but some had students who were bused. Varied as they were, the ten schools constituted a target group to be studied in detail to uncover what characteristics might have helped to set them apart on test scores.

Identifying Contrast Schools

Identifying characteristics of successful schools, no matter how carefully done, does not necessarily explain what those schools did to become successful. To find what they did to become successful would require a study of schools over an extended period of time as they passed from a failing status to a successful one. Without the advantage of such a study over time it is critical to identify which characteristics of successful schools are not also shared by unsuccessful schools. These characteristics might explain what the schools did to become successful because they are different.

In order to highlight factors on which the successful schools differ the staff identified a set of contrast schools which were not successful. Schools can differ in many ways, some of which are external to what the school does, so the contrast schools were selected deliberately

to be similar to the successful schools in terms of the poverty levels of their children. For each successful school another school was found that matched on the two poverty measures, Title I and free lunch percentages, and was similar also in racial composition and the proportion of bilingual students. Comparable to the successful schools on measures of poverty as well as racial and bilingual composition, the contrast schools had reading scores that were, on the average, 1.3 grade equivalents below national norms and thereby quite below the successful schools. (Note 2)

Studying the Schools

Since the two sets of schools differed on test scores but matched on a number of external measures, some other factors were needed to explain how they differed in the performance of their students.

Possible factors for study were the sets of variables that had been identified in one or another of the surveys, as, Coleman, Guthrie, and others, but those variables were not consistent correlates of student success. Further, those variables tend to be quantitative, as, per pupil expenditure, class sizes, and teacher aptitude, and they fall short of the qualitative dimensions that better describe what actually occurs in schools and that better describe the direct influences upon students. The studies by George Weber and others identified variables that go beyond the quantitative to include qualitative dimensions of schools, and these were judged to provide a richer source of factors to be studied.

The project staff drew upon the eight factors identified by George Weber to provide the labels for factors to be studied in this project, but it was deemed important to define the factors with a strong emphasis upon the processes involved. The Weber factors were richer than simple

quantitative dimensions because they dealt with school processes, but the project staff elected to stress procedures even more. Thus, for example, the staff went beyond the identification of additional reading personnel, a Weber factor, to the consideration of what roles and functions are performed by the reading personnel. The Weber factors were redefined in operational terms; some were altered; and some were split up. It is not important here to identify where and how changes were made because the factors studied can be described in sufficient detail that they stand alone for scrutiny. Below are given the descriptions of the factors used in this study.

A. Leadership

When there is strong leadership in reading instruction in a school, the staff will agree unanimously on who provides the leadership--whether it be the principal, a school reading specialist or someone from the central office--and observers will readily detect who provides the direction to the teaching of reading. A leader will display inspiration, empathy, and flexibility in providing the staff encouragement to do its very best.

B. Coordination

Good coordination in the teaching of reading means that students experience across and within grade levels activities that reinforce each other. Work at any grade is related to work in previous grades, and the several supplementary reading services, remedial, Title I, or learning disability, which take place outside the classroom are still related to what transpires inside the classroom. Coordination can be achieved simply by standardizing all reading activities, but a school with many varied learning activities must see that the activities are well "orchestrated".

C. Additional Reading Personnel

Additional reading personnel, that is other than regular classroom teachers, become significant when there is a variety of personnel: school reading specialists; Title I personnel; learning disability specialists; aides and others to work with students singly or in small groups. Even the school librarian can be an additional resource for reading. Good use of personnel means also that reading specialists share their expertise with teachers by being a resource for reading instruction in classrooms.

D. Atmosphere

Good atmosphere in a school means that people in the school are relaxed and without tension, that the school is operating in an orderly fashion, that the students are purposeful in their activities, and that any noise is normal with no evidence of disruptive clatter.

E. Individualization

A truly individualized program responds to individual differences in background, learning styles, and rates of learning by means of diagnostic procedures and prescriptions that differ by individual either in curriculum materials used or in study times required or in a combination of both.

F. Evaluation

Sound evaluation of pupil progress employs several measurement techniques, as, teacher constructed tests, curriculum tests, criterion-referenced tests, standardized tests, and so on. The evaluations are most effective when reading progress records follow the student from grade to grade and from teacher to teacher and when they are used in developing instruction strategies.

G. High Expectations

When the staff of a school has high expectations for students in general and for the students in that school in particular, high standards will be set for the students, and encouragement will be given for good performance.

H. Strong Emphasis on Reading

A school that places a strong emphasis upon reading devotes ample time to reading instruction and makes available a number and variety of reading materials. A priority for reading usually means that reading is taught at the beginning of the day when children are most alert.

I. Use of Phonics

A strong use of phonics in reading instruction means that decoding skills are developed in the early grades and that phonics is an integral part of the curriculum as represented in the instructional materials used, basal or supplementary.

J. Staff Training and Experience

A well trained and experienced staff will have had extensive formal education including graduate study and special courses in reading instruction either in college or in an in-service program and will have been working in education professionally for some time. There will be evidence that the staff stays contemporary.

K. Quality of Teaching

A good teacher will manage a classroom well such that students will be productive during study time and will be supportive of individual students particularly as they make mistakes or falter.

Visiting the Schools

The framework of factors to be studied made it clear to ERC that there should be many methods of collecting data, and, particularly, that interviews and direct observations should be made during visits to the schools. School visits can produce reports that are affected by reporter bias, so ERC ensured that no visitor would know whether the school was successful or not and that the visitor could not produce reports of observations that were biased by any personal expectations about factors of success. ERC further decided to use a team of visitors to each school so as not to be dependent upon a single visitor's perceptions. Each team of visitors had three members, some of who visited twice in order to provide a total of five visits to each school. One member of the team who visited twice was a reading expert selected to have general knowledge of the principles of reading instruction and, as well, familiarity with schools and how reading is actually taught. Thus, every school was visited by a reading expert. Two visits were made by a research associate from the staff of ERC. The fifth visit to each school was usually made by the assistant project director to give continuity across the different visiting teams. The five visits to a school were planned on different days to provide experiences that were as different as possible and to reduce the effects of special occasions, as, for example, field trips and assemblies, that occur in schools.

Five classrooms were selected randomly by ERC, and each was visited on two different occasions by a different observer. This produced a total of ten reading periods observed in each school. The five classrooms were selected to have one from each of the grades one through five.

Collecting Data

ERC used observation schedules in the classrooms that were designed to collect data relevant to the study factors. Simple data included the number of students and number of aides present. Notes were made about the evidence of a reading area in the room and about the availability and accessibility of various reading materials. Rankings and notes were made about the classroom atmosphere and about the interactions between teachers and students. On four different occasions during the reading period the observer systematically checked on each child in the room and indicated on a form what sort of activity the child was engaged in. When tallied, these observations showed, for example, how many children were involved in reading instruction with a teacher, with an aide, with other students, or alone. The tallies also showed when work was being done on non-reading activites. The combination of all the counts provided a Learner Activity Index showing how student time was distributed over the several activities.

In addition to observing classrooms, the visitors conducted structured interviews. Each of the five teachers who had been observed was interviewed after one of the observations, and the interview protocol provided several questions, particularly questions that related to leadership, coordination, and individualization. The reading expert interviewed the school reading specialist on one day and on another day the school librarian, whenever there was one. ERC staff visitors interviewed the building principal as well as the person or persons in the school district central office most responsible for the reading program in the school system.

Even though protocols were structured for interviews and observations, the visitors were encouraged to use additional space on the forms to describe any findings that were worthy of note. These non-structured comments revealed, for example, that in one school the staff expressed the feeling that the students were protagonists who must be controlled and that the staff was compelled to constantly monitor movement in the hallways.

The observations and interviews provided rich descriptions of the schools, and the study staff collected statistical data as well so that the combination of clinical and statistical data would give in the aggregate richer information than either source alone. The study staff prepared a number of questionnaires that allowed respondents to provide additional data about the multiplicity of factors being studied. The principal, the school reading specialists, and the teachers who had not been interviewed completed questionnaires that provided data about themselves as, for example, their training and experience and about the school as they perceived it.

Quite often the same data were requested of several respondents which allowed comparisons among responses that came from different sources. Differences between responses showed when there were different perceptions regardless of the actual facts.

There were several other forms that provided statistical data. On one, the teachers of grades one through three provided for each of three days a time log showing the distribution of teacher time devoted to different activities from which ERC could infer the amount of teacher time devoted to reading instruction. A second form elicited from the principal certain data about the school as, for example, racial mix, the

extent of bilingualism, student mobility, and the like, to verify the data that had been used to match the pairs of schools to make them comparable. A third form, a survey of knowledge about classical children's literature adapted from a standard test of this sort prepared by Charlotte Huck, was given to the teachers of grades four through six anonymously. The teacher scores were partial evidence for the training and experience factor.

Two survey instruments were administered to sixth grade students. One randomly selected group completed an inventory composed of forty-five questions which were designed to uncover not only positive attitudes toward reading but also indications whether or not reading was a preferred activity. Another group of sixth graders completed questionnaires about home backgrounds, indicating the presence or absence of applicances in the home, the educational levels achieved by parents, and information about parental aspirations and support of the child as a student. Such data related to verifying that the backgrounds of the students were similar in the two sets of schools. ERC recognized that while the students were not the most accurate source of the data, people still do act according to what they believe to be true even if that differs from what is actually true.

Instrumentation for Data Collection

Data related to each of the factors were clearly collected from many different sources. Table 1 provides a summary of how many items on various interviews, questionnaires, or other instruments were used to collect data on each factor. To illustrate how to interpret the summaries, consider leadership. The table shows that one question in each teacher

TABLE 1

NUMBER OF ITEMS IN EACH INSTRUMENT BY FACTORS

	CENTRAL OFFICE INTERVIEW	PRINCIPAL INTERVIEW	TEACHER INTERVIEW	READING SPECLST. INTERVIEW	LIBRARIAN INTERVIEW	PRINCIPAL QUESTIONNAIRE	READING SPECLST. QUESTNNRE.	TEACHER QUESTIONNAIRE	LITERATURE SURVEY	CLASSROOM OBSERVATION	LEARNER ACTIVITY INDEX	TEACHER TIME LOG
A. LEADERSHIP	-	-	1	-	-	3	2	3	-	-	-	-
B. COORDINATION	5	1	2	2	1	4	3	3	-	-	-	-
C. ADDITIONAL READING PERS.	2	-	-	-	-	4	3	2	-	1	-	-
D. ATMOSPHERE	-	-	-	-	-	-	-	-	-	4	-	-
E. INDIVIDUALIZATION	2	1	1	1	-	-	2	4	-	1	-	-
F. EVALUATION	1	1	-	-	-	1	1	9	-	-	-	-
G. EXPECTATIONS	-	-	-	-	-	11	11	11	-	-	-	-
H. STRONG EMPHASIS	1	3	3	3	3	-	-	5	-	8	-	9
I. USE OF PHONICS	-	1	-	-	-	-	-	2	-	-	-	-
J. TRAINING & EXPERIENCE	1	1	1	1	-	7	7	7	1	-	-	-
K. QUALITY OF TEACHING	-	-	-	-	-	-	-	-	-	12	8	-

interview dealt with leadership data as well as three questions on principal and teacher questionnaires and two on the questionnaire completed by reading specialists.

The instruments varied among each other with some natural differences in that open-ended questions most frequently appeared in interview protocols so that probes could be made to clarify and expand responses. Following is a listing of the questions used in the Central Office Interview protocol as an illustration. The actual form provided space for interviewers to enter responses, but this listing shows the questions and the foreseen probes.

CENTRAL OFFICE INTERVIEW

1. Do you have specific goals and objectives related to the reading program in your schools?

   _____ YES _____ NO

   Are they in writing?

   _____ YES _____ NO (IF YES) May I please have a copy?

2. How are revisions made to the goals and objectives?

3. Does the school system have a curriculum guide in reading?

   _____ YES _____ NO

   (If YES) May I please see or borrow a copy?

   (If YES) When was the curriculum last revised?

   (If YES) Who is responsible for revising and updating the guide?

4. Are the teachers given specific checkpoints (e.g., pages, books) that they should read by given dates?

   _____ YES _____ NO

5. Are there particular features of your reading program that you think others might find beneficial?

   _____ YES _____ NO

   (If YES) Would you please describe them:

6. Do you find any weaknesses in your reading program?

_____ YES _____ NO

(If YES) What?

7. Are there people in the schools besides you and classroom teachers who assist students with reading? Please describe.

_____ YES _____ NO

(If YES) Who determines their function?

8. Is there a reading diagnostic program in the school?

_____ YES _____ NO

(If YES) How does it operate?

Are all children involved?

How many specialists are involved?

Are special prescriptions made?

9. Do most of the teachers in your schools individualize the reading program for each child?

_____ YES _____ NO

(If YES) How?

10. What records are kept on reading performance of the students?

If possible, may I please have samples of the record sheets?

11. Please describe any in-service training programs in reading conducted for teachers.

Who conducts the program?

Are you involved in the program? In what way?

12. Do you draw upon resources outside the schools (consultants, colleges, institutions) to help in the reading program?

_____ YES _____ NO

(If YES) Please describe.

Questionnaires contained mostly questions with straightforward answers which required no amplification by respondents and, as such, were not unlike those used in other survey studies that relied entirely upon questionnaire responses. This study further relied upon observation instruments to provide other data. One such instrument was used to record four times during a reading class the numbers of students participating in different forms of activities according to the following outline of activities:

I. Academic

A. Skill learning

1. Reading

a. Alone_____

b. With others

1. Teacher_____

2. Aide_____

3. Student(s)_____

4. Combination of above_____

2. Non-reading (art, music, directed play)_____

B. Logistics (preparation for lesson)_____

II. Non-Academic (being disciplined, inactive, eating)_____

Finally, some of the instruments not only provided for but encouraged the observers to make comments about what they saw that fit no given structure. These comments provided clinical data--essentially anthropological data--about classrooms observed, the staff, and the school in general. (Note 3)

Analyzing the Data

The data collected about the twenty schools came from many different sources and were in different forms ranging from "hard" data, as, numbers, to "soft" data, as, anecdotes. Numerical and categorical, hard data could easily have been summarized statistically for comparison purposes while descriptive and anecdotal, soft data were more amenable to clinical analysis. The study staff elected not to conduct statistical and clinical analyses separately from each other, but instead to combine the two in order that the analysis would have the strengths of each combined. To combine the two approaches, the staff used collective judgments. These collective judgments were made factor by factor following a review of all the data collected about the schools by a team of at least four people, including staff who had visited the school because they had first hand experiences important to the judgment making.

A summary of all the data collected and arranged according to the study factors was prepared by one of the staff who had visited the school, and then was presented to each member of the group responsible for making the judgments. Following independent reading of the clinical and statistical data, the group assembled and discussed the data about each factor searching for clarification whenever the data were conflicting; frequently those who had visited the school were able to explain or amplify upon the data. Following the discussion each member of the group made a rating that indicated his or her judgment about whether the data about the study factor suggested that the school was either a successful or a contrast school. These ratings were made without knowing yet whether the school was a successful or a contrast school.

A rating scale of five points was designed to express the likelihood that the school was either a successful school or a contrast school. Following is a description of each scale value:

Scale Value	Description
5	On this factor alone this school definitely appears to be a successful school.
4	On this factor alone this school probably is a successful school.
3	On this factor alone this school could be either a successful or a contrast school.
2	On this factor alone this school probably is a contrast school.
1	On this factor alone this school definitely appears to be a contrast school.

When the ratings were all the same, that unanimous rating became the consensus rating. When the ratings differed, however, consensus was not achieved by simply averaging the ratings. Rather, further discussions served to clarify the data and new ballots were made. On a new balloting individual ratings were often changed. The cyclic process of balloting and discussing was repeated until there was unanimity by the group.

The group process not only resulted in unanimity but resulted in finer distinctions than originally planned when the groups introduced pluses and minuses to the five point scale to reflect their finer judgments. Beginning with a five point scale, the addition of pluses and minuses ultimately resulted in a rating scale of thirteen points. (Note 4)

The consensus rating approach was used to obtain ratings based upon the clinical and statistical data collected for each factor except the Use of Phonics factor. That factor was not based upon clinical data to the same extent as other factors, so a different approach was used. ERC

had one of the reading experts inspect the different basals and supplementary materials that each teacher reported using and classify them according to the extent to which they relied upon phonics and decoding. The staff then computed for each school the average of the expert's ratings, which were based upon the following scale:

Rating	Description
5	Mainly phonics.
4	Phonics mixed with other skills.
3	Some phonics.
2	Partially phonics.
1	No phonics.

Analysis of Ratings

In Table 2 are given the ratings for the twenty schools by each study factor. The ratings for the successful schools are grouped together, and the ratings for the contrast schools are also grouped together.

One way to analyze the ratings for the two groups of schools is to compute means for each group and compare them. Table 3 shows the group means computed by assuming that a plus (+) is equivalent to adding one-third of a rating point and that a minus (-) is equivalent to subtracting one-third for the group consensus ratings.

TABLE 2

RATINGS OF FACTORS BY SCHOOLS

	SUCCESSFUL SCHOOLS										CONTRAST SCHOOLS									
	SCHOOL CODE										SCHOOL CODE									
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
A. LEADERSHIP	2+	4	2-	2+	3	2	4+	2	2	1	3+	2+	1+	3	3+	3+	3-	1	2+	3+
B. COORDINATION	2+	4	2	3-	2-	3-	3-	2	2	1+	4	3+	2+	2+	2	2-	3-	1	1+	4+
C. ADDITIONAL READING PERS.	3-	2	3	3+	2+	2	4	3-	2	3-	4	4+	2+	4-	3+	3	3-	1+	3-	3
D. ATMOSPHERE	2+	4	4	4	2+	2+	2	4-	3+	2	3	4	4	3-	3+	3-	2+	3	3+	3+
E. INDIVIDUALIZATION	2	2+	2	3-	3-	2-	3-	2	2-	3	3-	2+	2+	3+	2+	2	2+	2+	2+	1+
F. EVALUATION	2+	3-	2	3	2	2	3	2+	3-	2+	3+	3-	2	3	3+	2	3	2-	2	2+
G. EXPECTATIONS	1+	3+	2+	3-	2	3	3	1	2	2	2-	2	2	2+	2	2	2-	3	2+	3+
H. STRONG EMPHASIS	2+	4	4	3+	2	3	4+	3	3+	3	4	3	3-	4-	3+	2+	3+	3-	3+	4+
I. USE OF PHONICS																				
Basal	2.9	3.0	3.1	3.0	3.0	3.0	3.0	3.0	1.7	2.2	3.1	3.0	3.1	2.9	3.3	2.7	3.1	2.7	2.1	3.0
Supplementary	2.8	3.1	3.2	3.7	3.4	3.6	4.2	3.2	3.6	3.7	2.9	3.0	3.8	3.8	3.7	3.6	3.7	3.7	3.6	3.9
J. TRAINING & EXPERIENCE	2	3-	4-	3+	2	4	4-	3	3	2-	4-	3-	3	3+	3-	1+	3+	2	4	2+
K. QUALITY OF TEACHING	2-	4	3+	4	3+	1+	2	2-	3	3+	4-	2+	3	4-	3+	2	4	3+	4	3+

TABLE 3

Mean Ratings of Groups of Schools

Factors	Successful Schools	Contrast Schools
A. Leadership	2.47	2.60
B. Coordination	2.33	2.50
C. Additional Reading Personnel	2.67	3.03
D. Atmosphere	3.00	3.17
E. Individualization	2.37	2.33
F. Evaluation	2.43	2.53
G. Expectation	2.17	2.13
H. Strong Emphasis	3.23	3.27
I. Use of Phonics		
Basal	2.79	2.90
Supplementary	3.45	3.57
J. Training and Experience	2.90	2.[illegible]
K. Quality of Teaching	2.77	3.27

None of the mean ratings of the successful schools is significantly different statistically from the corresponding mean for the contrast schools, but the total number of schools is so small that such use of statistical tests lacks sufficient power to detect other than really large differences between the two groups.

Even though the means taken factor by factor show no significant differences, as a group there is some suggestion that there is an overall difference, oddly enough in favor of the contrast schools. Among twelve possible comparisons of means that can be made from the data in Table 3 there are nine comparisons in which the contrast means exceed the successful means. This number of differences in favor of one group comes close to

being significant (probability of .07) and suggests that a comparison between the groups based upon all factors taken together rather than singly might produce significance.

To investigate the collective effects of the various factors the study staff used a more complex analysis procedure, the multivariate statistical technique of Discriminant Analysis, which considers the means and interrelationships of the factors taken together in a linear combination. This analysis, however, produced no evidence of differences between the two groups of schools. Since the linear model does not include interactions between factors, the staff made an effort to introduce product terms into the discriminant analysis so that main affects and interaction effects could be assessed together. Twenty schools provided too few degrees of freedom to include all possible product terms, so terms were introduced only if there was a suggestion that there was some interaction from inspection of the scores. The analysis of interactions by discriminant techniques suggested that High Expectations interacted with a combination of Individualization and Evaluation in such a way that the absence of High Expectation essentially discounted any effects of Individualization and Evaluation. This interaction effect was not significant--just as was the case with the comparison of means, the number of schools is quite small--but it suggests the hypothesis that the Individualization factor and the Evaluation factor that should accompany it do not produce positive effects unless the staff agrees that the students are capable of learning. If such a hypothesis is true, it may mean in operational terms that a staff with low expectations fails to diagnose weaknesses and prescribe assignments properly. The study provides no real evidence for this or for the hypothesis, however, because no data were found to be convincingly supportive.

Another form of analysis provides for comparisons between the two groups of schools on all the numerical and categorical data collected on the schools. Such comparisons naturally exclude the clinical data obtained, but nevertheless they are possible sources of meaningful differences between the two groups of schools. There were ninty-four data items for which it was possible to search comparisons between the two groups of schools for statistical significance. With this many items, about five are expected to produce differences apparently significant at the five percent level by chance alone, and one difference is expected by chance to be significant at the one percent level. In fact, there were fewer than five differences at the five percent level and none at the one percent level. These comparisons therefore failed to establish differences between the groups of schools just as all the other analyses had failed.

The study staff reviewed other data about the schools to see if there were possible explanations other than the study factors for the differences between the schools. For example, a review of new data collected on the general compositions of the student bodies did show that one pair of matched schools did not really match well on the proportions of bilingual students because the proportion had increased in the contrast school. In another pair the socioeconomic data had not provided a good match. In that pair the successful school was found to have, in fact, the relatively high number of low income families, but the incomes had a bimodal distribution resulting in a relatively high number of middle class families as well. In some four other pairs there were variations in student mobility that had not been known earlier because some data

were missing. In each instance the contrast school had a more mobile student body, and that fact was a possible explanation of score differences. Students transferring into a school obviously bring with them experiences from other schools which may or may not have helped them to read and so may depress or elevate a school's total performance, and conversely a school that loses its students to other schools has its effects on reading dissipated. As a result of this analysis of other data from the twenty schools, some evidence of contaminating external effects existed, but again there was not sufficient evidence to explain all the differences in test scores.

Thus far, the analyses used, however complex, failed to uncover consistent differences between the ratings of the two groups of schools. It does not follow, however, that all the ratings of the twenty schools are alike. This can be verified by direct inspection of the ratings given in Table 2, perhaps the simplest form of analysis. This inspection shows, for example, that the successful schools have a very mixed set of ratings. Some ratings of the successful schools are high, a fact which indicates that there is evidence of the presence of the study factors in the successful schools, but that evidence is scattered across the schools and across the factors. Ratings of 4, including 4- and 4+ as well as 4, are easily seen to occur in several rows and in several columns, and no consistent pattern is apparent either by schools or by factors. In fact, schools that have high ratings on some factors also have low ratings on others. This means that they appear to be successful schools according to some factors, and at the same time they appear <u>not</u> to be successful schools according to other factors.

Among the successful schools there are some with ratings so low that they appear not even to be successful schools according to all or most all of the factors. Successful school #1, for example, has ratings that uniformly are interpreted as meaning it looks not at all like a successful school. Successful school #10 has ratings that are almost consistenly low but not so extreme. Thus, some successful schools appear as though they should be contrast schools instead.

Not only do some successful schools appear instead to be contrast schools, but some contrast schools appear more like successful schools. Contrast school #11 has ratings that taken together suggest it is a successful school. In fact, contrast school #11 looks more like a successful school than do successful schools #1, #5, and #10..

Thus, in addition to the original two groups of schools the study has identified two other groups. One includes successful schools that look more like contrast schools. These are schools with students achieving national norms on tests, but otherwise these schools employ practices and procedures not judged to be different from those found in mediocre schools. The other new group differs because its procedures and practices were judged to be above the normal experience, but the students have low test scores. Then two new groups seem in some way to be exceptions to some rule or rules.

The four groups are mutually exclusive, but there are underlying connections between the groups that are illustrated in Table 4.

Table 4

Processes	Test Scores: positive	Test Scores: negative
positive	1. True Positive	2. False Positive
negative	3. False Negative	4. True Negative

The rows of Table 4 separate schools according to whether an inspection of their procedures and practices lead to a positive or a negative judgment. The columns classify schools according to test scores, which are some measure of the outcomes of the school operations. In cell #1 are schools with positive judgments of processes and with good test scores. These are labeled True Positive schools. In cell #4 are schools judged low on processes and low on test scores, and they are called True Negative schools. Cell #2 contains schools with positive approaches but with low test scores, so they are labeled False Positive schools. Finally in cell #3 are the schools with weak appearing practices and procedures with nevertheless good test scores, and they are called False Negative schools.

The twenty study schools had already been classified by the columns when they were selected, and the staff later classified them by rows in order to identify which schools fell in which cells. Schools were considered positive on processes if four or more of their ratings were above three and thereby in the range for which the judgment was that the school was a successful school. The negative schools had none, one, or two ratings above three, so their ratings were predominantly low and they thus had been judged to be not successful. Two of the twenty schools had three good ratings, not enough to be judged positive, but

too many to be negative, and those two were excluded. Table 5 shows how many study schools fell in each of the cells.

Table 5

	+	-
+	4	5
-	6	3

The existence of four groups makes possible different analyses of success in schools according to the overall level of their procedures and practices. Thus, interesting comparisons are possible between True Postive and False Positive schools to see if there are suggestions of determinants of success, and similarly False Negative and True Negative can be compared with the same objective.

As before, however, care must be taken that external influences are not operating to explain group differences. Accordingly, the project staff computed the means by cell of the poverty measures, percent of low incomes and percent of free milk or lunch. Tables 6 and 7 report these means:

Table 6

Mean Percent Low Income

	+	-
+	47%	39%
-	39%	36%

Table 7

Mean Percent Free Milk of Lunch

	+	-
+	58%	50%
-	51%	47%

Inspection of Tables 6 and 7 shows that in each row the successful schools had higher poverty indices than did the contrast schools, so test score differences cannot be explained by the usual model in which poverty is associated with low test performance.

Tables 6 and 7 also show that the True Positive schools, those high on test scores and high on ratings, have the economically poorest students among the four types of schools. This difference is of course among a specially selected set of schools from cities and not among a representative set of schools, but the existence of the difference does suggest that some schools can and do reverse the trend for low performance to be associated with high poverty.

Poverty differences do not help to characterize the schools in the four cells; other ways to characterize the schools are needed, and the study factors are a useful source of other ways. As an aid in reviewing factors, the ratings on the factors for these four schools are repeated in Table 8.

Table 8

Ratings of Factors for True Positive Schools

	School Code			
	2	3	4	7
A. Leadership	4	2-	2+	4+
B. Coordination	4	2	3-	3-
C. Additional Reading Personnel	2	3	3+	4
D. Atmosphere	4	4	4	2
E. Individualization	2+	2	3-	4-
F. Evaluation	3-	2	3	3
G. Expectations	3+	2+	3-	3
H. Strong Emphasis	4	4	3+	4+
I. Use of Phonics				
Basal	3.0	3.1	3.0	3.0
Supplementary	3.1	3.2	3.7	4.2
J. Training and Experience	3-	4-	3+	4-
K. Quality of Teaching	4	3+	4	2

Inspection of the ratings in Table 8 shows the four schools were not high on all the ratings, and that suggests that it is not necessary to be high on all factors to be a True Positive school. Thus, not all factors are necessary to achieve success. Further, there is not a large set of factors on which the four schools are uniformly high, and the schools are uniformly high only on Strong Emphasis on Reading.

There are three factors for which three of the four schools were judged to be high: Atmosphere; Training and Experience; and Quality of Teaching. The one school that was not high on Atmosphere is an open space school where there were low readings made of the purposeful and

quiet scales. That same school was rated low on Quality of Teaching. Since that school did not rely solely upon traditional teaching techniques because it uses, for example, student contracts as part of its program, the low rating in Quality of Teaching may not be too critical.

There are two factors for which three schools were judged to be low: Coordination and Individualization. Again it is the open school that is the exception on Individualization; the traditional schools are all low.

Overall, these schools are positive, but there are substantial variations among the ratings assigned them and even among the anecdotal reports from the visiting teams. Reports about the classrooms varied from "rooms colorfully decorated with student work; pleasant, friendly, yet structured and controlled atmosphere; and the presence of special personnel to observe classroom work of students who are candidates for work in the Learning Center" to "generally barren classrooms; detached, preoccupied teachers; and emphasis on recall with no especially probing questions being posed."

The different patterns of strengths and weaknesses among the successful schools suggests that these schools used different approaches to achieving success and that one model for success is not appropriate.

For the opposite extreme, the True Negative schools, the ratings are repeated in Table 9. These are schools judged to have low performing students because the schools do not demonstrate strengths in what they do, and the judgment is borne out by low test scores.

Table 9

Ratings of Factors for True Negative Schools

	School Code		
	13	16	18
A. Leadership	1+	3+	1
B. Coordination	2+	2-	1
C. Additional Reading Personnel	2+	3	1+
D. Atmosphere	4	3-	3
E. Individualization	2+	2	2+
F. Evaluation	2	2	2-
G. Expectations	2	2	3
H. Strong Emphasis	3-	2+	3-
I. Use of Phonics			
Basal	3.1	2.7	2.7
Supplementary	3.8	3.6	3.7
J. Training and Experience	3	1+	2
K. Quality of Teaching	3	2	3+

The ratings of the True Negative schools are indeed low, but they are not always all low on the same factors. Typically one or two schools have ratings below 3, but on four dimensions all three schools have ratings below 3. These are Coordination, Individualization, Evaluation, and Strong Emphasis on Reading, factors in which the three schools share weaknesses.

Among the other factors there are none for which all the True Negative schools show strength, but there is evidence to show that not all of these schools are totally lacking in good practices. The ratings in Atmosphere and the mean ratings in the Use of Phonics in supplementary materials, while rather mediocre, are not as low as are other ratings

for this group.

For the False Positive schools, those where good observations and judgments were made but good test scores were not, Table 10 summarizes factor ratings.

Table 10

Ratings of Factors for False Positive Schools

	School Code				
	11	14	15	19	20
A. Leadership	3+	3	3+	2+	3+
B. Coordination	4	2+	2	1+	4+
C. Additional Reading Personnel	4	4-	3+	3-	3
D. Atmosphere	3	3-	3+	3+	3+
E. Individualization	3-	3+	2+	2+	1+
F. Evaluation	3+	3	3+	2	2+
G. Expectations	2-	2+	2	2+	3+
H. Strong Emphasis	4	4-	3+	3+	4+
I. Use of Phonics					
Basal	3.1	2.9	3.3	2.1	3.0
Supplementary	2.9	3.8	3.7	3.6	3.9
J. Training and Experience	4-	3+	3-	4	2+
K. Quality of Teaching	4-	4-	3+	4	3+

These five False Positive schools all place a Strong Emphasis on Reading and display high Quality of Teaching. Three of the five have high ratings in Atmosphere, Use of Additional Reading Personnel, Training and Experience, and Leadership.

It is difficult to characterize these False Positive schools because there is evidence that they employ positive practices and yet their students do not perform well. It is difficult, also, to determine why there is no evidence that the positive practices do not result in good student performance unless, perhaps, it takes time before the positive practices lead to good results. The hypothesis that time is needed is consistent with Weber's assertion that his successful schools required as many as nine years to achieve success. So, one possible characterization of the False Positive schools is that they are changing the effects upon pupil performance and are thus in transition. It is unfortunate that this study was not longitudinal so that changes could have been part of the study, but from the beginning this was a study of schools at a fixed moment. Until additional data over time are available it can only be conjectured that some of these schools have made some changes in processes but that the effects of these changes must await the time necessary before students can reflect those changes in their performance.

The last cell labeled False Negative contains schools whose procedures do not fare well upon observation but whose students do well on tests in reading. Table 11 contains the factor ratings for these schools.

These schools generally show low ratings, and all six are low on Coordination, Additional Reading Personnel, and Evaluation. Five of the six are low on Leadership, Individualization, and High Expectations, and all but two are rated as low in Atmosphere. The few high ratings that exist among these schools are dispersed over several factors, and so no single factor appears high in all these schools.

The pattern of low ratings among these schools suggests that their students should not perform well on reading tests, but they do. There

Table 11

Ratings of Factors for False Negative Schools

	School Code					
	1	5	6	8	9	10
A. Leadership	2+	3	2	2	2	1
B. Coordination	2+	2-	3-	2	2	1+
C. Additional Reading Personnel	3-	2+	2	3-	2	3-
D. Atmosphere	2+	2+	2+	4-	3+	2
E. Individualization	2	3-	2-	2	2-	3
F. Evaluation	2+	2	2	2+	3-	2+
G. Expectations	1+	2	3	1	2	2
H. Strong Emphasis	2+	2	3	3	3+	3
I. Use of Phonics						
Basal	2.9	3.0	3.0	3.0	1.7	2.2
Supplementary	2.8	3.4	3.6	3.2	3.6	3.7
J. Training and Experience	2	2	4	3	3	2-
K. Quality of Teaching	2-	3+	1+	2-	3	3+

are a number of possible explanations for this strange difference between expectancy and actual results. There may be a lack of reliability and accuracy in the observations and ratings of the schools, but the use of many different specialists and the establishment of consensus among them makes this an unlikely explanation. Another possible explanation is that the factors rated have no bearing at all upon the results of testing and that something else is needed to explain the good results. The study staff reviewed the data collected on all the schools and found that these six schools had common attributes other than the low factor

ratings. Common themes that prevailed in these schools were emphases upon discipline and drill. The observers found several instances in which the staffs looked upon the students as adversaries, demanding of them obedience and quiet. In the classrooms observed there were times when entire classes were occupied with worksheets, or teachers spent extensive time in drill and practice activities.

The fact that these are drill and practice schools suggests more than the obvious conclusion that drill and practice in basic reading skills can result in students achieving well in tests of those same basic reading skills. Put another way, these schools teach the skills the tests measure. But these schools were judged to be weak on a number of factors that involve some good practices in schools, and the test scores do not reflect their weaknesses. This suggests the possibility that the tests themselves are inadequate to measure all the behaviors that make up reading.

Just as Coleman and others were wrong to use measures of quantity to characterize the inputs of schools so is it wrong to use solely quantitative measures of school outputs. It is not new to say that there is more to reading than what reading tests measure, but the presence of the False Negative school in this study does point up the fact that some schools probably fail to help their students in the qualitative aspects of reading. Since, however, there were no other measures of outcomes it is impossible to verify that these schools do have failings despite their good test scores.

Comparisons Among the Four Types of Schools

Beside the foregoing analysis of the characteristics of the schools in each of the four cells another analysis involves comparisons among the four classes of schools to identify how they differ and how they are alike. Table 12 provides a summary of the distribution of the ratings in each cell. When all the schools in a cell or all but one have factor ratings that are high (above 3), that factor is listed in the category "High" for the column that corresponds to the group. Similarly, when all or all but one of the factor ratings in a group are below 3, that factor is listed in the "Low" category. When the ratings are divided between high and low, the factors are listed under "Medium." In several instances the factor ratings are not evenly divided but show a tendency (for example, four out of six or three out of five) to be high or low, and the factors are listed as "Tending High" or "Tending Low."

Since the Phonics ratings are based upon a different scale, the summary of those ratings was achieved differently. A school rating for Phonics was considered to be high if it was more than 0.1 above the median for all the schools and low if it was more than 0.1 below the median. In each category of schools the Phonics ratings are quite mixed and the summary places all Phonics-Supplemental ratings by category as medium and all but one category as medium for Phonics-Basal. Thus, the Phonics ratings do not serve to differentiate among the four categories of schools, and, in essence, the four categories are quite alike in the use of phonics.

The four categories of schools are quite alike also in ratings of Coordination and of Individualization. In every instance the Individualization ratings are low, and in all but one case the Coordination ratings were

Table 12: Grouping of Factors by Level of Ratings
by Category of Schools

Level of Ratings	Positive		Negative	
	True	False	False	True
	Type 1	Type 2	Type 3	Type 4
HIGH	Emphasis on Reading Quality of Teaching Atmosphere Training & Experience	Emphasis on Reading Quality of Teaching		
TENDING HIGH		Atmosphere Training & Experience Leadership Add'l. Rdg. Personnel		
MEDIUM	Leadership Add'l. Rdg. Personnel Evaluation Expectation Phonics-Supplemental Phonics-Basal	Evaluation Phonics-Supplemental Phonics-Basal	Training & Experience Quality of Teaching Emphasis on Reading Phonics-Supplemental Phonics-Basal	Atmosphere Quality of Teaching Phonics-Supplemental
TENDING LOW		Coordination	Atmosphere	Phonics-Basal
LOW	Coordination Individualization	Individualization Expectation	Coordination Individualization Expectation Leadership Add'l. Rdg. Personnel Evaluation	Coordination Individualization Expectation Leadership Add'l. Rdg. Personnel Evaluation Training & Reading Emphasis on Reading

low. It is interesting to note, therefore, that the four classes of schools and all the schools individually were judged to be low on Coordination and Individualization. Further, these two factors, as is the case with the Phonics factors, do not differentiate the classes of schools and can be ignored for purposes of contrasting the schools to see how they differ.

The positive schools, true and false, naturally have higher ratings than do the negative schools by virtue of the process by which the positives and negatives were identified, but the positive schools differ more on some factors than on others. The ratings of the positive schools are uniformly high on Emphasis on Reading and show the greatest difference over negative schools, a difference of more than one on the five point scale of ratings. The positive schools are high on the Quality of Teaching factor and show almost a whole point difference in ratings over the negative schools. On Leadership the positive ratings were not all high, but because the rating of the negative schools were so low the differences were more than one point on the five point scale.

The true positive schools have high ratings on Atmosphere and the Training and Experience of staff while the false positive schools show only a tendency to be high on these factors. On the other hand, the false positive schools rate somewhat higher on Leadership and the use of Additional Reading Personnel. Thus, the false positive schools show some evidence--by Leadership, Strong Emphasis on Reading, Quality of Teaching, and Additional Reading Personnel--of good practices but that Atmosphere and Training and Experience have yet to be high. These same schools also have low ratings of Expectation, lower than the ratings for true positive schools and, in fact, more like the ratings of the negative

schools. Given a school functioning in ways that should help students to learn it is hard to understand why the Expectations are so low.

There could be several explanations for the low Expectations of the false positive schools. First, there could be a circularity about Expectations such that a staff aware of low pupil performance consciously or unconsciously adjusts expectations to that low level. This circularity could explain why the true positive schools have higher Expectations--their students do better on tests. However, this explanation suggests that the false negative schools whose students do well on tests should have other than the low Expectations that they have, and yet another explanation is needed. Perhaps the staffs of the false negative schools believe their students are not good readers even if the test scores are good, and that possibility is consistent with the earlier assertion that there are qualitative aspects of reading that are not measured by standardized tests.

A second possible explanation for the low Expectations among false positive schools is not unrelated to the first nor is it exclusive of it. If, as suggested earlier, it takes time for efforts made by a school to improve reading to result in good pupil performance in reading tests, then the circularity principle suggests that the Expectations are also in transition and will rise over time. Still it is difficult to understand why these schools do not yet show evidence of setting high standards for their students to achieve.

Table 12 shows that the negative classes of schools have no ratings above mediocrity and have a large number of low ratings. Both classes of negative schools share low ratings on Expectation, Leadership, use of Additional Reading Personnel, and Evaluation. In addition, the true

negative schools are lowest of all on Emphasis on Reading and Staff Training and Experience, but they do have slightly higher ratings on Atmosphere. With only these exceptions, the two classes of negative schools look quite alike--even more alike than do the two positive classes of schools. The true negative schools do indeed look negative, but it is hard to find an explanation for the false negative schools who also look negative but whose students do well on tests of reading. The factor differences are in Training and Experience and in Emphasis on Reading, but it appears that the increased Emphasis on Reading took the form of drill and practice in those skills measured by standardized tests and in little else. Once again there is the suggestion that the false negative schools, while helping students to do well on standardized tests, do little to help students achieve qualitative outcomes on reading.

Impact of the Four Types of Schools

The analyses of the ratings of the schools in the four cells first to characterize each cell and then to see how the four types of schools are alike or different has helped to show how complex is the characterization of schools. It is not enough to characterize schools as merely good or bad nor is it enough to simply characterize schools as good or bad on a set of dimensions. Schools may be good or bad on test scores, but test scores do not adequately represent all the outcomes of a school, and other outcome measures are critical. Such measures, while not available to this study--or even to many other studies for that matter--involve the qualitative aspects of reading. Some schools may be good or bad on quantitative measures--standardized tests--quite independent of how they do on qualitative measures of outcomes. Similarly schools differ on their inputs to the learning process and may differ on qualitative and

quantitative measures of inputs. Thus is the characterization of schools quite complex, perhaps even more complex than the four cells described here.

The four cells, however, serve to describe two dichotomies, one for how well a school is following good practices and one for its outcomes. One cell represents good outcomes and good practices; another represents bad outcomes with the absence of good practices. The other cells represent schools that somehow do not match outcome to expectation: the false positive school falls below expectation; the false negative school exceeds its expectation.

The four cells become useful not only for characterizing the different relationships between outcome and expectation, but they help to provide two important inferences concerning the schools whose outcomes do not match expectation. The false positive schools whose outcomes fall short of expectation appear to be schools in transition. Of course, their transition can logically be in either direction: they may be on their way to achieving successful outcomes; or they may be dropping off in some or all of their good practices. Further study of such schools over time is necessary to establish the direction of change, and such a study will also help to pinpoint how different strategies for change operate. Since true positive schools have different patterns of excellence in the study factors it is likely that their success was achieved by different strategies. A study of schools in transition can help to identify how different strategies for different schools can result in movement toward or away from excellence.

The other set of unusual schools, the false negative schools, have good test scores, better even than would be expected from inspection of

their processes. Their narrow dedication to drill and practice does seem to result in good performance on reading tests, but the natural question is whether these schools would fare as well on outcome measures that go beyond elementary skill levels to assess reading behavior, learning by reading, and the like. This suggests further that were those outcome measures available for those schools they would no longer appear to be good on all outcomes. But to say they may not be entirely good requires a judgment about what outcomes or objectives should be met by a school. Some schools may in fact aspire to teach only the basic skills as measured by standardized tests and not try to impart the more qualitative aspects of reading. For those schools good test scores may be enough to them. Other schools may aspire to more than what standardized tests measure, but such schools must be able to describe what they hope to achieve in terms sufficiently behavioral that measurements can be made by having students display those behaviors. For those schools new tests and measures are needed that differ from existing standardized tests, and that, of course, is what criterion-referenced tests are for. Preferred outcomes can differ from school to school and so can the measurement of those outcomes.

Though schools may differ in their aspirations it seems reasonable to assume that the true positive schools represent the richest form of goodness of the schools in this study, but even they are not high on all factors which fact shows that they could be improved. For example, they have low ratings on Coordination and Individualization which show room for improvement. There are four factors (Leadership, Expectations, Additional Reading Personnel, and Evaluation) on which the true positive schools are given mixed or, on the average, medium ratings. Still, these four factors give an interesting suggestion about the dynamics of some school

factors because while they appear inconclusive--as if they are not essential to a school's achieving success--the ratings of the true positive schools are still higher than those of the negative schools. These same four factors account for all of the low factors of the false negative schools except for the Coordination and Individualization factors which do not differentiate among the four classes of schools. And all but two of the low factors among the true negative schools are similarly accounted for. Even though the ratings of the four factors for the true positive schools are medium and inconclusive they are still much better than the low ratings for the negative schools, and differences between magnitudes become important quite apart from the magnitudes themselves.

Looking at the magnitudes of differences has been an approach used in other studies, such as the study by the New York State Department of Education (1974) in which a successful school and an unsuccessful school were compared. That study concluded that leadership, atmosphere, and emphasis on reading were critical factors, factors that to varying degrees are supported by this study. Other factors reported in the Weber study (additional reading personnel, evaluation, and expectation) also receive some support by this study. This study does suggest, moreover, the importance of quality of teaching (which was not deemed essential in other studies) and of staff training and experience, both factors related to the teacher in the classroom.

It may not be important, however, to note areas of agreement and of disagreement nor to try to draw inferences from those areas because this study has shown that successful schools differ among themselves in their patterns of factors. Thus, it may not be appropriate to search for unique factors or patterns of factors that separate good and bad schools but instead to concentrate on the fact that differences between good and bad schools do exist and are discernable. Just as it is recognized that

schools differ and that schools differ in how they achieve success, the critical concern should be with the process by which a school finds out what kind of a school it is and then develops strategies for improvement in the teaching of reading.

Implications for Further Work

A school could determine what kind of a school it is by using directly the approach used in this study. It could enlist outside observers who would visit the school, use the instruments developed for this study, develop the ratings of factors by consensus, study outcome measure, and finally identify which of the four cells the school falls into. There are a number of reasons why this approach is not appropriate, however. First, it is not necessarily best to enlist outside observers in the process because then the school and its staff become passive actors to be observed precisely when it is best to involve them in the process of finding, ultimately, how to achieve improvement. Thus, the staff can be directly involved in collecting some of the data, particularly those which are statistical. The clinical data require often a type of objectivity that may be difficult to expect from self-inspection, but those data can be collected by peers from other schools by means of some collaborative efforts. Schools, particularly those in a given district, now collaborate in workshops and meetings in which ideas are shared. Even across district lines schools now collaborate, for example, to judge each other for accreditation by the New England Association of Schools and Colleges, but that collaboration is quite formal and potentially punitive. A form of collaboration somewhere between the informal district meeting of teachers and the accreditation sessions is needed. Such a collaboration would involve several schools helping to observe each

other and would involve the collective staffs in a form of introspection and concern for what they are doing that in and of itself is good and healthy.

Second, it is not appropriate to use directly the instruments from this study because the instruments should be made more sensitive to the four types of schools. Further, the instruments should be expanded to deal, as we shall see later, with the collection of additional data about strategies and outcomes and in addition, the staffs should be trained in how to use the instruments.

Just as with the use of peers to make observations and collect certain data, it is desirable to use peers to develop the ratings of factors by consensus rather than relying solely upon outsiders. Still, it will be important to have the rating teams trained and practiced in the consensus process.

To study outcome measures as they were in this study--and in most other studies as well--by relying upon standardized tests may be inadequate. The schools themselves must determine what outcome measures are important by first deciding what objectives they have for reading performance and expressing these in behavioral terms that allow for measurement. For some of such objectives standardized reading tests may be appropriate, and for others some existing instruments that are not standardized may be appropriate. Examples of such instruments are those used in the national assessment program of the Education Commission of the States on the instruments already in assessment use in the Commonwealth by the Massachusetts Department of Education. These existing instruments would be very useful for some outcome measures and not for others, and other techniques of measurement may be necessary. Some such techniques may

include unobtrusive measurements of reading behavior, for example the use of books in the library. Care must be taken that when appropriate such techniques reveal qualitative behaviors and not just quantitative. Thus, it is important to go beyond just counting how many books are checked out of a library. But, this introduces additional decisions by the schools because it is not sufficient to make judgments about difficulty levels of books taken out or the topics covered in the books taken out, because individual interests vary so much that given books have different effects upon different people. It is better that the schools search for evidence that student reading shows evidence of the transfer of reading skills to other forms of learning, to language development, or even to simple enjoyment.

When a group of schools has finally determined how each of the schools stands in terms of its practices and in terms of its outcome measures, it is then important to discover what the effects are of different strategies for change. To accomplish this task the schools then become laboratories and the school staffs become the researchers who will follow the schools over time, particularly those schools identified as in transition. The continuation over time of the observing of the schools will allow the researchers--the school staffs acting collaboratively--to investigate what emphases and what changes the schools are making. Instruments can help to find out what strategies the schools believe they are following. Observations over time will substantiate whether or not they are doing as they say they are doing or identify any changes of which there is no direct awareness.

When the schools succeed in using themselves as laboratories to find over time what kinds of schools they are and what kinds of schools they are becoming by virtue of different strategies then the schools become better able to define for themselves strategies appropriate for them to achieve the particular success they seek. This form of action research seems most appropriate for schools to use to achieve success for several reasons. First, schools may appropriately define for themselves different definitions of success; no one definition should be imposed externally. Second, schools may elect different strategies to achieve success; no strategies should be imposed externally. Finally, the possibility that schools become their own agents for change and improvement is not only consistent with the policy of local control of schools but it places the control in the schools where the capacity to determine the most meaningful strategies for success exist.

NOTES

Note 1 — In the tenth city there was soon to be a new superintendent, and the incumbent preferred not to make a commitment to participation in the study for his successor.

Note 2 — The contrast schools were not the schools with the lowest test scores since such schools did not always match on poverty or other measures.

Note 3 — All the instruments used in this study are not reproduced in this report because the findings of this study result in the recommendation to be noted later that the instruments should be refined to account for additional data, and the instruments should therefore not be used directly. Copies of the instruments are, however, available upon request made to Educational Research Corporation, 85 Main Street, Watertown, Massachusetts 02172.

Note 4 — It was not necessary to use 5+ or 1- because ratings of 5 and 1 already expressed a conviction of certainty.

REFERENCES

Coleman, J.S., Campbell, E.Q., Hobson, C.J., McPartland, J., Mood, A.M., Weinfeld, F.D., & York, R.L. Equality of educational opportunity. Washington, D.C.: U.S. Department of Health, Education, and Welfare, U.S. Government Printing Office, 1966.

Guthrie, J. A survey of school effectiveness studies. In U.S. Department of Health, Education, and Welfare (Ed.). Do teachers make a difference? Washington, D.C.: U.S. Government Printing Office, 1970.

State of New York, Office of Education Performance Review. School factors influencing reading achievement: A case study of two inner city schools. Albany, New York, 1974.

Weber, G. Inner-city children can be taught to read. Washington, D.C.: Council for Basic Education, 1971.