DOCUMENT RESUME

ED 188 106 CS 005 388

TITLE Report on the Special Studies of Selected ECE Schools with Increasing and Decreasing Reading Scores: A Report to the Joint Legislative Budget Committee and the Fiscal and Policy Committees of the Legislature as Required by Item 283 of the Supplemental Language to the 1977-78 Budget Act.

INSTITUTION California State Dept. of Education, Sacramento. Office of Program Evaluation and Research.

PUB DATE 80

NOTE 78p.

AVAILABLE FROM Publications Sales, California State Department of Education, P.O. Box 271, Sacramento, CA 95802 ($1.50)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.

DESCRIPTORS Case Studies; Classroom Environment; Early Childhood Education; Elementary Education; Primary Education; *Reading Achievement; *Reading Failure; *Reading Improvement; *Reading Research

ABSTRACT

This report describes the efforts of the California State Department of Education in response to a 1975-76 evaluation report which indicated that in some schools, third grade reading scores were declining--specifically in schools whose entering students averaged below the twentieth percentile on the 1973-74 Entry Level Test (ELT) after three years in early childhood education. Two research questions are discussed: "What are the characteristics of the schools with students with ELT scores below the twentieth percentile?" and "What circumstances are associated with the decline or improvement in third grade student reading achievement?" The first chapter presents the background information that inspired the studies reported; the second chapter reports on the procedures, sample, and findings from a special case study conducted by the Department of Education; and the third chapter presents the procedures, sample, and findings from a study of classroom instructional processes conducted by Stanford Research International. (HCD)

Report
on the Special Studies of Selected ECE Schools with Increasing and Decreasing Reading Scores

A Report to the Joint Legislative Budget Committee
and the Fiscal and Policy Committees of the Legislature
as Required by Item 283 of the
Supplemental Language to the 1977-78 Budget Act

California State Department of Education
Wilson Riles-Superintendent of Public Instruction
Sacramento, 1980

This legislative report was prepared by the Office of Program Evaluation and Research, California State Department of Education, and was published by the Department, 721 Capitol Mall, Sacramento, CA 95814. Any questions regarding the report should be addressed to the Office of Program Evaluation and Research (phone: 916-445-0297).

Distributed under the provisions
of the Library Distribution Act

1980

Contents

TABLES

ILLUSTRATIONS

I. Background

This report describes efforts of the State Department of Education in response to a finding in the Evaluation Report of ECE, ESEA Title I, and EDY, 1975-76, which indicated that in some schools, third grade reading scores, as measured by the California Assessment Program (CAP), were declining. Specifically, the findings indicated that "in schools whose entering students averaged between the 21st and the 99th percentile on the 1973-74 Entry Level Test (ELT), grade three reading achievement improved markedly beyond predicted levels after three years in ECE." In contrast to these increasing trends, "in schools whose entering students averaged below the 20th percentile on the 1973-74 Entry Level Test, grade three reading achievement declined relative to prediction after three years in ECE."[1]

TABLE V-9

Changes in Residual Scores (Weighted Averages) on Grade Three Reading Achievement Tests After One, Two, and Three Years of Participation in ECE, by Three Levels of Performance on the 1973-74 *Entry Level Test*

Number of years in ECE	Changes in residual scores, grouped by percentile rank on 1973-74 *ELT*			All ECE schools (weighted average)
	1–20	21–60	61–99	
One year	.00	.05	+.06	.00
Two years	−.15	+.05	+.06	−.01
Three years	−.24	+.19	+.18	+.03

NOTES: Residuals were standardized to have a mean of zero and a standard deviation of one. Changes in residuals after one year of ECE were obtained by first calculating the average residual for ECE schools in the year before they entered ECE, calculating their average residual in their first year in ECE, and then subtracting the former from the latter. Similarly, changes in residuals after two and three years in ECE were obtained by subtracting the preprogram residuals of ECE schools from their residuals after two and three years, respectively, in ECE.

The numbers of students and schools included in tables V-9 through V-12 varied because of the different numbers of schools involved in ECE in each year. In all cases the number of students for whom data were analyzed was greater than 6,200.

TABLE V-10

Changes in Residual Scores (Unweighted Averages) on Grade Three Reading Achievement Tests After One, Two, and Three Years of Participation in ECE, by Three Levels of Performance on the 1973-74 *Entry Level Test*

Number of years in ECE	Changes in residual scores, grouped by percentile rank on 1973-74 *ELT*			All ECE schools (unweighted average)
	1–20	21–60	61–99	
One year	.04	.07	+.07	.02
Two years	.15	+.02	+.14	−.02
Three years	.14	+.14	+.19	+.02

NOTES: Residuals were standardized to have a mean of zero and a standard deviation of one. Changes in residuals after one year of ECE were obtained by first calculating the average residual for ECE schools in the year before they entered ECE, calculating their average residual in their first year in ECE, and then subtracting the former from the latter. Similarly, changes in residuals after two and three years in ECE were obtained by subtracting the preprogram residuals of ECE schools from their residuals after two and three years, respectively, in ECE.

The numbers of students and schools included in tables V-9 through V-12 varied because of the different numbers of schools involved in ECE in each year. In all cases the number of students for whom data were analyzed was greater than 6,200.

[1] Evaluation Report of ECE, ESEA Title I, and EDY, 1975-76. Sacramento: California State Department of Education, 1977, p. 36.

The findings, which were displayed in tables V-9 and V-10 in the evaluation report for 1975-76 and are reprinted here, were interpreted with the following statements: "The improvement in residual scores[2] indicates that, on the average, ECE seems to be associated with improved residual scores in reading in a majority of schools--a group of schools whose students averaged from the 21st to the 99th percentiles on the Entry Level Test. On the other hand, ECE seems to be associated with declining residuals in a much smaller group of schools whose students had lower average levels of learning readiness when they entered grade one."[3]

Although the finding about the decline in reading scores occurred in the schools below the 20th percentile on the 1973-74 Entry Level Test, not every school within the group had had third grade reading scores each year increasingly lower than the score predicted for that school. Some schools among the population had seen third grade achievement scores increase relative to their predicted score over the three-year period in question, while others were unchanged. Of the schools in the lowest 20 percent on the 1973-74 Entry Level Test, 110 schools had increases in residual scores, 146 schools had decreases, and 121 schools had scores that were unchanged.

While the overriding purpose of the Department's inquiry and the focus of the research which would take place were to find out how the average scores came to decline for the schools below the 20th percentile on the Entry Level Test, information about schools with increasing scores in the same group was important to explore. Two types of research questions were raised.

The first question was "What are the characteristics of the schools with students with ELT scores below the 20th percentile?" To respond to this question, statistical data, available through the California Assessment Program, were compiled and statistical profiles were developed. This information was examined by Department staff and provided additional background for the study described in this report.

Perhaps the most important question to be dealt with in this study was "What circumstances are associated with the decline or improvement in third grade student reading achievement?" Two complementary research strategies were selected to answer the question : (1) a traditional correlational study of within classroom variables conducted by Jane Stallings of SRI International; and (2) a series of case studies of 16 schools--eight schools with increasing CAP third grade scores and eight schools with decreasing scores.

The purpose of the SRI study was to identify classroom processes related to changing test scores among the schools in the lowest 20 percent on the Entry Level Test. Trained observers spent time in classrooms recording how materials were being used, the activities which occurred, the organization of groups, the numbers of teachers and aides involved in the instructional process, the instructional strategies, and the behavior control patterns observed. The 48 classrooms that the observers visited were in 14 schools that were in the lowest 20 percent on the 1973-74 ELT. Eight schools visited by SRI observers were also included

---

[2]"A residual score" is the difference between the predicted score and the actual score.

[3]Ibid.

in the Department of Education case study. From the observational data, the study yielded classroom profiles that described instructional processes that are related to classrooms with increasing student reading achievement and those processes related to classrooms with decreasing student reading achievement.

While the research strategy employed in Dr. Stallings' study captured relationships between processes within classrooms and student achievement, it did not look at processes at the school level.[4] Therefore, a second research strategy was used. The case studies facet of the research strategy was designed to complement the strengths and weaknesses of the SRI classroom observations. The greatest strength of the case study approach was that it facilitated the emergence of complex, interdependent, and unique explanations of score changes.

Any methodology chosen to investigate the changing scores has limitations. One weakness of case studies is that there is less control over the emergence of an incorrect explanation. A second weakness is that no statistical estimates can be made of how widespread any subsequent findings are in schools not studied. Judgments about the frequency of these findings in schools across California must be based solely on the intrinsic nature of the explanations themselves. Each school is a study in itself. Although the case study methodology is less sensitive to commonalities across schools, the advantage gained is in the sensitivity to uniqueness and to the complexity and interaction of various school processes.

The purpose of this research was to enhance the Department's understanding of the score changes so as to improve educational services for children. All schools in the study shared not only the "lower 20" characteristic; all were also early childhood education (ECE) schools. Thus, the studies sought to explain the "how" and the "why" of the 1975-76 Evaluation Report finding in a context which could give direction for improvements and refinements in state leadership and assistance to these schools.

This report consists of three chapters. Chapter I presents the background information which inspired the studies reported. In Chapter II the procedures, sample, and findings from a special case study conducted by the Department of Education are reported. Chapter III presents the procedures, sample, and findings from a study of classroom instructional processes conducted by SRI International.

---

[4]Jane Stallings and others, Early Childhood Education Classroom Evaluation. Prepared for the Office of Program Evaluation and Research, California State Department of Education. Menlo Park, Calif.: SRI International, 1977.

II. Case Studies

The case study reported in this section was one of the research strategies selected for investigation of the decline in third grade reading scores in schools whose students scored in the lowest 20 percent on the Entry Level Test (ELT). A case study approach has the advantage of capturing the unique complex interactions among the varieties of circumstances that are found in schools.

PROCEDURES USED IN THE STUDY

This case study was a combined effort of the Elementary Education Program Management Unit and the Office of Program Evaluation and Research of the State Department of Education. The special study team consisted of seven program consultants and four evaluation consultants. Two program consultants were selected for their bilingual skills and shared the responsibility for the eight schools in the sample which had high proportions of students who spoke a language other than English.

The program-evaluation mix, with changing pairs of team members, was chosen to provide maximum coverage to the schools and to encourage interaction between professional colleagues with different specialties. However, the roles of the two team members were not different within the schools.

Training of the Study Team

The training of the study team members took place over a two-month period immediately preceding the actual study visit period (which was late March through early May, 1977). Among the major activities of the training period were a review of literature, special in-service training in a variety of areas, exercises to make team members confront their individual biases and beliefs, development of the study procedures, and selection of the schools for participation in the study.

Literature related to case study methodology was made available to the team members to serve as a background and catalyst during the training period. The literature dealt with the use of case study methodology in education, specific observation techniques and concerns, and descriptions by other researchers who have used the case study method.

During the training, special attention, in the form of in-service workshops, was devoted to certain areas in which the team members felt they needed augmentation of their knowledge or experience. Among the topics specially addressed were issues relating to limited- and non-English-speaking (LES/NES) pupils; black, minority, and poverty concerns; and interpretation and use of California Assessment Program (CAP) data.

Since team members recognized that each person held beliefs, attitudes, and biases that could influence his or her perceptions, a series of exercises was incorporated in the training. During the exercises each person made explicit his or her notions about changing test scores, philosophy of a good educational program, and ideas of what early childhood education (ECE) was. Team members arrived at some collective understandings and were able to see how their individual beliefs might influence perceptions during the observation periods.

After discussing several procedures that could be used in the school visits, the team agreed to combine two approaches: as blotters, team members would absorb information without screening or interpreting it; while in the investigative mode, they would follow leads and gather evidence to support or reject possible explanations. A framework (reminder sheet) would be used for reference only. The team members decided not to use a formal instrument during their school visits for two reasons: (1) to avoid limiting their perceptions in looking for causes and factors; and (2) to encourage candid responses at the schools without creating anxieties about being evaluated.

Selecting Schools for the Sample

Schools were selected for the sample from the population of schools in the lowest 20 percent on the ELT in 1973-74, including schools whose students had improved reading scores. Definitions for schools with "increasing" and "decreasing" scores were developed, based on criteria chosen arbitrarily.[1] Some criteria were used because of their apparent resemblance to the aggregated score changes; other criteria served to expand or narrow the pool of potential study sites.

The criteria for schools with decreasing scores were as follows:

1. Schools must have been in the lowest 20 percent on the Entry Level Test (ELT) in 1973-74.

2. A school's third grade reading scores must have been lower in 1975-76 than prior to the school's entrance into ECE.

   (a) For Phase I schools (those that entered ECE in 1973-74), their 1972-73 percentile ranks had to be greater than their 1975-76 percentile ranks.

   (b) For Phase II schools (those that entered ECE in 1974-75), their 1972-74 raw scores had to be greater than their 1975-76 raw scores.

3. A school's third grade scores must have decreased at least four raw points from 1973-74 to 1975-76.

4. A school's obtained third grade reading score must have been below the school's prediction band in 1975-76.

The criteria for schools with increasing scores were as follows:

---

[1]The selection of schools to be included in the study was subject to a number of constraints. The selection was not a sampling of schools for whom the original finding--that of increasingly larger negative discrepancies between schools' predicted third grade reading scores and their obtained scores--was true. The original finding was based on scores and residuals (differences between predicted and actual scores) aggregated across schools in the lowest 20 percent of the 1973 Entry Level Test (ELT). Because the decline was found to be steeper in ECE schools than in non-ECE schools, inclusion of both types of schools in the sample might have been desirable. However, both the selection factors that determined the entry of schools into ECE and the problem of gaining entrance for a Department of Education study into nonfunded schools led to the decision to include only ECE schools in the study.

1. Schools must have been in the lowest 20 percent on the ELT in 1973-74.

2. A school's third grade reading score must have been higher in 1975-76 than prior to the school's entrance into ECE.

   (a) For Phase I schools, their 1972-73 percentile ranks had to be lower than their 1975-76 percentile ranks.

   (b) For Phase II schools, their 1973-74 raw scores had to be lower than their 1975-76 raw scores.

3. A school's third grade score must have increased at least four raw score points from 1973-74 to 1975-76.

4. A school's obtained reading score must have been within or above the school's prediction band in 1975-76.

5. A minimum of 20 third grade students must have been tested in 1975-76.

Thirty schools that met the decreasing criteria became the list of decreasers; the 24 schools which met the increasing criteria became the increasers. In order to ensure that the final samples of schools with increasing and decreasing scores were matched for percentage of students speaking other than English, each list was divided into "high bilingual" and "low bilingual" schools (bilingual meaning "other language"). High bilingual schools were those whose California Assessment Program (CAP) bilingual ranking was at the 76th percentile or above; the actual percentage of students speaking a language other than English was 24 percent or more. Low bilingual schools ranked at or below the 75th percentile; less than 24 percent of their students spoke a language other than English.

When the lists were dichotomized by bilingual ranking, the list of schools with decreasing scores contained 20 high bilingual and 10 low bilingual schools; the list of schools with increasing scores contained 13 high bilingual and 11 low bilingual schools. Using a random numbers table, schools were drawn from each of the four categories (decreasing-high bilingual, decreasing-low bilingual, increasing-high bilingual, increasing-low bilingual) until four schools plus one alternate had been selected from each category. (When one school declined to participate, the alternate for that category was used.)

The final sample of 16 schools all met the originally designated definitions; the percent of bilingual students was balanced. Of the 16 schools, ten were located in the southern part of the state, four in northern California, and two in the Central Valley. Four schools were in the same school district, but no other district had more than one school represented. Thirteen of the 16 were kindergarten through grade six schools; two were kindergarten through grade three schools; and one school was a kindergarten through grade eight school. The schools ranged in size from 300 to 1,200 students, with the median school size slightly under 600. Two bilingual schools were located in northern California; both Central Valley schools were high bilingual; and the remaining four high bilingual schools were in southern California. Six schools with decreasing scores and three schools with increasing scores were Phase I schools (those that entered ECE in 1973-74); the remaining seven were Phase II schools (those that entered ECE in 1974-75). All 16 schools were multifunded.

The demographic data gathered by the team members on their visits indicated that most of the schools had heavy concentrations of minority and lower socio-economic level students and that many of their families received Aid to Families with Dependent Children (AFDC). Several schools were predominantly black; the percentage was as high as 96 percent and 98 percent in two of the schools. The Hispanic[2] population in several schools was as high as 89 percent, with even higher concentrations during the times when the migrant farm worker population was present. One school had a Filipino population of 26 percent; smaller concentrations of such other ethnic groups as Asian, American Indian, and Samoan were noted in several other schools.

Changing school populations was a characteristic noted in several of the schools visited, the most frequent change being the increase in the number of Hispanic pupils. Influxes of pupils from Mexico occurred in a number of schools. A shift from a majority of white pupils to a majority of black pupils had taken place in a few schools, and in one school the population shift was from a preponderance of black pupils to white. High transiency rates were found in five schools.

The observers noted varying conditions at the schools visited, ranging from a pleasant atmosphere and well-cared-for plant and grounds in a few of the schools to others which were poorly kept, with barred windows, poorly equipped playgrounds, and an oppressive atmosphere. Four schools were in inner-city neighborhoods, three in rural communities, and the others in various types of urban or suburban areas.

Visiting the Schools

Final planning for the school visits was conducted jointly with representatives of each school district in which a selected school was located. A meeting with district liaison persons was held during the training period. The group decided that the procedure for the team's entering the schools would be tailored individually for each school. It was stressed to district personnel that the teams wished to maintain a low profile during their school visits; that the selected schools would not be cast in a negative light; that the data collected would be confidential, with school names coded; and that the research role of the teams would be clearly different from that of the monitor and review (MAR) visits.

Each school was visited by a two- or three-member team for a total of four days. The assignment of observers to schools was made so that no pair of team members would visit more than one school. The four-day visit was divided into two visits of two days each, the visits separated by a week or weekend. (Easter vacation intervened between the two visits in the case of eight schools.) One bilingual team member was assigned to each of the eight high bilingual schools.

---

[2] In most communities in California, the overwhelming majority of persons in this category are of Mexican descent.

The observers' approach to the two visits was differentiated: During the first two days at each school, team members spent their time together utilizing the "blotter" approach; and during the second two days they worked together or separately, following a more investigative procedure. The team entered the school in the manner arranged with the district liaison and/or the school principal. A packet of information had been requested for each school, but no schedules or particular activities were prearranged. By talking informally with teachers, parents, children, principal, and whoever else was present at the school, the pair attempted to reconstruct the history of the changes in the test scores at the school. Team members observed the present situation and learned from it, but they continually asked questions about and made reference to the past, because it was test scores from previous years that were being studied. The goal was to absorb as much information as possible, but to provide no feedback or evaluative judgments. At the conclusion of the second day, the pair of observers organized their information and impressions from the first two-day visit and developed tentative plans for the second two-day visit. At this time, the pair opened the packet containing the school's test data.

When the pair of observers returned to the school for the second two-day visit, their activities were usually more focused than during the first two days. They observed or interviewed separately or together, depending upon the size of the school, time constraints, and the particular avenues of investigation. The afternoon of the fourth day was spent writing the school case study. (All team members kept written notes of their visits in order to provide documentation for the case studies.)

Several aspects of the school visit procedure were designed specifically to provide a measure of reliability (to ensure that another team, following the same procedure, would come to the same conclusions): changing partners for each school; separating the visit into two two-day periods; emphasizing the "blotter" approach during the first visit; being free of an instrument or form; looking at test data only after the second day; including in the school case studies any disagreements between members; and documenting in writing all findings. The training process also contributed to the attempt to provide reliability through the emphasis on bias clarification.

Identifying the Limitations of the Study

A major limitation of the case study methodology is in the sphere of generalization: Its findings cannot be generalized to the total school population in California. However, the findings may have implications across the state in terms of the services provided to children by the school, the district, and the State Department of Education. The implications are broad enough to be meaningful to more schools than the small sample. While the findings could be contrary to the experiences of any one school or group of schools, the same could be said of a study based on large-scale, statistically sophisticated data analysis.

In comparison to a statistical study in which the significance of findings is dependent upon the frequency of occurrence, the significance of the findings and implications in this study is not related to the frequency with which an event occurred. In other words, finding a leadership problem several times did not make it of greater significance than something else which occurred less frequently.

Another limitation of this case study method was the problem of reconstructing past events regarding test score changes. There was a wide range in the quality and quantity of the historical data gathered in the schools, depending on the degree of staff turnover and the accuracy of the memories and the biases of the staff members interviewed.

In addition to the limitations inherent in the use of a case study methodology, several procedural limitations were encountered: the arbitrary nature of the school selection criteria; the difficulty of identifying a "trend" in test changes; the pressure of an unrealistic time schedule on the persons making the visits; and administrative constraints, such as allocation of time and personnel work load.

MAJOR FINDINGS AND IMPLICATIONS

The findings of the study and the implications of those findings for schools are organized in this section in a series of related themes: (1) the intent of ECE; (2) leadership and management of change; (3) expectations; (4) staff development; (5) teaching and learning in the instructional program; (6) curriculum; and (7) evaluation.

The topical themes developed in this section are a result of the case study methodology, the purpose of which was to capture the complexities and the interactions among those situations at schools that combine to affect the direction of student achievement. The themes emerged from the 16 individual case studies as those factors that appeared to operate and interact in unique ways within schools, but with common threads across schools. Since the methodology for this study was one in which many possible organizational frameworks for the findings and implications would be possible, the particular topics included in this section are only one group of the potential themes which could be addressed.

Within each topical theme are two sections. The first section presents the findings; the second section consists of one or more implications derived from the findings. The implications are based on the findings across the 16 case studies. All the limitations of these studies should weigh against too hasty an acceptance of the implications. It should suffice to say that one of the important tools which the study observers used was their individual and collective common sense, and this implies that the reader would do well to do the same when reflecting on the implications.

Throughout this section, the school names are coded. The eight schools with increasing test scores between 1973-74 and 1975-76 are shown as In1 through In8, and the eight schools with decreasing test scores are shown as De1 through De8. The eight schools which had high bilingual populations are indicated with a subscript "b," as follows: $In1_b$, $In2_b$, $In4_b$, $In7_b$, $De1_b$, $De4_b$, $De6_b$, and $De8_b$.

The quotations used throughout the findings sections were taken directly from the individual school case studies that were written by the observers soon after their visits to the schools.

The Intent of ECE--Findings

The translation of the intent of ECE into educational experiences for children was found to break down in various ways; the communication breakdowns were common in schools with decreasing test scores but were also present to a lesser degree in schools with increasing scores. In most of the schools with increasing scores, however, there was a general sense of educational purpose that supported some of the specific purposes of ECE and provided a more reliable means for appraising the utility of changes in the way these schools did things. Those who conducted the study identified the following four kinds of failure of schools to understand the intent of ECE, and the four will be examined here:

1. Know-how without purpose. Teachers, principals, parents, and children did things because they believed someone else (e.g., the state) required them to be done rather than doing things to achieve the intended purpose, which they were often unaware of. For example, some saw the school-level plan as a proposal required to procure outside funding rather than as the documentation of a planning process, the purpose of which was to improve educational effectiveness.

2. Purpose without know-how. Teachers, principals, and parents sympathized with major exhortations, such as "meet the needs of each child," and they wanted to accomplish the intent, but they didn't know how to do it.

3. Wrong purposes. Incentives that were originally designed at the state level to reinforce the intent of ECE were misused on unrelated local purposes often rooted in school or district politics or to the personal agendas of people in key positions.

4. No purpose and no know-how. Some information was not available to the people with operational responsibility, such as the principal, teachers, and aides. Communications from the state, the district, and the staff development programs tended to be abstract, impersonal, and difficult to relate to problems of the classroom and the school.

Know-how without purpose. In school De2, a school with decreasing scores, the observers concluded that the programs which were "designed around ECE requirements tended to emphasize reading and math ... done in relative isolation from application of the related skills 'which resulted in' a fragmented learner who may have acquired numerous 'reading skills' but could not apply them. There was a failure to connect the know-how of the reading skills to the purpose of reading."

A more extreme, but similar, case existed at school De5, where for "four days every week children spend their entire day in reading, written language, math, and PE, ... on the fifth day they have 'ECE' day ... to meet the ECE requirement for a balanced curriculum." Reading scores have been declining at this school since the heavy reading emphasis began.

In school De7, "the teachers thought that ECE required multi-age grouping, and the administration gave them little choice in the classroom arrangement." A standardized testing program was operating as a local evaluation, too, but "its results were not known to teachers nor used in the program."

At school $De8_b$, "reportedly in compliance with federal and state regulations against segregating program participants, the pull-out ESL (English-as-a-second-language) program was replaced by sending the bilingual aides into regular classrooms. At the same time, the ESL teacher was moved to serve grades 4 to 6 only. She was not replaced in K-3. Thus, there was a ... reduction in services to LES/NES children." In 1972, "the District Title I people told the teachers (at $De8_b$) that they had to individualize .... They were required to use a criterion reading system .... The district told them this was in obedience to state guidelines ..., it was just recordkeeping unrelated to the basal reader instructional program." When asked what ECE meant at their school, $De8_b$, "the most common answers were 'more aides' and 'learning centers.'"

The effects of these changes in pupil performance are discussed later under the heading "Teaching and Learning in the Instructional Programs."

Purpose without know-how. In school De3, several administrative and teaching staff mentioned "individualization" when asked about the impact of ECE. But "none (of them) felt comfortable in giving a definition of this term as used at the school." The observers at De3 concluded that "the 'correspondence course' effect of relying on the materials to do the teaching as a means of individualizing pace may be traced to the insufficient definition of individualization which reaches the school. Both the Department of Education and district communications to the teacher have concentrated on the role of skills continuums and the curriculum itself and have been silent on the role of the teacher. While there are very good reasons for the state and district to stay away from specifying the teacher's role, in a weak school like De3, this silence has the effect of encouraging the drift of accountability toward materials and the child's own study skills and motivation and obscuring the role of the teacher in teaching the skill to the student."

School $In2_b$ suffered declines in test scores in the first years of ECE and then rebounded and surpassed its scores at entry into ECE. The initial decline was associated with major changes in the first year: "There was an emphasis on individualization. (It) was defined as an instructional program permitting each child to work at his/her level of success while enjoying school .... It was considered the responsibility of the teacher to make these program elements work together to provide success for every child." In subsequent years the teachers did acquire the know-how to back up this purpose, but they didn't begin with it.

Wrong purposes. A number of examples were recorded of ECE being used by key local people to serve the wrong purpose, but only those examples that helped explain test score changes were recorded in the school case studies. Several of those are cited here. In school De3, for example, "The principal had manipulated the reelection of the advisory committee chairman through absentee balloting--even though the chairman was not a nominee for reelection. (Parents) ... complained of hiring practices by the principal as related to aides." At the same school, the principal told the observers that the only important change she could identify that was related to ECE was the assignment of a school site ECE coordinator who could help with the staff. "Primary teachers who were questioned about the role of the ECE coordinator could not describe what this person did."

At school De7, the teachers complained that "the aides were hired because they were favorites of those in control .... In this school, all of the administrators continually go to meetings and conferences, but the teachers are woefully uninformed."

No purpose and no know-how. In some schools a lack of purpose encompassed the whole educational process. When one teacher in school De2 was asked about the new programs, the teacher replied, "I don't really understand what they all mean, but maybe I haven't really tried."

In school De7, one of the major intents of ECE--to serve as an umbrella for coordinating all funding sources at the school--was not known, nor was there any articulation between SB 90 and the rest of the program.

At school De3, ECE was viewed "as little more than a limited source of funds ... the school had received Title I funds for several years previous to ECE."

Sufficient purpose and know-how. Some schools, mostly those with increasing scores, offered examples of the intent of ECE coming to life in fruitful ways. School $In2_b$ was turned around by parent involvement and planning. "At the end of the first and beginning of the second year of ECE, a few parents expressed concern about the instructional programs .... The staff listened and considered the parents' point of view carefully. In restrospect, the staff felt that the opinions of these parents contributed to decisions to make some of the changes described during the first two years of ECE .... During this school year (1974-75) changes were made to allow teachers to spend more time together in total school planning." The changes in the school which came out of these processes boosted scores considerably.

At school In4, teachers evaluated their instructional management system as "too cumbersome," and they designed a more suitable one.

At school In5, a good sense of purpose was manifested by classroom placements designed "to match children with the next teacher, according to children's needs and teachers' styles." Also at school In5, the parents played an active role in resolving the negative effect of discipline problems on learning.

At school In6, the observers reported "integration of multiple funding with prior and ongoing approaches to instruction and curricula ...."

Implications of Findings on Intent of ECE

The implications of the findings of the study as related to the schools' understandings of the intent of ECE follow:

- School, district, and state accountability procedures should be focused more sharply on the substantial purposes of education in general and ECE in particular, rather than on forms and procedures--no matter how fair they may be. Attempts to implement accountability have strayed into rewarding "looking good" behavior, such as keeping records on students but not using the records in some cases.

- Supportive services to teachers, aides, principals, and other school staff need to be more carefully designed for their respective users. Many of these school personnel need help in "how to do it" in terms recognizable to them. They need to know what to do in their own worlds, no matter where the support services originate--school, district, or state. Teachers, parents, and principals need to be better informed about ECE--what the options are, what the responsibilities are, what services are available, and generally what to expect. More information specifically directed to these groups is needed.

- Communication--whether written or oral, from the state, the district, or within the school--needs to be more frank and personal and needs to emphasize the primary role of human beings in educational processes. Too much of the information received by the teacher, the parents, and the children is couched in the abstract language of "programs," which evades the fundamental issues of what people are to do. In some cases, this evasion has resulted in pseudo communication, which tells the doer nothing except that someone else thinks the doer has been told to do something.

Leadership and Management of Change--Findings

The importance of leadership in managing or not managing change effectively was demonstrated at two levels: (1) district-level leadership, which served at three schools with increasing scores as a supportive force in the introduction and implementation of new programs and at a school with decreasing scores as an interfering or disruptive force; and (2) school-level leadership, which emerged frequently as a positive or negative factor.

In four of the schools with increasing scores, the building principal clearly exerted a positive leadership role; in two other schools, groups of teachers assumed school-level leadership in managing change; and in one school, the resource teacher emerged as the leader. Characteristics of positive leadership appeared to be widespread involvement and delegation of decision-making responsibility to those who must implement the decisions, mutual trust and respect for competence and judgment, anticipatory planning, and support for ongoing skills development and learning.

By contrast, at most of the schools with decreasing scores, either a leadership vacuum or negative leadership characteristics were noted. Authoritarian principals created an atmosphere of repression at two schools. In other schools, the absence of effective leadership was illustrated by lack of planning for changing school populations (particularly large influxes of Spanish-speaking pupils) until a crisis situation developed and by failure to exercise leadership in matters pertaining to selection and implementation of curricula and classroom organization.

School $In1_b$ provided an example of ongoing leadership and support from the school district: "In this small district there is easy access to district office services and a continuing dialogue between the schools and the administration." School $In2_b$ also benefited from the district's stance: "The district supported ECE by carefully selecting the staff for that first program year. The school was assigned a new principal, who was identified as being the most capable in the district of giving strong and positive leadership. School $In4_b$ was unique in the study in that the superintendent and principal were the same person. (He) ... started his assignment the same year that the ECE program was initiated. Staff turnover has been low ...."

Conversely, the district role could disrupt the school's functioning, as seen at school De7. There had been a "lack of leadership and support from the previous administration, and the school board ... had exacerbated the divisions in the community .... There have been three district superintendents, two building principals, and two resource teachers in the past four years." The board and several district administrators also exerted their influence by selecting a reading program and enforcing its exclusive use by district teachers. While the observers noted that the present superintendent appeared to have had the support of the teachers and to be attempting to rectify some of the district's problems, he was terminated by the school board.

The building principal, the traditional school-level leader, clearly exerted a positive role at schools $In2_b$, $In4_b$, In6, and In8. When the role of the reading teachers at school $In2_b$ needed examination, "Discussions about the possible role changes for those teachers were initiated by the principal and joined by staff and parents." In discussing the evaluation of the planned changes at school $In2_b$ which led to improved pupil achievement, the observers commented that "these changes were made because the principal and instructional staff were constantly asking themselves and each other how they could provide better services to children." Another function which the principal could encourage was planning. At school $In4_b$, "The teachers do extensive planning ... daily time is available ... released time is available upon request for within grade level planning ...." The staff at school $In4_b$ spent a year making plans for their school. "Under the leadership of the principal, this was a united staff effort ... requiring considerable extra time and commitment." The principal of school In6 served as a spark or initiator of needed change. "He suggested that when he sees a school problem, he'll make suggestions to the staff to work it through, they'll drag their feet, he'll lay out a proposal, they'll react and finally develop a strategy themselves .... When the principal came to the school six years ago, he challenged the staff to improve the children's achievement ... pledged his support of whatever approaches their professional judgment led them to undertake." The administrator at school In8 sets the tone for the school climate and the positive relationships among adults as well as between adults and children.

In schools where administrative leadership was weak, other avenues of leadership emerged. Both at schools In3 and In5, a core of teachers banded together and assumed the leadership role in planning, curriculum development and implementation, and school organization.

The resource teacher emerged as a leader at school In3. He introduced a reading program whose mastery, through the joint efforts of the staff and the coordinator, "became the heart of their staff development and program implementation."

The contrasts between positive leadership and no leadership were most apparent at those schools that had a changing pupil population, particularly an increasing number of LES/NES pupils. Two schools with increasing scores planned for these changes. At school $In4_b$, an influx of migrant students each fall increased the number of LES and NES students. Extensive planning time was allotted to teachers for planning for the use of aides and cross-age tutors, with the principal reviewing teacher plans every three weeks. At school In6, the number of Spanish-speaking children had increased slowly over the last five years. To accommodate the slowly changing needs, all children at the school were receiving instruction in Spanish. By contrast, at school

De2, where the population had changed very slowly over 15 years, the perceptions of the staff and the principal about an increasingly problem-ridden student population had stymied effective planning for the changing needs of the population. At schools $De4_b$ and $De6_b$, both with rapidly increasing LES/NES populations, the principals had encouraged their staffs to learn Spanish, but only as the changes became overwhelming. At school $De8_b$, although the relatively well-educated Hispanic population was being replaced by many LES and NES children from Mexico, the school's programs have only recently begun to acknowledge the changing student needs.

Positive leadership in selecting and implementing a new reading program was illustrated by school $In1_b$, where the reading program adopted four years ago received continuing district support and ongoing supervision. Similarly, at school In3, the core of teachers who selected and implemented the reading program were assisted by the reading coordinator, who became the ECE coordinator. At school In4, the teachers "are in the process of developing a new reading continuum more closely aligned to their newly adopted ... reading program." A fourth example came from school In6, where the principal expected the staff to take responsibility for the curriculum and classroom management ... clear that the school has a philosophy of reading ... developed in response to the principal's challenge." At school In8, "Teachers are encouraged to use whatever means ... most compatible ... to achieve reading goals ...." At schools with decreasing scores, decisions about curriculum were made arbitrarily at a level removed from the school staff. At school De5, a principal who administered the school in an authoritarian manner introduced the reading program in 1970-71, and no evaluation or review of the program took place until 1977. The implementation of the reading program did not have positive outcomes, because teachers had not had a role in the decision making. In school De7, it was found that administrative interference had occurred in a responsibility that should have been left to the teaching staff. The reading system, whose exclusive use was not enforced until this year, was one which some teachers believed "was not appropriate for the entry level skills of the children."

An example of effective leadership in implementing changes in classroom organization was found at school In5, where the multiage groupings were carefully planned by the core of teachers who assumed a leadership role at that school, "with children staying with the teacher until they passed to the next age span ... teachers try to match children ... according to children's needs and teachers' styles." At school De2, on the other hand, every teacher had a different perception about how children were placed in classes, causing frustration and friction among staff members. A multiage project at school De2 was implemented after a core of teachers convinced the principal to support the project. Only this group of teachers was enthusiastic, however, and the split between the project teachers and upper-grade teachers was not healed by the principal. His style was more to oversee and allow, but not to direct, guide, or attempt to influence others. At school $In2_b$, when the same K-3--4-6 split occurred, "The upper-grade teachers initiated discussions which resulted in a school reorganization and commitment to hold all meetings and in-service sessions as a total staff." At school De7, where district and administrator interference were negative influences on curriculum, there was also an adverse effect on school organization: "the teachers thought that ... ECE demanded multiage grouping ... the class/teacher match was done solely by the principal ... reasons for the changes in ... assignments ... were perceived as disciplinary ...."

Implications of Findings on Leadership and Management of Change

The implications of the findings of the study as related to the schools' leadership and management of change follow:

- Changing school programs means, above all, changing people's ways of doing things. The importance of good leadership, regardless of its source (principal, reading coordinator, group of teachers, or district), is fundamental. The potential damage from poor leadership is equally important. Good leadership entails, among other things, close contact and involvement with those who have operational responsibility for implementing decisions, such as teachers and aides; mutual trust and respect; anticipatory planning; and support for ongoing skills development. These and other good leadership characteristics should be fostered through personnel practices and policies, especially the appointment of school principals, and through staff development programs specifically designed to spread good leadership know-how to school, district, and state people in leadership roles.

- The leadership potential of a core group of teachers at a school should be acknowledged and supported when possible, perhaps through the participatory planning features of ECE.

- The role of the district in improving school programs needs to be clarified.

Expectations--Findings

One phenomenon which emerged from the school visits was that teachers' expectations for and beliefs about the children they taught often seemed to influence what pupils would learn and that administrators' expectations of teachers' performance seemed to affect that performance.

Teachers' expectations of their pupils were reflected in various ways: their perceptions of those children and their abilities, their selection of curricular material for those children, and their standards for pupil performance. In schools where the children were seen by the teachers as having poor oral language skills and "many problems," they did poorly; while at a school where a teacher referred to non-English-speaking children as being bright and learning quickly, the children performed well. Schools where the curriculum presented low-level tasks, rote response, and little else had decreasing scores; while at schools that had selected more comprehensive curricula and where mastery and competence were expected and demanded, scores were increasing.

Where teachers were held accountable for their pupils' performance and given accurate information about that performance, the teachers' behavior was more likely to have had a positive effect on pupils' test scores.

While the children at the schools visited varied in background and characteristics, the differences in perceptions and beliefs about the children were greater than the diversity of the children. The interaction of student characteristics with teacher (and other adult) attitudes and expectations is one which plays a role in the improvement or decline in reading performance, according to the case studies obtained.

In one school with declining test scores, De2, teachers described the children as having many more problems than in the past. There were comments about poor language skills, more family problems, fewer experiences--overall, children brought fewer experiences to school than children had in the past. In another school with declining scores, De6b, which had a very high proportion of non-English-speaking children, teachers commented on the children's lack of experiences upon entering kindergarten. By contrast, at school In7b, where scores increased, the teachers commented, in referring to non-English-speaking children, "Oh, these children are bright. They learn quickly." In a similar vein, observers attributed a score increase at school In4b to "more children attending preschool"--children perceived by teachers as "better prepared to come to school." A stark contrast between two schools with similar populations occurred. At a decreasing school (De5), children were described as being extremely limited in background and experiences, particularly in oral language. At the other school, In3, children were expected to perform at high levels of excellence; teachers demanded and obtained high quality performance. No one described to the observers the limitations of children, although much emphasis was placed on skills needs and skills development for adults.

The expectation that children cannot learn the more complex, integrated skills was illustrated at a school with very low and decreasing scores (De5), where a reading curriculum limited to one set of skills was implemented school-wide. School staff acknowledged the curriculum's inadequacy (children were not learning to read), but they viewed the program as positive because the children were succeeding at it.

The clearest statement of positive expectations came from school In5, where teachers said, "We know we can teach these children." The teachers believed children could learn and that they, the teachers, could teach them. Similar attitudes were expressed at school In3, where reading scores increased with the acknowledgment of pupil "teachability."

The phenomenon of test score decline cannot be attributed solely to low expectations, but rather either to "lowering" expectations or to interaction between expectations and other circumstances occurring at the schools. The interaction occurred most often in schools with changing pupil populations. At schools with decreasing scores, De2, De4b, De6b, and De8b, the changes in student characteristics were seen as offering fewer capable students, of whom less in the way of achievement could be expected. At schools De3, De5, and De7, on the other hand, the interaction was between low expectations for pupils and a newly introduced idea of individualization of instruction. "Each child at his (or her) own pace" seemed to translate to "each child at his (or her) slowest pace or lowest level."

As teachers' expectations for children's learning seemed to influence the direction of that learning, so did the expectations for teacher behavior on the part of those in positions to hold teachers accountable seem to influence teacher behavior. At De3, a school with decreasing scores, teachers were quite sure their students' scores were increasing because of information given the

school by the district office. No one had told teachers that their students were achieving less than in the past; no one held the teachers "accountable." At school In6, on the other hand, the principal clearly acknowledged the staff's role and his expectation that they could and would provide for children to achieve at higher levels than they had in the past.

Implications of Findings on Expectations

The implications of the findings of the study as related to the teachers' expectations for the children they taught and the administrators' expectations of teacher performance are as follows:

- Expectations for what children can learn need to be raised in many schools, especially those with high minority populations. This includes the teachers' expectations for:

  1. The type of activity (reading and writing words, sentences, and paragraphs rather than just filling in the blanks, rote response to low-level skills, games, puzzles, and expensive media attractions)

  2. The pace of each child's progress (each child making optimal progress consistent with the child's development rather than each child as slow as he or she feels like)

  3. The quality of children's work (regardless of type or level of work, the teacher should expect good workmanship from the child rather than half-hearted, sloppy efforts)

Staff Development--Findings

Three basic types of staff development programs were observed in this study:

1. Ongoing in-service training closely tied to the instructional program. This type of program was found to be effective in impacting on staff behavior and classroom practices in five of the schools visited. The programs had a strong "how to do it" emphasis; and they were an integral part of the main-line operation of the school, with extensive classroom follow-up. Although all five in-service programs had an impact on staff behavior, in three schools there was a positive effect on pupil performance; while in two schools the impact was negative.

   The characteristics that distinguished the programs that contributed to pupil performance were the following: the training adapted the new program to teachers' preexisting instructional practices; the program was adequately comprehensive, and the teaching staff was committed to the program. By contrast, the two ineffective staff development programs of this type were enforced on a reluctant and antagonistic teaching staff by an authoritarian administration.

2. In-service training designed to improve staff capacity to serve ethnic and language minorities. Most of the schools studied had significant minority populations, but at only four of them did observers note in-service training designed to improve staff capacity to serve ethnic and language minorities. Courses in multicultural understanding were

mentioned by observers as having had a positive effect on LES/NES achievement. Three schools offered Spanish-language courses for teachers.

3. <u>In-service programs on an assortment of topics without specific relationship to the teacher's responsibilities</u>. These were nonspecific presentations that were deemed to be insufficiently integrated into the instructional program, and the presentations were usually designed for a general audience of educators. Their impact on teacher behavior and performance was negligible.

Although not included in one of the above categories of staff development programs, the informal exchange among teachers as they worked together to plan the K-3 program was found to be a significant outgrowth of ECE. The world of the self-contained classroom has been opened up to fellow teachers, aides, and parents. Where things went well, this fostered an increase in the sharing of ideas, techniques, materials, and commitment, which were beneficial to all. Where things went poorly, this opening-up process bred insecurity, retrenchment, and shared excuses.

One reason for the effectiveness of the ongoing in-service training that was closely tied to the instructional program is the fact that it was people specific; that is, aides were trained for aides' jobs and teachers were trained for teachers' jobs. Training covered details of what to do in the classroom in frequent sessions that continued throughout the school year.

In school $In1_b$, the reading program that was adopted was developed by another district. It was a comprehensive program that included extensive in-service training provided by the developer district and backed up by the new reading specialist at school $In1_b$. In school In3, "Many of the teachers at the school attributed the improved reading of the children to the brand X reading program and specifically to the reading specialist who taught them to use it .... One teacher challenged the ... specialist to demonstrate (the program) in her classroom for two weeks. This he did and won her respect. She diligently learned the system ..., and--by many accounts--does it better than anyone." Small groups of In3 teachers and aides attended in-service sessions before and after school three times a week until they learned the system. Unlike schools $In1_b$ and In3, school $In4_b$ did not have a dominant reading system. There was more variety from teacher to teacher but a common emphasis on the use of aides and cross-age tutors during reading instruction. "While there were pre-service and in-service training programs for both the aides and the tutors operated by the ECE coordinator, the major training of these people was by the teacher they worked with ... daily time is available for joint planning." These daily sessions had the effect of training the aide for the next day's work. At school De5, a commercial programmed reading system had replaced almost all of the traditional curriculum (except on Fridays). "Extensive in-service (training) was provided by the publisher's representative for teachers at the outset of the program." School De7 was similar to De5 in that a single commercial reading system replaced almost all of the curriculum.

Each of the five schools with strong ongoing staff development programs made the programs part of its self-management and provided significant follow-up in the classrooms. For example, in school $In1_b$, the reading specialists who

conducted the initial in-service training returned regularly to observe the implementation of the program in the classroom. In school In3, (the specialist) "would go into classes to observe the implementation of various skills from the staff development, correct any misconceptions, and base the next workshop on these emerging needs ... the teachers were taught and taught until they were competent ...." In school $In4_b$, since much of the in-service training was teacher-to-aide, the in-service trainer was observing the trainee on a daily basis.

In school $In1_b$, the feedback to teachers from the in-service training staff was "tactful and to the point." In school In3, the reading specialist who conducted the in-service training "emphasized to us that he learned as much from the teachers as they learned from him. He encouraged them to keep doing all the good things they were already doing, and he showed them how to incorporate their favorite lessons into the (program)." In school $In4_b$, where each teacher worked with her aide and tutors, "teachers select their aides and aides select their teachers. There has been a stable aide-teacher relationship."

On the other hand, in the two schools with decreasing scores, the reading programs were installed and operated in a more authoritarian manner. As a result, the staff development was an instrument of control over staff behavior. When there was resistance to the administrative controls, there was resistance to the reading program and the in-service training, which was part of its enforcement.

Implications of Findings on Staff Development

The implications of the findings of the study as related to staff development are as follows:

- Staff development at the school level should be more closely tied to the particular instructional program in use, backed up by observations in the classroom and by feedback to teachers and aides on their performances. The emphasis should be on "how to do it" and be people specific: Aides trained for aides' jobs, teachers trained for teachers' jobs. The sessions work best when they are frequent and conducted by someone who can demonstrate the substance of each lesson with children.

- Staff development should be built into the main-line operation of the institutions, rather than operated as an adjunct or external project. Staff development, staff management, and staff accountability are all part of the main line of responsibility, not extra projects.

- In-service training should help persons adapt new programs and skills to the preexisting practices rather than merely enforcing compliance to a new program.

- Staff development programs should not be designed to focus the teacher's attention narrowly on any single curricular area to the exclusion of others.

- Staff development programs, in addition to teaching know-how, should foster a clear sense of purpose and commitment related to the particulars of the program.

- Staff development activities that are not specifically related to the planned instructional program of the school should receive a lower priority and lower allocation of resources than more fruitful activities closer to the school. Attending conferences, workshops on tangential topics, and workshops of a general nature and conferring with outside consultants not familiar with the school program are activities which should supplement rather than supplant the primary staff development effort.

- State, county, district, and school people should work together to identify and disseminate the kinds of information that is needed by teachers and principals.

- State, county, and district workshops should have a balance of know-how and purpose in the content area.

Teaching and Learning in the Instructional Program--Findings

Attempts to individualize the instructional program with the advent of ECE had positive effects on pupil achievement when people were responsible for learning, and they had negative effects on achievement when people abdicated their responsibility for learning to programs.

Four characteristics of the teaching-learning practice were observed in schools where test scores increased: (1) initial teaching was done by people rather than by media, material, or learning station; (2) diagnosis of pupil needs was an integral part of teaching and learning, not an external add-on; (3) prescription included teaching by the teacher (or aide) rather than just matching to curriculum levels; and (4) follow-up monitoring of the children's ability to use new concepts was a part of the instructional program. When misuse of these practices occurred, decreasing pupil performance also occurred.

Schools at which learning centers were used to reinforce skills already introduced by the teacher were successful in increasing pupil performance, while schools at which teachers relied on curricular materials to do the teaching (as in a correspondence course) were unsuccessful. When children's work went uncorrected or uninterpreted for long periods of time, the effects of early misunderstandings were reinforced by usage, resulting in poor pupil performance.

Initial teaching by people rather than by media, material, or learning station. It was commonly observed in schools with increasing scores that new subjects were taught by the teacher interacting directly with the children in some way. For example, in school $In1_b$ "... all skills are initially taught by the teacher in her station." In school $In2_b$ there was a change in the role of the teacher from the first year of ECE (when scores were lower) to this last year (when scores were up). At first "... the child spent a lot of time working on his or her own, or with an adult .... The child was expected to take a great deal of responsibility for his or her instructional program and learning .... While the independent, self-motivated child really thrived in this learning environment, the majority of the students were not progressing as fast as they should have." The parents, and later the teachers, realized this scheme was not working, so now the teachers "provide all initial reading skill instruction, and students use centers only for reinforcement." In school In3,

the teachers worked with the children until each one of them demonstrated understanding of the new subject matter. In school In6, a parent who moved from another school said she liked this school, because "they use people instead of machines with the children to individualize instruction."

In contrast to the schools with increasing scores, the teachers in schools with decreasing scores tended to rely on materials, media, and learning stations to teach their lessons. Teachers in these schools were characteristically preoccupied with managing pupil movements or assigning curriculum materials. School De5 had a highly specified system of instruction: "Children at this school spend four days a week in a commercial, programmed language and reading program. Their entire morning's activities consist of working in workbooks or on worksheets, filling in blanks, circling choices ...." In school De7 the observers got the impression that "teachers, in general, spent more time controlling than teaching. An inordinate amount of time each day is spent on (dittoed materials) with little or no teacher direction." In school $De8_b$, "Children race through the workbooks and follow-up exercise without really learning how to read."

Diagnosis as a natural part of teaching and learning, not an external add-on. The heart of the diagnostic process in many of the schools with increasing scores was found to consist of the teacher paying close and frequent attention to how each child responded to instruction. Formal assessment devices supplemented this natural process but did not replace it. After instruction by the teacher in school $In1_b$, the children worked at various follow-up "stations" where "no child's work goes unchecked." School $In2_b$ switched to a split schedule, which "meant even more direct and carefully monitored instructional time for students in reading." In school In3, "Transactions between the teacher (or an aide) and the child regarding the correctness and quality of the child's work were frequent and timed harmoniously with the child's tempo of work .... There was an unusual emphasis on using teachers' listening to the student as part of the instructional transaction ...."

In school In5, it was reported that "while working in directed teaching situations with a small group of children, teachers were well aware as to the activities of children working independently and in small groups in other parts of the room. They extended help to those who needed it and redirected the energies of those whose attention may have wandered."

In school In8, the teacher's monitoring of student progress was supported by a specification of the particular skills for which the teacher was responsible. This was supplemented by weekly feedback from the reading specialist on progress in specific skills.

In schools where diagnostic procedures were not integrated into the teaching and learning processes, teachers were apt to complain about "all that record keeping and paperwork" that goes into skills inventories and pupil profiles. These records were perceived as being required by ECE and for ECE rather than for teaching and learning. At school De5, with its commercial programmed system, control and responsibility for diagnostic decisions were vested almost entirely in a mechanistic, programmed learning system. In school De7, questioning the children revealed to the observers that they were doing independent

work they were not prepared for. Yet, the teachers were not aware of this. In school $De8_b$, the pupils' work was left unchecked for long periods of time, according to some teachers.

Prescriptions included teaching by the teacher (or aide) rather than just matching to curriculum levels. In many schools with increasing scores, the prescription flowed naturally from the ongoing diagnostic attention teachers paid to children. In school In3, for example, when examining the child's work in progress, the teacher "would correct any mistakes and engage in some dialogue on the substance of the mistake."

On the other hand, in some schools with decreasing scores, the prescriptions were often supposed to be "self-teaching" materials without the direct instructional interaction with the teacher. In school De3, the staff relied more and more on the curricular materials to do the teaching (as in a correspondence course). In school De5, the "failure to master the content of a step or level at 100 percent means that a child must complete more exercises and activities of the type he or she could not do on the test." In school De7, while the teachers "did review the (children's) work daily, ... children were then to correct their errors, usually without teacher assistance...."

Follow-up monitoring of the children's ability to use new concepts. In the classrooms of many schools with increasing scores, once the children demonstrated initial understanding, they used the new concepts and skills in a variety of contexts. During this phase of learning, teachers and aides observed the children closely, identifying problems which frustrated particular children and intervening with timely reteaching and follow-up. In school $In1_b$, after initial teaching by the teacher, the children "move to work with aides, ... tutors, parents, ... to follow up the skill. The follow-up activities may include workbooks, tapes, art projects, or free reading."

School $In2_b$'s scores began to improve when, among other things, "centers began to be used to reinforce skills already introduced by the teacher rather than to teach a skill for the first time, as in the previous year." In school In3, "Teachers linked experiences throughout the day to reading and writing skills, whether during 'key word' or 'story writing' or free reading or social science or other areas of curriculum." Children in school In5 were observed to benefit from "the follow-up teaching done in the reading and math resource rooms." Children stay until they pass a criterion test on the objective of the lesson. In school $In4_b$, pupils "received immediate assistance from either an aide or a cross-age tutor when working on reading-related activities ...."

In contrast, teachers at school De3 reported that "they couldn't get around as much as they thought they should to see what progress children were making in their assignments, or even if they were doing them correctly. Sometimes a week or more would go by ...."

At school De5, the children's school life was dominated by a single commercial programmed reading program. There was little opportunity for them to use the skills being taught in any context other than the commercial materials, which did not include books other than workbooks; and, consequently, the teacher had no chance to monitor the use of skills.

Implications of Findings on Teaching and Learning in the Instructional Program

The implications of the findings of the study as related to the teaching and learning in the instructional program are as follows:

- Initial teaching should be done by the teacher (or some other person) rather than by media, material, or learning station.

- Diagnosis should be a natural part of teaching and learning rather than an external add-on. The teacher should pay close and frequent attention to how each child responds to instruction, and the teacher should use formal assessment devices as an occasional check rather than relying too much on formal assessment. The children's work should be used as one of the primary sources of diagnostic information.

- Prescriptions should usually include teaching by the teacher rather than being limited to selection of appropriate materials.

- There should be follow-up monitoring of a child's ability to use new concepts in a variety of contexts, such as using reading skills to read a science book. Reteaching should be available when problems which frustrate particular children are identified.

Curriculum--Findings

A significant finding which emerged from this study was that in schools in which the curriculum consisted entirely or mostly of reading, reading scores were declining, whereas in schools in which pupils had an opportunity to use and apply reading skills in a range of curricular areas, scores were increasing.

Three characteristics distinguished the curriculum of schools with increasing scores: (1) teachers felt ownership of the curricular programs they used (pupil achievement suffered when teachers were not committed to the program they were teaching); (2) the content of the curriculum was appropriate to the particular needs of the pupils (it was sufficiently comprehensive in scope, could be adapted to meet the needs of LES/NES pupils, for example); and (3) the implementation of the curriculum was accompanied by effective staff development and follow-up activities.

The seemingly paradoxical finding that reading scores declined in schools in which the curriculum consisted entirely or mostly of reading was illustrated at school De5, a school with decreasing scores. There, strong authoritarian leadership led to the introduction of an instructional program consisting almost exclusively of reading. Reading "instruction," however, consisted almost entirely of children completing worksheets, workbooks, and other independent activities without teacher or aide instruction. While teachers did monitor student work, the result of a teacher and child checking work was nearly always the assignment of another worksheet, workbook, or activity--rarely the intervention or explanation of a missed idea or a new concept.

In addition to a curriculum consisting entirely of reading and "instruction" occurring to a limited extent, the "all reading" curriculum at school De5 was itself very much unbalanced, with emphasis entirely on decoding and phonetic

analysis skills. This one-sided emphasis was in marked contrast to the curriculum at school $In1_b$, for instance, where all aspects of reading were carefully and comprehensively taught; this school was utilizing state textbooks, plus their own writing, phonics, and reinforcement techniques. At school In6, "a broad emphasis on reading is accomplished with the library, field trips, and projects that relate reading to other parts of the child's world ...."

A contrast with the breadth at school In6 occurred at school De7, where "the previous principal ... made regular visits to classrooms to be certain that only the approved program was in use ... and ... teachers were to teach only reading and mathematics."

Although the presence of a particular system or lack of one did not appear to influence reading scores, the degree to which teachers felt ownership of the programs was significant. At one school with decreasing scores, $De8_b$, "The teachers recall being very hostile to the dictate that they ... use a 'criterion reading' system selected by the district ... it was never a prominent part of the children's day, nor was it a major feature of the way teachers organized the curriculum or their instruction."

At school In3, teachers all used the same reading skills sequence, which they had enthusiastically adopted after some initial resistance to the program. Having seen its successful use demonstrated, teachers learned. On the other hand, the same program, used in another school ($De4_b$), was inconsistently and unenthusiastically implemented.

At school In6, most teachers used a commercial system, which they had adopted to meet their pupils' needs; this school staff had had a year to plan, review, select, and develop a reading program appropriate for their pupils and in keeping with teachers' styles and preferences. Consistency between classrooms or grade levels seemed less important to increasing test scores than a belief on the part of teachers that whatever they were using was their own choice and under their control.

At school In6, where test scores increased, the reading program in each class reflected the teacher's judgment, with materials including basal readers, former state adopted textbooks, and four different commercial programs.

There were no alternatives for teachers at school De7, where all supplementary materials and library books had been auctioned off after the purchase of the reading management system, which the teachers thought was unsuitable for their pupils.

The curriculum for LES/NES students was seen as less than adequate in some schools with increasing scores as well as in schools with decreasing scores. In the schools with high LES/NES enrollment where test scores had decreased, the absence or inadequacy of bilingual programs was evident, particularly in cases where the number of LES/NES pupils was increasing. On the other hand, at the high LES/NES schools with increasing scores, the population changes were not as rapid; and the attitudes of the school staffs were much more favorable toward the ability of the LES/NES students to learn. Nevertheless, even the schools with increasing scores often failed to provide bilingual instruction, leading the observers at one school to comment that pupil reading comprehension might well drop by the sixth grade, despite score increases in the early grades.

Implications of Findings on Curriculum

The implications of the findings of the study as related to curriculum are as follows:

- Students should use and apply the skills of reading in a range of curriculum areas. Schools with a curriculum consisting almost entirely of reading suffered declines in reading performance.

- Within the reading curriculum itself, a balanced variety of reading skills should be taught. Scores declined when there was overemphasis on decoding and phonetic analysis and underemphasis on integrated skills, such as comprehension and story writing.

- Steps should be taken to ensure teacher ownership of newly adopted curriculum programs. When teachers did not believe in the program, it was not implemented well, or at all! Teachers should be involved in selecting and adopting new programs.

- Teachers should understand how to use the curriculum and the purpose of each of its parts. This is a staff development need.

- The curriculum should meet the needs of the particular students enrolled, especially students with limited- and non-English-language development. Attempts to individualize the curriculum often emphasized systems that were designed to manage children working at different paces but failed to provide appropriate curriculum to children whose individual needs were other than individual rate of learning, such as not speaking English.

Evaluation--Findings

Four problems with evaluation were common in schools with increasing scores as well as schools with decreasing scores: (1) school personnel had minimal information or misinformation about the tests and test results; (2) schools failed to use what information was available; (3) because monitor and review (MAR) ratings and test scores measure different aspects of school functioning, schools sometimes had difficulty knowing how to integrate and interpret both types of information; and (4) the means available for identifying and assessing limited-English-speaking and non-English-speaking (LES/NES) children and for evaluating bilingual education programs were inadequate or nonexistent.

School personnel at several schools could not recall the California Assessment Program (CAP) reading tests, and no teachers questioned had seen test results from CAP testing. Many teachers and administrators did not distinguish between the CAP tests and other standardized tests they administered. In two instances, school personnel believed test scores were decreasing when, in fact, they had increased; in another school, the reverse was the case. Observers found only one instance in which a program had been modified as a result of either the monitor and review evaluation or the CAP test results.

A serious problem at the schools with high bilingual populations was the fact that pupils whose primary reading instruction was in Spanish were tested in reading achievement by an English-language instrument.

Confusion was evident at several schools about "what tests were what." Personnel at few schools could recall the California Assessment Program (CAP) reading tests; at school $In1_b$, "It required showing them a copy of the test (CAP) before they even remembered giving it." At other schools, teachers and administrators did not distinguish among the various standardized tests administered; "state tests" most often meant the commercial standardized tests required by the district.

In addition to the confusion about the tests, observers also found that school personnel were frequently vague or misinformed about pupil outcomes from testing. At schools In5 and $In7_b$, teachers believed the test scores were decreasing, and the observers had to explain the tests and show the outcomes before teachers would believe them. At school De3, on the other hand, a district presentation of standardized test results led the teachers to believe with great certainty that the scores at that school were improving, when, in fact, they were not.

The observers found the CAP reading test to be very "low profile" in all cases; school administrators knew little if anything of the outcomes, and the district offices apparently did not provide interpretive help to the schools. At school De7, although the CAP test profile showed clearly the lack of balance in the instructional program, no use was made of the information. Children succeeded on phonetic drills and scored low on comprehension, a pattern of which no one at the school was aware, apparently because no one had looked at the test printouts.

School personnel were also ill-informed about the content of tests used. At school De5, the teachers claimed that children's poor performance was based, in part, on "the content of the test, which is seen in some ways as not relevant to these children." When asked what the content of the test was, however, the teachers did not know.

Administrators and teachers also tended to be uninformed about outcomes of standardized tests other than the CAP; they looked at test scores but did not look at subskill results or at individual pupil results. An example of the positive use to which such information could be put was at school $In2_b$, where teachers became aware of the needs of students and modified the instructional program accordingly.

School $In2_b$ was also an example of the difficulty which schools face when two forms of evaluative information provided by the Department appear to be contradictory. Since MAR data and student achievement are designed as independent measures of school functioning, schools seemed to have difficulty in knowing what direction to take for program improvement when different ratings seem to point in different directions. The ratings given on the monitor and review (MAR) had been very high in school $In2_b$'s first year in ECE. However, by the school's third year in ECE, while the modifications they had made in the instructional program led to improved student reading achievement as measured by CAP, their MAR ratings had decreased sharply. The opposite situation occurred at school $De6_b$; a high MAR rating in 1976 accompanied low and decreasing test scores. Whatever the aspects of the school's program were that earned them favorable MAR scores, they did not lead to improved student reading achievement. School De 5 had had a favorable MAR rating in 1974, at which time no mention had been made about the inadequacy of the school's curriculum. While schools $In7_b$ and

De2 could both recall having had a monitor and review, neither could cite any information which had led to program modification or improvement. School De7, monitored shortly before the special study team's visit, had received ratings of adequate or average, but team observers two months later noted a curriculum consisting entirely of reading and math.

Procedures for assessing the English language competence of children whose primary language was other than English were erratic and apparently unreliable. At school $De8_b$, three separate lists of LES/NES children contained different numbers as well as different names of children. At school $In1_b$, while school personnel reported that 67.8 percent of the children were bilingual, very few children were identified as LES or NES. From the information at school $In2_b$, it was doubtful that LES/NES students were accurately assessed to determine their language dominance. At school $In7_b$, no language dominance assessment took place prior to 1977, despite reports that up to 50 percent of the children were LES or NES.

Even when children were adequately assessed for language dominance--a rare occurrence--other assessment was lacking. At school $De6_b$, although every child was assessed for language dominance and over 70 percent of the children were identified as LES or NES, no assessment of Spanish language competence or reading skills was carried out. Instruction took place only in Spanish in as many as six first and second grade classes at this school, but assessment of children consisted of standardized testing in English. Neither the district nor the state had provided incentive or means for testing Spanish language competency or reading achievement. A similar absence of assessment procedures occurred in all the sample schools identified as "high bilingual."

Implications of Findings on Evaluation

The implications of the findings of the study as related to evaluation are as follows:

- In-service training is needed on how to use tests and test results in the district, in the school, and in the classroom.

- The strategic implementation parts of ECE--planning, implementation, internal and external program quality review, including MAR and evaluation--need to be more explicitly connected to each other. The ongoing interaction among these parts has been overlooked in some schools.

- The link between program management and evaluation needs to be more functionally explained at the school level. This would entail dealing with the reality that the program that is managed, improved, and evaluated consists of people and their behavior. At the school and classroom levels, this fact is so plain that talk of program is often enough to mark one as out of touch. This was especially so in schools with decreasing scores.

- In some schools with full bilingual programs, there is a heavy emphasis on learning in the primary language in the early years and learning English in the later years of elementary school. There should be some way of assessing the success of the primary language learning and the success of teaching English later. Currently, there is minimal evaluation in these areas. Instead, there is English language testing of children who have been instructed in Spanish, for example.

III. Classroom Process Study

Two complimentary research strategies were selected to answer the most important question that was examined in the study of third grade reading scores: What circumstances are associated with the decline or improvement of third grade student reading achievement? One of the strategies was to conduct a series of case studies, and the results of that examination were reported in Chapter II. The second strategy was to conduct a traditional correlational study of variables within classrooms, and the State Department of Education commissioned Stanford Research Institute (SRI) International to make that study. Jane Stallings of SRI International was in charge of the study, and a summary of the results is presented in this chapter.[1]

The SRI International study of early childhood education (ECE) classrooms was designed to be a replication of previous research in elementary schools; in particular, the SRI National Follow Through Evaluation (Stallings and Kaskowitz, 1974) served as a model for key aspects of the study. In accordance with the design of the follow through study, data collection procedures, instrumentation, and analytical techniques previously found to be effective were adopted for use in ECE classrooms. A brief description of the procedures follows.

PROCEDURES USED IN THE STUDY

Fourteen elementary schools in 11 California school districts were selected for the SRI study by the staff from the Office of Program Evaluation and Research, California State Department of Education. All of the schools had participated in early childhood education since 1973 or 1974. The schools for the study were selected from those whose 1973-74 Entry Level Test (ELT) scores were in the lowest 20 percent of all schools.

The criteria for the selection of schools with increasing or decreasing scores were the same as those used for the selection of schools in the State Department of Education study.

The criteria for schools with decreasing scores were as follows:

1. Schools must have been in the lowest 20 percent on the ELT in 1973-74.
2. A school's third grade reading scores must have been lower in 1975-76 than before the school's entrance into ECE.
   a. For Phase I schools (those which entered ECE in 1973-74), the 1972-73 percentile rank had to be lower in the spring, 1976, than in the spring, 1973.
   b. For Phase II schools (those which entered ECE in 1974-75), the 1973-74 percentile rank had to be lower in the spring, 1976, than in the spring, 1974.

[1] Full details of the sample, instrumentation, classroom observation procedures, and objectives of the analysis are contained in the technical report of the study, which is available upon request from the California State Department of Education, Office of Program Evaluation and Research.

3. A school's third grade scores must have decreased at least four raw score points from 1973-74 to 1975-76.

4. A school's third grade mean score on the California state assessment Reading Test must have been below its prediction band in 1975-76.

5. A minimum of 20 third grade students must have been tested in 1975-76.

Schools within the decreasing and increasing score categories were then classified as having either low or high bilingual enrollment. If their bilingual percentile rank on the California Assessment Program bilingual index was below the 75th percentile in the distribution of all schools in the state, the school was classified low bilingual; if the school was above the 75th percentile, it was classified high bilingual. The final sample of schools was then chosen randomly within each of the four categories (decreasing high bilingual, decreasing low bilingual, increasing high bilingual, and increasing low bilingual) to produce a balanced distribution of bilingual schools. The final sample of schools used in the SRI study included the eight schools that also were included in the State Department of Education study.

SRI contacted the schools and arranged for recruitment of third grade teachers from each school. This effort resulted in the recruitment of the targeted number (three) of third grade teachers from each school, with the following exceptions: in one school where SRI expected to recruit six teachers, only four third grade teachers were available (all of whom participated); at two other schools four teachers instead of three wanted to participate, and SRI felt it would be unfair to exclude one arbitrarily; and one school had only two third grade classrooms. Two teachers in different schools were sick during data collection activities and were unable to be present for the observations in their classrooms. The initial sample and the reduced sample distributions are presented in Table 1.

The criteria for schools with increasing scores were as follows:

1. Schools must have been in the lowest 20 percent on the ELT in 1973-74.

2. A school's third grade reading score must have been higher in 1975-76 than before the school's entrance into ECE.

   a. For Phase 1 schools, the 1972-73 percentile rank had to be lower than the 1975-76 percentile rank.

   b. For Phase II schools, the 1973-74 percentile rank had to be lower than the 1975-76 percentile rank.

3. A school's third grade score must have increased at least four raw score points from 1973-74 to 1975-76.

4. A school's obtained reading score on the California state assessment Reading Test must have been within or above its prediction band in 1975-76.

5. A minimum of 20 third grade students must have been tested in 1975-76.

Classroom Observation Procedures

The SRI classroom observation instrument was structured to provide a description of activities and interactions that occur in classrooms. Through the use of well-defined codes, the instrument yields a record of classroom activities, the physical environment, the organization of groups, and the interactions among the teacher, aides, and children.

The classroom observation instrument contains three major sections:

1. Classroom summary information section
2. Physical environment information section
3. Classroom observation procedure, which consists of three parts: the classroom checklist, five minute observation preamble, and five minute observation.

In a single observation day, the first two sections of the instrument are completed once; and the third section is completed four times an hour for five hours a day. From these coded data, variables are formed which describe classroom instructional processes.

Four classroom observers were selected on the basis of their personal ability, education, attitude, and experience in working with children. Two of these observers had used the SRI classroom observation instrument for the Teacher Effectiveness Study at Stanford University. The other two observers did not have previous observing experience, but both had the educational qualifications and had participated in field-based research studies.

All four observers were given pretraining materials to study prior to attending the training session. Trainees were expected to arrive on the first day of training with a knowledge of the codes and a general understanding of the system.

Both observers whom SRI had previously trained to collect observation data had demonstrated competence in the use of the system. The observation system used in the Stanford study was exactly the same as that used for the State Depart-

Table 1

Distribution of Sample Used in SRI Study

Number in initial sample			Number in reduced (final) sample		
Teachers	Sites	Total	Teachers	Sites	Total
2	1	2	2	3	6
3	9	27	3	7	21
4	3	12	4	3	12
5	0	0	5	0	0
6	1	6	6	1	6
Total	14	47	Total	14	45

ment of Education study. These two observers had been using the observation system throughout the 1975-76 school year and had been checked frequently for reliability. Therefore, it was decided that a two-day refresher session would be sufficient training for these two observers. Immediately after completing this two-day session, they were assigned to classrooms to conduct observations. While these two observers were observing for the data collection, a five-day training session was held for the other two observers who had not had previous experience with the SRI observation system.

The prevailing strategy of the training session was to expose the trainees to as many different classroom situations as possible and to allow for discussion as to how these events would be coded on the observation system.

Most of the training time consisted of showing trainees video tapes of elementary classrooms, coding the tapes with the observation system, and answering any questions which arose from the coding. On the second through fourth mornings, trainees conducted practice observations in elementary classrooms in the Menlo Park/Palo Alto area. These practice observations were approximately one and one-half hours long. The trainees then returned to the training room and discussed any coding problems that arose during that time.

In addition to the coding from video tapes and the classroom observations, trainees were assigned coding exercises (dialogues of classroom situations) to complete in the evening. Each morning the work from the previous evening was reviewed and discussed.

Throughout the training session, reliability checks were conducted. The trainees and trainer would code a video tape, and an individual observer would make a tally of the frequency of each code. This procedure helped the trainer learn which codes required more clarification.

On the fifth morning the reliability tapes were shown. These were tapes of classroom situations that the trainees had not previously viewed.

Reliability tests of the two new observers were conducted on the last day of the five-day training session (pretest reliability). The two experienced observers were examined for proficiency during their two-day session. The reliability of all four observers was recorded again two weeks after they started data collection (post-test reliability).

The afternoon of the fifth day was devoted to the logistics of the study. Observers were given the observation schedules and materials required. At this time, the importance of the confidentiality of the observations was stressed. They were told not to make judgmental comments to teachers, positive or negative, regarding the environment in the classroom. Observers were instructed to inform the school administration of their schedule when they arrived at the school and to be as unobtrusive as possible in the classrooms.

Data Collection Procedures

The primary goal of the data collection activities was to collect data that accurately described classroom settings and teacher and student behaviors. To do this, it was important that all field activities were carried out consistently. A second concern was that schools and classrooms be disturbed as little as possible by SRI staff so that disruption of instructional activities would be minimal.

The data-gathering activities were conducted through a series of visits to each site. The schedule of visits was as follows:

Site visit	Major purposes	Date
1	Introduce SRI and the study recruit teacher for sample	March, 1977
2	Classroom observations	March-April, 1977
3	Ethnographic observations of classrooms	April-May, 1977
4	Collect attendance data	June, 1977

Observation data. All visits were coordinated with local school principals and district officials who played important roles by advising SRI of desirable times for visits and by contacting teachers in advance of those visits. At the first meeting a date was set for the next visit by the classroom observer. The data collection period was typically an entire week within which teachers chose among themselves which days would be most appropriate for observation in their respective classrooms. This procedure was used to allow for field trips, school assemblies, and other activities that took students out of the classroom. At schools with three participating teachers, five days were spent at the site, two full days each for two of the teachers and two full days for the third teacher. At all other schools (e.g., those with two, four, or six participating teachers), there were two days of observation per teacher. Each observation "day" consisted of the first five hours of the school day (if there were five hours), including recess periods and physical education classes but excluding lunchtimes. Children were not observed individually, but they were all identified on each classroom checklist.

Several weeks after the classroom observations had been finished, the SRI project leader visited a sample of participating schools to make ethnographic observations of third grade classrooms. One day was spent at each site visited to observe and answer questions about the study from the school principal, staff specialists, or classroom teachers.

Attendance data. A final site visit was required in the four sites where the collection of class attendance data was incomplete. Some teachers were unable to provide the data before they left for the summer vacation. SRI staff visited these sites and, with the assistance of district personnel, obtained the relevant data. In the other ten sites, complete attendance data were forwarded to SRI at the end of the school year.

Test data. Initially, SRI intended to use the achievement test data collected by each school in the fall, 1976, and the spring, 1977, as its dependent variables. The SRI personnel thought these test scores would be available from districts long before the California Assessment Program scores could be obtained from the State Department of Education. During June and July, SRI staff contacted district directors of research to obtain the required classroom level pretest and post-test scores for fall, 1976, and spring, 1977. Although data were received from all sites, the diversity of the tests used and dissimilar district testing schedules precluded the use of these scores in the SRI analysis.

An additional problem was that four districts tested only in the spring of each year. Interpolation of fall test scores in these districts from the two spring data points was not feasible.

For the reasons cited above, the decision was made to use the California Assessment Program (CAP) test data as the dependent achievement measure in the analysis. CAP files were assessed to locate matched samples of children for each classroom in the study for spring, 1976, and spring, 1977. Since this request required special handling by the testing and data control personnel in the State Department of Education, these data were not obtained until late September. The data from one school and from three classes in another school were missing entirely. Although school level scores were available for those schools, the classroom scores could not be computed, because the appropriate box to identify students in SRI's study had not been marked.

The findings from this study included descriptions of the participating schools in terms of a number of background variables as well as the main focus of the study, which was the analysis of the relationship between classroom level instructional process variables and student reading achievement gains.

Descriptions of participating sites. Descriptions of the sample have been developed with data collected from a number of sources: classroom observation instruments, categorical program applications, and evaluations on file in the State Department of Education. These data are summarized in Table 2. Descriptions of the data and comparisons between sample statistics and statewide values on the variables follow; summary statistics for the sample values are shown in Table 3.

Size of school. The size of schools in the sample ranged from 345 to 1,107 students. Despite this wide range, the sizes of the mean and standard deviation suggested that most of the sample schools were within a comparatively narrow enrollment range of 500 to 800 students; for schools with seven or nine grade levels, the size differentials of same-year cohorts (e.g., third grade) were significantly reduced. In fact, all but two schools were in a third grade enrollment range of 75 to 110 students. These data supported earlier subject assessments of substantial homogeneity in the sample across size of site, type of school, and size of third grade cohort variables. This conclusion held for all but two or three large inner-city schools.

Percentage of low-income families. Sample selection criteria ensured that the attendance areas chosen would contain substantial percents of low-income families. (See tables 2 and 3.) The 14 schools were distributed relatively evenly over the sample, which ranged from 17 percent to 68 percent low income. If anything was surprising in these numbers, it was that the mean percentage of low-income families for the low-achieving group was less than 40 percent.

Number of primary languages other than English. This measure was provided to complement the percent of LES/NES students and the percent of bilingual indexes, because two populations of students with roughly the same percentages of LES/NES students pose very different instructional problems if their subpopulations vary in the number of languages they speak. For example, it would seem that one school in which students speak seven primary languages, in addition to English, presents a challenge different from that faced at a school in which the only primary languages are English and Spanish. This difference appears despite the fact that the first school's percentage of LES/NES children is lower than that at the second.

Table 2

Descriptors of Schools Included in SRI Study

School by coded number	Number of students (1)	Percent, low income (2)	Percent, LES/NES (3)	Mean Entry Level Test score (4)	Socio-economic status index (5)	Percent, AFDC children (6)	Percent, bilingual (7)	Mobility index (8)	Number of primary languages other than English (9)	Length of principal's tenure (years) (10)
101* Value (Rank)+	714	57	2	24.97 (17)	1.84 (25)	25.0 (83)	39.9 (86)	33.2 (43)	4	10+
201* Value (Rank)	727	44	2	26.27 (28)	1.96 (34)	48.3 (96)	4.9 (26)	37.5 (60)	1	3
301 Value (Rank)	661	22	28	19.32 (1)	1.86 (25)	17.4 (70)	61.7 (95)	43.8 (79)	5	1
302 Value (Rank)	918	17	25	23.69 (10)	1.95 (33)	15.7 (66)	59.7 (94)	40.5 (70)	9	10+
401* Value (Rank)	755	68	17	20.38 (2)	1.14 (1)	75.4 (99)	25.4 (75)	52.6 (93)	9	3
501 Value (Rank)	1107	63	47	19.01 (1)	1.43 (5)	18.0 (71)	94.6 (99)	39.6 (67)	4	4
601 Value (Rank)	700	25	2	23.92 (11)	1.39 (4)	42.7 (54)	23.8 (73)	45.9 (84)	5	4
602 Value (Rank)	869	58	0	24.30 (13)	1.41 (5)	46.8‡ (96)	2.5 (15)	29.2 (28)	2	5

Descriptors of Schools Included in SRI Study

School by coded number	Number of students (1)	Percent, low income (2)	Percent, LES/NES (3)	Mean Entry Level Test score (4)	Socio-economic status index (5)	Percent, AFDC children (6)	Percent, bilingual (7)	Mobility index (8)	Number of primary languages other than English (9)	Length of principal's tenure (years) (10)
603* Value (Rank)	672	48	6	24.05 (12)	1.52 (8)	46.8 (96)	9.0 (44)	34.3 (47)	1	10+
701 Value (Rank)	519	21	35	20.94 (3)	1.35 (3)	13.5 (60)	57.8 (94)	34.8 (50)	1	5
801* Value (Rank)	667	30	0	23.48 (9)	1.41 (11)	18.1 (72)	64.3 (95)	29.4 (29)	1	4
901* Value (Rank)	483	36	15	25.75 (23)	1.50 (7)	25.7 (84)	56.3 (93)	42.0 (75)	3	2
1001* Value (Rank)	359	18	29	27.33 (41)	2.41 (73)	17.6 (71)	34.6 (83)	18.5 (4)	7	4
1101 Value (Rank)	345	38	5	24.61 (15)	1.52 (8)	47.9 (96)	16.8 (64)	40.0 (69)	2	1

12

15

*This school had third grade reading scores that had increased since its entry into the ECE program.
+The ranks given (in parentheses) are state percentile ranks.
‡These are 1975-76 data.

Sources of data: columns 1 and 10, California Public School Directory; columns 2 and 3, Consolidated Applications; columns 4 through 8, California Assessment Program; and column 9, Consolidated Evaluation.

The range of languages over the sample was from one to nine; the relatively high mean and standard deviation strongly suggested that one possible source of the low achievement scores of children at these schools was a language barrier between them and the dominant English-speaking culture.

Number of years of principal's tenure. Since turnover of key staff was a problem reported by some schools with low-income students, SRI decided to examine the extent to which these schools experienced a change of principals. The results were encouraging; though no statewide comparison data were available for this report, sample schools appeared almost uniformly to attract individuals who were willing to stay and lend a measure of stability to an otherwise fluid situation. (See mobility index in tables 2 and 3.) This finding, coupled with SRI's overall subjective assessment of the principals as "conscientious, competent educators," provided one note of encouragement to the future of these schools.

Percent of families receiving Aid to Families with Dependent Children (AFDC). Values on this variable ranged from a low of 13.5 percent to a high of 75.4 percent, with the schools clustering at two different levels: eight schools had 13--25 percent AFDC, while five others had between 40 percent and 50 percent. (See Table 2.)

Table 3

Sample Values on School-Level Variables

Variable	$\bar{X}$	Standard deviation	Range	Median
School size	678.2	208.4	345--1107	686.0
Percent, low income	38.9%	17.6%	17--68%	37.0%
Percent, limited- and non-English-speaking students	15.2%	15.3%	0--47%	10.5%
Mean Entry Level Test score*	23.4	2.57	19.0--27.3	23.87
Socioeconomic status index	1.63	.33	1.14--2.41	1.52
Percent with Aid to Families with Dependent Children (AFDC)	32.8%	18.5%	13.5--75.4%	25.4%
Percent, bilingual*	39.4%	27.2%	2.5--94.6%	37.3%
Mobility index*	37.2	8.4	18.5--52.6	38.6
Number of primary languages other than English	3.9	2.9	1--9	3.5
Number of years of principal's tenure	---	---	1--10+	4

*Index developed by California Assessment Program, State Department of Education. For a description, see Interpretive Supplement to the Report on the Reading Test, Second and Third Grades, 1977, Sacramento: State Department of Education, 1977.

The interpretation of the percent of AFDC families was difficult, however, because of the established relationship between racial/ethnic status and the propensity of this group to apply for welfare. Thus, the measure was not a good proxy for the socioeconomic status index (SES), and it was not strongly associated with seemingly comparable SES measures.

Percent of schools with bilingual populations. This variable was used as one of the stratifying factors in the sample selection and, thus, was a dimension of more than average interest. Both the range (2.5 to 94.6 percent) and the standard deviation (27.2) were quite large on this measure, suggesting wide variability among sites. Closer inspection, however, revealed that three schools had comparatively small numbers of students who were bilingual, while the remaining 11 schools had large bilingual populations that placed those schools well beyond the state average. Thus, much of the variability appeared to be within a range where a subpopulation exerted a significant influence on the environment of a school and in which differences of even 10 percentage points in the student body composition marked differences of degree, not differences of kind

Mobility index. Student mobility is a long-standing complaint of teachers at low-income schools, who often see a seat in their classroom filled by three or four different students in the course of a year. This problem was compounded for SRI's sample schools by two characteristics of the state of California: its overall high transiency rate (which is the product of complex social factors) and the agricultural orientation of the state economy.[2] The data confirmed the seriousness of this problem: eight of the 14 schools were well above the median state school mobility index value, and three others were very close to the state median. One surprise, however, was the finding that three participating schools had extremely stable populations; in one case the school had a very low SES ranking. However, the overall picture confirmed observations in previous studies that schools with low-income/low-achieving student bodies tend to have greater student mobility than other categories of schools.

Percent of limited- and non-English-speaking students (LES/NES). At one site no students qualified as limited- or non-English speaking (LES/NES), and at four others 6 percent or less of the student body was LES or NES. On the other hand, several sites had substantial LES/NES populations, including one where nearly one of every two children was in that category. For this reason, the distributional statistics must be interpreted with caution. (For example, the standard deviation is almost as large as the sample mean.) The sample can more appropriately be divided into two groups: (1) one in which the proportion of LES/NES students was considerably greater than in the general population and where these students had a major impact on the cultural and educational environment of the school; and (2) a second group of schools in which the students were only a small percent of the population.

Mean Entry Level Test (ELT). The Entry Level Test is given to all first grade children in the state. For the most part, the relative position of the entry level for first graders in the SRI sample of schools did not change substantially in 1976-77 from the level of entry scores reported for first graders

[2]Three schools were in agricultural areas, and others received children of migrant farm workers during the period when farm work was scarce.

in 1973-74. Three schools, however, had moved well above the lowest 20 percent of schools in the state. One school's mean score changed so much that the 1973-74 score had to be considered either an anomaly or a change in the composition of the student body. The range across the sample was not great, especially when the outlier school was removed from the calculation of descriptive statistics ($\bar{X}$ = 23.13; S.D. = 2.41).

Socioeconomic status (SES) index. School values on the SES index ranged from a low of 1.14 (1st percentile) to 2.41 (23rd percentile). Again, standard distributional statistics are misleading here because of the wide range and nonsymmetrical distribution of the variable values. There was again an outlier that was clearly not typical of the rest of the schools. The remaining schools tended to group themselves into two categories: the nine schools at or below the 10th percentile in the state and the four that ranged from the 25th to 33rd percentiles. Thus, there were significant differences from school to school on this measure; and no single statistic captured this diversity. It should be noted that the degree of this diversity is difficult to explain in the face of the relative homogeneity across sample schools on achievement measures and the percentages of children from low-income families who were discussed earlier.

RELATIONSHIPS AMONG CLASSROOM INSTRUCTIONAL PROCESSES

Several analyses were carried out to study in detail the instructional processes being used in Stanford Research Institute (SRI) International's sample of classrooms. First, SRI compared schools categorized as "increasers" with those categorized as "decreasers." Next, SRI looked at the correlations of process variables and reading scores. Finding that classrooms differed significantly in their pretest scores and gain scores, the SRI group compared subgroups within the sample by using a discriminant function analysis. Finally, SRI looked at correlations of absence rate and instructional process variables.

Comparison of Schools

A finding in the Evaluation Report of ECE, ESEA, Title I, and EDY, 1975-76, indicated that in schools whose students scored at or below the 20th percentile on the Entry Level Test in 1973-74, third grade reading test scores had declined, relative to their predicted scores over the three-year period from 1973-74 to 1975-76.[3] Although the aggregate finding for that group of schools indicated a decline in scores relative to prediction, individual schools could be identified which had patterns of scores that increased over the period in question, as well as schools in which scores declined. A question of interest became, "What circumstances are associated with decline or improvement in third grade student reading achievement?"

---

[3] Evaluation Report of ECE, ESEA, Title I, and EDY, 1975-76. Sacramento: California State Department of Education, 1977.

Table 4

Analysis of Variance for Seven "Increaser" and Seven "Decreaser" Schools in SRI Study

Variable number	Description	Increasers Group I (n = 20)		Decreasers Group II (n = 25)		F	P
		X	S.D.	X	S.D.		
Number absent		11.17	5.18	10.40	2.69		
Number enrolled		29.41	11.57	29.62	2.65		
Pretest		14.27	3.53	11.32	3.65	6.196	.017
Post-test		18.20	3.25	14.48	3.87	10.148	.003
17	Total class duration	5.44	0.36	5.07	0.58	6.343	.015
19	Stationary desks in rows	0.55	0.51	0.29	0.46	3.086	.083
20	Assigned seating for at least part of the day	1.00	0.00	0.50	0.51	19.091	.000
24	Child selection of seating	0.45	0.60	0.92	0.93	3.726	.057
26	Games, toys, play equipment used	1.25	0.64	1.96	1.52	3.787	.056
38	Audiovisual equipment used	0.85	0.81	1.75	1.07	9.504	.004
39	General equipment, materials present	8.35	0.88	9.21	2.21	2.668	.106
40	General equipment, materials used	4.45	2.48	2.46	1.86	9.230	.004
44	Total number different resource categories coded "present" over three days	16.75	2.31	18.75	4.79	2.908	.092
50	Cooking and sewing supplies present	0.05	0.22	0.29	0.46	4.529	.037
51	Cooking and sewing supplies used	0.05	0.22	0.25	0.44	3.363	.070
66	Numbers, math, arithmetic	14.32	7.25	15.48	7.64	0.261	.618
67	Reading, alphabet, language development	31.86	26.88	32.49	12.87	0.010	.88
70	Drill and practice	0.11	0.47	1.76	3.81	3.709	.058
73	Active play/music and dance	4.47	5.11	1.52	2.12	6.633	.013
79	Transitional activities	6.06	5.32	8.29	3.61	2.712	.103
85	Teacher without children	10.86	10.84	3.39	5.18	8.980	.005
90	Overall teacher occurrences	27.45	7.92	35.46	13.05	5.754	.020
91	Aide without children	14.16	14.24	7.25	10.46	3.432	.068
95	Aide with large group	6.93	11.95	25.47	23.34	10.334	.003
96	Overall aide occurrences	13.30	9.07	19.50	13.03	3.220	.076
100	Volunteer with small group	2.39	6.10	11.56	23.61	2.853	.095
103	Adult without children	16.24	11.11	7.28	7.09	10.495	.002
107	Adult with large group	33.74	16.79	42.95	19.14	2.819	.097
108	Overall occurrences of adults	45.80	15.46	58.75	19.21	5.899	.019
117	Large group of children independent	22.00	13.13	15.23	10.39	3.647	.060
125	Two children with aide--math	0.00	0.00	0.97	2.06	4.466	.038
133	Two children with any adults--math	0.43	1.08	1.50	2.39	3.367	.070

Table 4--Continued

Analysis of Variance for Seven "Increaser" and Seven "Decreaser" Schools in SRI Study

Variable number	Description	Increasers Group I		Decreasers Group II		F	P
		X	S.D.	X	S.D.		
139	Large group of children	26.94	29.20	13.87	19.31	3.16	.079
140	Approximate number of children involved in math for all days observed	98.24	55.89	136.46	73.60	3.638	.060
143	One child with teacher--reading	0.08	2.40	1.57	1.80	5.525	.022
155	One child with any adults--reading	1.05	2.87	2.51	2.78	2.918	.091
159	One child independent--reading	-0.17	7.15	2.88	2.72	3.730	.057
160	Two children independent--reading	0.02	9.77	4.58	5.24	3.887	.053
169	Large group with teacher--social studies	33.82	44.51	14.20	33.56	2.776	.099
185	Large group of children independent--social studies	5.25	14.09	0.38	1.87	2.816	.097
194	Large group with aide--science	0.00	0.00	5.24	14.44	2.623	.109
206	Large group of children independent--science	11.25	30.86	0.00	0.00	3.21	.077
215	Large group of children with aide--arts and crafts	0.00	0.00	3.54	8.62	3.355	.071
225	Two children independent--arts and crafts	1.69	6.44	11.63	24.05	3.213	.077
227	Large group of children independent--arts and crafts	24.78	40.05	3.75	11.06	6.089	.017
252	Sewing, cooking, pounding--longitudinal	0.00	0.01	0.03	0.04	5.841	.019
254	Dramatic play, dress-up	0.01	0.02	0.00	0.00	3.204	.077
262	Average number of adults i the classroom--longitudinal	1.19	0.32	1.38	0.37	3.094	.082
271	Teacher observing	17.65	14.77	4.93	6.40	14.572	.001
274	Aide Participating	0.99	3.38	6.83	7.77	9.726	.003
276	Aide not involved	1.21	2.97	6.39	8.87	6.20	.016
287	Number of FMOs with sewing, cooking or pounding as beginning activity	0.05	0.22	0.58	1.32	3.196	.078
292	Number of FMOs with social interaction as beginning activity	2.15	2.23	0.79	1.64	5.401	.024
296	Number of FMOs with transitional activities as beginning activity	2.00	2.10	2.92	1.41	2.965	.089
355	Adult open-ended questions to children, nonacademic	0.08	0.13	0.00	0.01	8.539	.006
358	All child responses	9.07	1.98	10.09	2.09	2.701	.104

Table 4-- Concluded

Analysis of Variance for Seven "Increaser" and Seven "Decreaser" Schools in SRI Study

Variable number	Description	Increasers Group I		Decreasers Group II		F	P
		X	S.D.	X	S.D.		
368	Child responses to adult open-ended questions	0.08	0.11	0.02	0.05	5.260	.026
371	Child extended response to adult open-ended question	0.03	0.07	0.00	0.01	3.149	.080
387	Child general comments to adults	0.57	0.51	0.19	0.24	10.283	.003
389	Adult general comments to children	1.06	0.91	0.51	0.53	6.164	.016
393	Child corrective feedback	0.00	0.01	0.01	0.02	3.845	.054
394	All adult acknowledgment to children	2.80	1.22	2.02	1.30	4.127	.046
397	Adult acknowledgment, other task-related	0.84	0.47	0.49	0.31	9.032	.004
405	All adult corrective feedback to children	2.58	1.47	3.88	1.82	6.546	.014
406	Adult positive corrective feedback, academic	1.01	0.69	2.13	1.41	10.406	.003
417	Children attentive to adults, academic	0.00	0.00	0.02	0.05	5.720	.020
418	Adults attentive to children, nonacademic	2.42	1.77	1.07	1.02	9.945	.003
424	Positive behavior, children to adults	0.11	0.24	0.32	0.51	2.953	.089
425	Child expressions of unhappiness	0.44	0.34	1.15	1.06	8.464	.006
426	Adult expressions of unhappiness	1.82	1.08	3.92	5.93	9.278	.004
430	Total adult effect	3.44	1.60	7.21	5.74	8.052	.007
431	Total child effect	1.01	0.59	1.64	1.16	4.896	.031
441	Adult communication or attention focus, large group	4.04	3.66	1.53	2.19	7.857	.007
442	All children nonverbal	1.88	0.60	2.64	1.15	7.060	.011
452	Adult open-ended questions to children (Var. 355a + 356a and/or Var. 486a + 487a)	0.10	0.13	0.04	0.09	4.100	.047
457	All adult positive correcting feedback (Add Vars. 406a, 408a, 410a and/or 537c, 539c, 541c.)	1.19	0.88	2.33	1.46	9.516	.004
460	All children positive affect (Var.422a + 424a and/or Var. 553c + 555c)	0.11	0.24	0.32	0.51	2.953	.089
464	Child attentive (Add Vars. 416a, 417a, 446a and/or Vars. 547c, 548c, 577c.)	0.00	0.01	0.03	0.07	3.276	.074
471	Adults attentive to large group (Subtract sum of Vars. 420a and 421 from sum of Vars. 418a and 419a and/or subtract sum of Vars. 551c and 552c from sum of Vars. 549c and 550c.)	1.39	1.27	0.55	0.75	6.626	.013

Process variable differences. To study the question cited above, SRI first compared the 20 classrooms within the schools identified as "increaser" schools (schools whose third grade reading scores had increased) and the 25 classrooms in the schools identified as "decreaser" schools (schools whose third grade reading scores had decreased). One-way analyses of variance were computed to compare the frequency of process variables used within one set of classrooms with the frequency of those processes used within the other set of classrooms. Statistical differences were found at the $p < .10$ level in 76 out of 390 variables. (See Table 4.)

The "increaser" schools had longer school days, and the classrooms were a little more structured: the classrooms had more stationary desks in rows and assigned seating for students (variables 19 and 20); they used toys, audiovisual materials, cooking and sewing equipment, or dramatic play less often than the "decreaser" schools did (variables 26, 38, 39, 44, 51, and 254).

Math and reading occurred with about the same frequency in both groups (variables 66 and 67). Within those academic activities, the "increaser" schools had teachers who worked less frequently with only one or two students at a time in math or reading than teachers in the "decreaser" schools did (variables 125, 133, 143, 155). Adults in classrooms of "increaser" schools also worked less frequently with large groups of children and worked more often with small groups than adults in "decreaser" schools did (variables 108, 194, 215, 267, 168).

In the "increaser" schools, teachers were observed more often to be not involved with children, grading papers and the like than teachers in the "decreaser" schools were (variables 85, 91, 103), and children in the "increaser" schools were observed to be working independently in reading, math, science, social studies, and art (variables 139, 155, 159, 160, 189, 106, 227). Teachers in "increaser" schools were observed more often in social conversations (variables 196, 355, 387, 389, 417) and less often in instructional activities (variables 296) than teachers in "decreaser" schools were.

More open-ended questions were asked in the "increaser" schools. The children were responsive and more often gave elaborated or extended responses (variables 355, 368, 371, 452) than children in "decreaser" schools did.

The adults in classrooms of "increaser" schools provided more acknowledgment to children but less corrective feedback (variables 294, 397, 405, 406, 457) than adults in "decreaser" schools did.

The children in classrooms of "increaser" schools were a little less attentive (variable 464). They displayed less unhappiness, less positive affect (variable 406), and less affect in general (variable 431) than their counterparts in the "decreaser" schools did.

The findings from the comparison of "increaser" and "decreaser" schools were mixed. The variables found to be associated with success in reading in other studies did not discriminate between the "increaser" and "decreaser" schools. The consistent pattern of the direct teaching techniques was not found in either of these two groups of schools.

Two problems with this analysis should be mentioned. First, in spite of great effort to select a matched sample of schools, some demographic differences in the two groups were identified. (See Table 5.) "Increaser" schools had a higher percent of low-income children and more families receiving AFDC payments. "Decreaser" schools had more bilingual and limited- and non-English-speaking children, a higher mobility rate, and lower first grade entering school scores. SRI found it impossible to eliminate the effect of these differences with a sample so small when the direction of differences was inconsistent across the various SES measures.

Second, large and significant variance in the classroom performance of teachers and students could be expected within each school, as well as differences between the schools, in the "increaser" and "decreaser" samples. That is, assignment to the "increaser" and "decreaser" groups on the basis of school level information ignores the variance of classroom process and student achievement within schools. This was substantiated by the fact that standard deviations were sometimes greater than means for several variables.

The SRI observation study of ECE focused on classroom environments, processes, and interactions. Data were collected in classrooms, and the unit of analysis was the individual classroom. Since student achievement measured by the California Assessment Program's Reading Test was the dependent variable for much of the analysis, mean scores by classroom were used in this study. SRI found considerable variation in the instructional processes used within the "increaser" and "decreaser" subgroups of schools. SRI personnel were also interested in seeing whether there were differences in the pretest, post-test, and gain scores between these two groups and if there were differences in these variables within the schools themselves.

Test score differences. The classroom means and gain scores for each class are portrayed in tables 6 and 7. Scores for schools labeled "increaser" schools by the Department's Office of Program Evaluation and Research are presented in Table 6, and those for "decreaser" schools are shown in Table 7.

Before examining the results in detail, SRI raised the following questions to structure the investigation:

1. Do classrooms within individual schools differ significantly from each other and from the school mean, or are these measures approximately the same?

2. Considering the school and classroom status and gain scores, how do the students in these schools compare with other California third grade reading students?

The first question concerns intraschool variability and, implicitly, the appropriateness of using average school statistics to describe heterogeneous classroom populations. Using the data in Table 6, SRI compared the school means for spring, 1977, with the individual class means to determine whether significant differences existed. The fourth column in Table 6 presents the post-test reading scores (spring, 1977) for all third graders in each sample classroom as well as the overall mean for all third graders at each school.

Table 5

"Increaser" and "Decreaser" Subsample Values on School-Level Variables

Variable	Decreaser Schools				Increaser Schools			
	$\bar{X}$	S.D	Range	Median	$\bar{X}$	S.D.	Range	Median
School size	731.3	256.6	345--1107	700.0	625.3	147.3	359--755	672.0
Percent, low income	34.9	18.8	17--63	25.0	43.0	16.8	18--68	44.0
Percent, limited- and non-English speaking	20.3	18.2	0--47	25.0	10.1	10.6	0--29	6.0
Percent, educationally disadvantaged students	72.6	17.4	47--87	80.0	70.3	21.0	39--92	79.0
Mean Entry Level Test score*	22.26	2.43	19.01--24.61	23.69	24.60	2.28	20.38--27.33	24.97
Socioeconomic status index	1.56	0.24	1.35--1.95	1.42	1.71	0.41	1.14--2.41	1.61
Percent with Aid to Families with Dependent Children	28.9	16.0	13.5--47.9	18.0	36.7	21.2	17.6--75.4	25.7
Percent, bilingual*	45.3	32.1	2.5--94.6	57.8	33.5	22.3	4.9--64.3	34.6
Mobility index*	39.1	5.6	29.2--45.9	40.0	35.4	10.6	18.5--52.6	34.3
Number of primary languages other than English	4.0	2.7	1--9	4.0	3.7	3.2	1--9	3.0
Number of years of principal's tenure	-	-	1--10+	4.0	-	-	2--10+	4.0

*Index developed by California Assessment Program, State Department of Education. For description, see Interpretive Supplement to the Report on the Reading Test, Second and Third Grades. Sacramento: California State Department of Education, 1977.

Table 6

Pre, Post, and Gain Reading Scores of Classrooms in "Increaser" Schools

Increaser schools	Grade	Reading score, as a percent			Matched number of students	Sub-group
		Pre	Post	Gain		
School 101		61.6	71.9	10.3		
Teacher 01	3	61.9	80.6	18.7	21	5
Teacher 02	3	57.6	81.4	23.8	17	5
School 201		54.0	74.9	20.9		
Teacher 01	2-3	79.0	89.0	10.0	4	7
Teacher 02	2-3	58.0	77.0	19.0	8	5
Teacher 03	2-3	43.3	63.6	20.4	11	3
School 401		43.9	59.9	16.0		
Teacher 01	3	52.2	68.0	13.8	18	4
Teacher 02	3	36.0	52.9	16.9	9	2
Teacher 03	3	44.0	55.8	11.8	16	3
Teacher 04	3	36.0	48.6	12.6	13	2
School 603		54.4	70.9	16.5		
Teacher 01	3	53.8	71.1	17.2	13	4
Teacher 02	3	54.6	73.7	19.2	14	4
Teacher 03	3	56.0	75.8	19.8	18	4
School 801		47.6	65.0	17.4		
Teacher 01	3	50.8	67.4	16.6	14	4
Teacher 02	3	41.7	63.3	21.7	12	3
Teacher 03	3	50.1	74.1	24.0	17	4
School 901		63.3	64.3	1.0		
Teacher 01	3	63.0	65.7	2.7	12	6
Teacher 02	3	65.0	70.5	5.5	16	6
School 1001		75.4	90.4	15.0		
Teacher 01	1-3	73.2	88.6	15.4	7	7
Teacher 02	1-3	81.2	92.0	10.8	7	7
Teacher 03	1-3	84.0	97.3	13.3	6	7
State average		67.7	81.7	14.0		

n = 20 classrooms with pretest and post-test scores.

Table 7

Pre, Post, and Gain Reading Scores of Classrooms in "Decreaser" Schools

Decreaser schools	Grade	Reading score, as a percent			Number of students	Sub-group
		Pre	Post	Gain		
School 301		55.7	65.4	9.7		
Teacher 01	3	63.7	83.4	19.7	13	5
Teacher 02	3	56.0	57.7	1.7	14	6
Teacher 03	3	47.2	72.0	24.8	5	4
School 302		59.8	73.6	13.8		
Teacher 01	3-4	56.4	79.1	22.7	9	4
Teacher 02	3	59.5	74.4	14.9	15	5
Teacher 03	2-3	58.3	80.0	21.7	7	5
School 501		39.8	46.7	6.9		
Teacher 01	3	70.7	70.7	0	3	6
Teacher 02	3	--	--	--	-	
Teacher 03	3	--	--	--	-	
Teacher 04	3	16.6	34.8	18.3	7	1
Teacher 05	3	32.7	32.7	0	6	1
Teacher 06	3	--	--	--	-	
School 601		41.1	60.6	19.5		
Teacher 01	3	41.6	59.7	18.1	15	3
Teacher 02	3	34.8	57.4	22.6	14	2
Teacher 03	3	43.6	62.0	18.4	10	3
School 602		51.6	66.3	14.7		
Teacher 01	3	--	--	--	-	
Teacher 02	3	--	--	--	-	
Teacher 03	3	--	--	--	-	
Teacher 04	3	--	--	--	-	
School 701		46.3	57.7	11.4		
Teacher 01	K-3	28.0	36.0	8.0	2	1
Teacher 02	K-3	30.7	44.0	13.3	3	2
Teacher 03	K-3	54.7	57.3	2.7	3	6
Teacher 04	K-3	35.0	43.0	8.0	4	2
School 1101		51.6	66.9	15.3		
Teacher 01	2-3	41.2	61.7	20.6	7	3
Teacher 02	2-3	57.8	58.7	.9	9	6
State average		67.7	81.7	14.0		

In seven of the 14 schools, significant differences existed between class and school means. A significant difference is defined arbitrarily here as approximately 10 percentage points, the size of the standard deviation for the sample of schools in California. Thus, in one-half of the schools in the sample, differences of at least one standard deviation existed between achievement levels of students in different classes. Consider the example in Table 8.

Table 8

Comparison of Scores in Two Schools, by Classrooms

Classroom scores, by percent, in two schools

School X	Percent	School Y	Percent
Teacher 01	84.0	Teacher 01	59.7
Teacher 02	77.0	Teacher 02	57.4
Teacher 03	63.6	Teacher 03	62.0
Average	74.9	Average	60.6

The interclassroom variability in the two schools in Table 8 is clearly not equal; significant differences between classroom scores can be found in School X but not in School Y. Correspondingly, the value and meaning of the school-wide "average score" is different at the two schools. In School Y this statistic is comparatively representative of achievement in individual classrooms, but at School X it is more misleading than useful because it glosses over important differences between classrooms and accurately describes only one of the three classrooms in the school. The classroom is the level at which instruction takes place and at which teaching strategies, techniques, and environments impact upon the student. To aggregate up from classrooms to schools, especially in situations similar to that of School X, means that these classroom effects and levels of achievement are cancelled out.

The second question SRI addressed was how classrooms in its sample, selected on the basis of the schools' low ELT scores, compared with scores in other California third grades. The mean score for California third graders in the spring, 1977, was 81.7 percent correct; when those students were second graders a year earlier in the spring, 1976, the mean score was 67.7 percent correct. Thus, the "average" California third grader had to gain 14 percentage points to maintain the same position in the score distribution for that cohort group. How did student gains in SRI's sample compare? These data were presented in the fifth column in Table 7.

The results were striking. Of the 32 classrooms for which SRI had pre- and post-test matched scores for five or more students, 21 classrooms gained more than the state average of 14 percentage points.[4] Eight of these classrooms were in

[4]School 701 had kindergarten through grade three in one classroom. Only three or four third grade students had pre- and post-test scores. These classrooms and one in school 501 were not used in the analysis.

schools classified as "decreaser" schools. Again, the contrast between school and class scores was illuminating: In several schools significant differences existed between the school gain and individual class gains. A dramatic example is presented below:

School X	
Teacher 01	19.7%
Teacher 02	1.7
Teacher 03	24.8
Total school	9.7%

SRI noted that not all third grade teachers at every sample school participated in the study; the absence of one or two classrooms helps explain, for example, why the school means in tables 6 and 7 cannot be computed from the scores of the classrooms presented. A second caveat was that SRI's classroom means were based upon a matched sample of children, whereas the school means represented entire cohorts of children that changed from year to year. Thus, SRI's classroom means were based on a subsample of children that may or may not have been representative of the students at that school. The importance of this possibility was mitigated, however, by the following two considerations:

1. None of the sample schools reported substantial demographic changes (such as a desegregation order) that would have affected their California Assessment Program scores between the spring, 1976, and the spring, 1977. This would suggest that students who left the classroom were replaced by students who were of similar background and achievement status.

2. Examination of the characteristics of students in the matched sample of students indicated that they did not differ systematically from the overall school population on either the state socioeconomic status index or percent bilingual index.

These considerations do not then change the conclusion that significant differences existed between the rates of gain registered by classes within schools and that the students in 21 classrooms identified as low-achieving gained more than the average California student during the third grade year.

Instructional Processes Related to Reading Achievement

To study the instructional processes related to reading, SRI initially planned to perform only partial correlation and stepwise regression analyses for the ECE sample, as was done in the Follow Through Observation Evaluation of 1972-73 (Stallings and Kaskowitz, 1974). However, the data collected for the ECE study indicated that pretest scores varied widely among classrooms; classroom averages in the spring, 1976, ranged from an average 30.7 percent correct to 84 percent correct. With such variations in the groups of students teachers received in the fall of 1976, teachers could be expected to use different instructional processes to teach them. Instructing a class of students who entered the third grade with very low reading abilities is likely to require different teaching techniques from instructing students who enter third grade with average or above average

Table 9

Partial Correlations* and Two-Tailed Pearson Product Moment Correlations Between Reading Scores and Classroom Processes (n = 38)

Variable number	Description	Pearson correlations		Partial correlations	
		r	p	r	p
15	Ratio of students high to T+A	-.39	.01	-.43	.001
16	Ratio of students high to T+A+V	-.38	.01	-.38	.01
18	Movable tables and chairs	.38	.01	.11	.50
22	Teacher assigns students to groups	.45	.001	.46	.001
24	Children select work groups	-.32	.05	-.24	.15
25	Games, play equipment present	.30	.05	-.43	.001
28	Instructional materials used	.32	.05	.25	.15
36	Instructional charts used	.29	.05	.10	.50
40	General equipment used	.36	.05	.38	.01
45	All resource material categories used	.36	.05	.34	.05
49	Achievement charts used	.44	.01	.27	.10
54	Sandbox, water equipment	-.32	.05	-.19	.25
62	Group sharing-time	-.55	.001	-.42	.001
67	Average occurrence of reading per child	.30	.05	.31	.05
79	Transition activities	-.29	.05	-.13	.40
81	Teacher out of the room	-.39	.01	-.36	.05
88	Teacher with small group	.34	.05	.32	.05
89	Teacher with large group	-.31	.05	.15	.30
112	Large group with any adult	-.40	.01	.03	--
143	One child with teacher--reading	-.36	.05	-.38	.01
155	One child with any adult--reading	-.34	.05	-.37	.05
242	Percent activities on academic subjects	.56	.001	.41	.01
244	Percent of group time-sharing over total day	-.55	.001	-.42	.001
249	Percent of reading occurring over total day	.31	.05	.32	.05
271	Teacher observing the group	.33	.05	.22	--
272	Teacher not involved with children	-.37	.01	-.40	.001
293	Number of social interactions that begin observation	-.48	.01	-.43	.001
342	Adult to child--verbal	.41	.01	.32	.05
343	Child initiates verbal interaction	.41	.01	.41	.01
350	Child asks questions	.36	.01	.33	.05
352	Adult asks nonacademic questions	-.40	.01	-.41	.01
354	Adult asks academic questions of individual child	.32	.05	.30	.05
359	Child response to nonacademic question	-.36	.05	-.40	.01

*Holding spring, 1976, test scores constant

Table 9--Concluded

Variable number	Description	Pearson correlations		Partial correlations	
		r	p	r	p
362	Child response to academic question	.35	.05	.30	.05
367	Adult response to child question with direction question	.39	.01	.23	--
374	Adult instruction--academic	.34	.05	.30	.05
376	Adult instructs a small group	.32	.05	.30	.05
386	Child task-related comments	.27	.05	-.04	--
394	All adult acknowledgment to students	.37	.05	.46	.001
395	Adult acknowledgment for academic	.35	.05	.49	.001
412	Feedback, to response, to academic question	.34	.05	.42	.001
420	Adult works with small group	.41	.01	.29	.05
429	Child negative behavior	-.32	.05	-.25	--
435	Total academic interactions	.35	.05	.43	.007
440	Adult communication to small groups	.43	.001	.41	.01
451	All adult direct academic questions	.33	.05	.29	.05
453	All adult responses to child question with a question	.39	.01	.23	--
455	All adult instruction	.35	.05	.25	--
461	All child negative affect	-.33	.05	-.24	--

abilities. Partial correlations or stepwise regressions using the total sample of 38 classrooms was likely to mask interesting process/outcome relationships in subgroups of classrooms. Therefore, SRI developed an analysis plan that included the following steps:

1. Two-tailed Pearson product moment correlations between classroom process variables and spring, 1977, reading test scores

2. Partial correlations between classroom process variables and reading achievement, holding spring, 1976, scores constant

3. Stepwise regression of spring, 1977, reading scores, entering spring, 1976, reading scores first, then classroom process variables

4. Scatterplot of classes based on pre- versus post-test scores to determine achievement subgroups within the total sample

5. Discriminant analysis to verify the utility of the achievement subgroups on the basis of differences in classroom processes

6. One-way analyses of variance and orthogonal contrasts on selected input variables to further examine intergroup differences

7. Summary group descriptions

Pearson product moment correlations and partial correlations. SRI's first step in examining the relationships between classroom processes in ECE classrooms was to compute two-tailed Pearson product moment correlations between the post-test and the process variables and to compute partial correlations, holding spring, 1976, reading test scores constant. While partial correlations were the most important correlation to consider when searching for relationships between student growth and instructional processes, Pearson correlations relate a wider set of variables that describe how instructional processes are related to student outcome, regardless of the student's Entry Level Test scores. SRI thought it would be of interest to examine both sets of correlations.

Since SRI used the same observation system that was used in its follow through observation study, SRI personnel used selected process variables constructed for that study in their ECE analysis. A total of 253 activity and interaction variables were used. Out of those, 49 Pearson product moment correlations were significantly correlated with outcome achievement scores at $p < .05$. The 49 variables identified in the Pearson product moment analysis were then used in the partial correlation analysis. Of these, 31 variables were significantly correlated ($p < .05$) with achievement. (See Table 9.) Because of the chance of collinearity of variables, these correlations had to be interpreted with caution; however, if these findings are similar to the findings of other studies, they become more believable. The findings from the analysis follow:

1. Organization of groups--In the ECE classrooms, the lower the ratio of students to adults, the higher were the achievement test scores (variables 15, 16; see Table 9). The correlation was even higher for the partial

correlations ($r = -.43$).[5] The ECE aides apparently were functioning well as teaching assistants.

How teachers organized their classrooms, used the aides, and grouped children had a relationship to reading achievement. These data suggest that teachers should assign students to working groups and should not allow them to select their own groups or places to sit during working time (variable 22).

2. Small groups--The most efficient way in ECE to distribute the adults' time seemed to be working with small groups (variables 88, 376, 420, 440). Approximately 29 percent of the reading time in all classrooms was allocated to teaching small groups. The scores were higher in classrooms where more time was spent with small groups.

3. Individual child--In classrooms where teachers, aides, or volunteers worked more often with one child at a time, than with more than one child, the post-test scores were lower (variables 143, 155). This does not imply that teachers or aides should not work with one child at a time or that individual tutoring is not beneficial. It does imply that within the normally functioning classroom, it is not efficient for the instructors to allocate very much of the reading time to working with only one student at a time. While the mean for ECE adults working with one child was low (approximately 2 percent of their reading time) teachers in some classrooms were spending 6 to 8 percent of their time with individuals. What is likely to happen when the adults spend more time with individuals is that other children will not have the necessary direction or supervision to stay on their reading tasks.

4. Large group--In the Pearson correlations (variable 89), a negative relationship was found between reading post-test scores and teachers teaching large groups, but the negative relationship was not found in the partial correlations. This suggests that when the effect of initial scores is eliminated, large group instruction is neither positive nor negative in its relationship to achievement. ECE teachers were, on the average, allocating approximately 27 percent of the reading time to instructing the large group. Large groups were defined as eight or more children up to the total group minus 1.

5. Teacher time without children--Some of the time (26 percent), teachers were not involved with the children. They graded papers, prepared lessons, talked with other adults, prepared art materials, or were out of the room. Where this happened often, the scores in those classrooms were lower (variables 81, 272, 293) than in those classrooms where it did not occur often.

---

[5] The follow through findings were just the opposite; the more adults in the classroom and the lower the ratio of students, the lower were the test scores. However, when the ratio of students to adults was as high as 18:1, as it was in the follow through study sample, the effect of the part-time aide may not be sufficient to raise test scores.

6. Activities occurring--The more group sharing time that occurred, the lower the post-test scores were (variable 62). This does not imply that there should not be any group sharing time. The average amount of time spent in group sharing was approximately 20 minutes a day (6 percent of 5.5 hours), but some classrooms spent twice that much time while others spent much less. The caution here is not to spend so much time in group sharing or discussions that too little time is spent in reading. The more reading that occurred, or the more time that was spent on that task, the higher the test scores were. This is indicated by variables that describe the amount of time students spent reading or doing reading-related activities (variables 67, 242, 249, 435).

7. Materials present and used--Several materials or resources had positive relationships with the post-test scores. Significant Pearson correlations include instructional charts, achievement charts, and games (variables 25, 28, 36, 49). When the effect of pretest scores is removed, the games variable has a negative relationship to gain (variable 25). This suggests that in classrooms with more able students, games have a positive effect on scores; while in classrooms with less able students, such games have a negative effect. Only general equipment and instructional materials are significantly related to gain in the partial correlations. Audiovisual equipment or television did not show a significant positive or negative relationship with the test scores in either set of correlations.

8. Interactions--A very directive method of instruction had a positive correlation with the post-test score. In this method, the teacher provided instruction and asked the children direct questions about the subject. The children responded; and the teacher provided feedback, letting the child know whether the response was acceptable. Variables 354, 362, 374, 395, and 412 describe this kind of interaction. Acknowledgment and corrective feedback for children's responses to academic questions had especially high correlation with reading achievement (.49 and .42).

9. Children questioning--When children in classrooms showed more verbal initiative by asking questions, the classroom scores were higher (variables 343, 350). If the adults responded with a direct question rather than giving the answer when children asked academic questions, the reading scores were higher (variable 367) on the Pearson correlations.

10. Distraction--Off-task behavior had a negative effect on the post-test scores. When adults and children were more often involved in nonacademic interactions (variables 352, 359, 293) than in academic interactions, the reading test scores were lower. Also, the Pearson correlations suggest that in classrooms where children exhibited negative or disruptive behavior, the test scores were lower (variables 429, 461).

Stepwise regressions. Since so many variables were significantly correlated with reading scores, stepwise regressions were computed to determine which variables accounted for the most variance in predicting reading achievement scores.

The pretest score accounted for 79 percent of the post-test score variance. Three process variables explained another 10 percent of the variance. Two of the variables suggested a structured situation where the teacher assigned students to groups and the verbal interactions were academically oriented. The variable

"Teacher working with one child in reading" had a negative relationship. Where that occurred with greater frequency, it was related to lower test scores and explained a small portion of the variance. (See Table 10.)

Table 10

Stepwise Regression of Post-Test Reading Scores, Holding Pretest Scores Constant

Variable number	Description	R^2	R^2 Change	b
	Pretest	.79	.79	.79
22	Teacher assigns students to group	.85	.06	.13
435	Total academic interactions	.87	.02	.16
143	One child with teacher in reading	.89	.02	-.14

Since the pretest accounted for so much of the post-test score, SRI wanted to see what variables were related to the post-test score if the pretest was not entered in the equation. By rerunning the stepwise regressions without the pretest, SRI tried to determine (1) the change in the predictive power (R^2) of the new equation; and (2) those variables related to post-test via the pretest (i.e., the pretest as a moderator variable; see Table 11).

Without the pretest, the process variables explain 62 percent of the variance. The variable that describes the percent of time children were observed to be involved in the academic activities of reading or math explains 31 percent of the test score variance when the pretest is excluded from the analysis.

Interactions that are social in content explain 13 percent of the test score variance. This variable indicates that the teacher and students were off the task. Where this occurred more frequently, the scores were lower. Movable tables and chairs allowed for small group arrangements. This variable accounted for a 9 percent variance. The teacher observing the group was a variable that indicated the teacher was supervising the children while they worked independently.

Summary of correlations and regressions findings. These preceding correlational findings are similar to other findings regarding the direct approach to teaching reading and off-task behavior. These findings were similar to those of Stallings and Kaskowitz (1974), Brophy and Everston (1975), McDonald and Elias (1975), and Rosenshine (1977). Clearly, it was beneficial for teachers to keep children on their tasks, and this seemed to be best accomplished in environments where the teacher assigned children to seats and to groups, provided instructional materials, worked with the children, taught them in small groups, and used a question-response-feedback technique for instruction.

Table 11

Stepwise Regressions of Post-Test Reading Scores and Process Variables

Variable number	Description	R^2	R^2 Change	b
242	Percent of activities that are academic	.31	.31	.39
293	Number of observations started with social interaction	.44	.13	-.34
18	Movable tables and chairs	.53	.09	.38
271	Teacher observing students as they worked	.62	.09	.31

Subgroups Determined by Pretest, Post-Test, and Gain Scores

To examine further the relationships between classroom processes and achievement, SRI personnel wanted to see whether they could define the instructional processes most relevant to producing test score gains for particular groups of classrooms. The SRI group hypothesized that a teacher of a classroom with a post-test/pretest difference of 15 percent and an initial test level of 40 percent might differ in approach to teaching from a teacher whose class had an identical gain but started with a pretest level of 70 percent. Therefore, a scattergram plotting post-test versus pretest was created to attempt a clustering of classes on the basis of pretest level and amount of test score gain. (See accompanying Figure 1.)

The determination of achievement groups required several steps. First, a "no gain" line was defined (post-test--pretest = 0) where there was no difference between post-test and pretest score means. Second, to be consistent with other studies of this type, a "gain" line was set at one standard error above the no-gain line (Berlinen and Ticknoff, 1975). The no-gain group, then, became those classrooms with post-test/pretest differences falling between the standard error gain line and the no-gain line. Third, the classrooms were assigned to seven groups on the basis of pretest levels. The scattergram resulted in the groups of classrooms exhibited in Table 12.

A special note is required on the a priori determination of groups IV (medium pretest/gain) and V (high-medium pretest/gain). The 14 classrooms involved all exhibited similar gain, but the variance of the pretest scores was substantial. Therefore, in order to obtain the most appropriate comparison group for Group VI (high-medium pretest/no gain) the 14 classrooms were split into two pretest groups: Group V was statistically equivalent to the pretest level for Group VI, while Group IV had a lower average pretest. A comparison of groups V and VI indicated no statistical differences in the test scores of the two groups at the beginning of the third grade year:

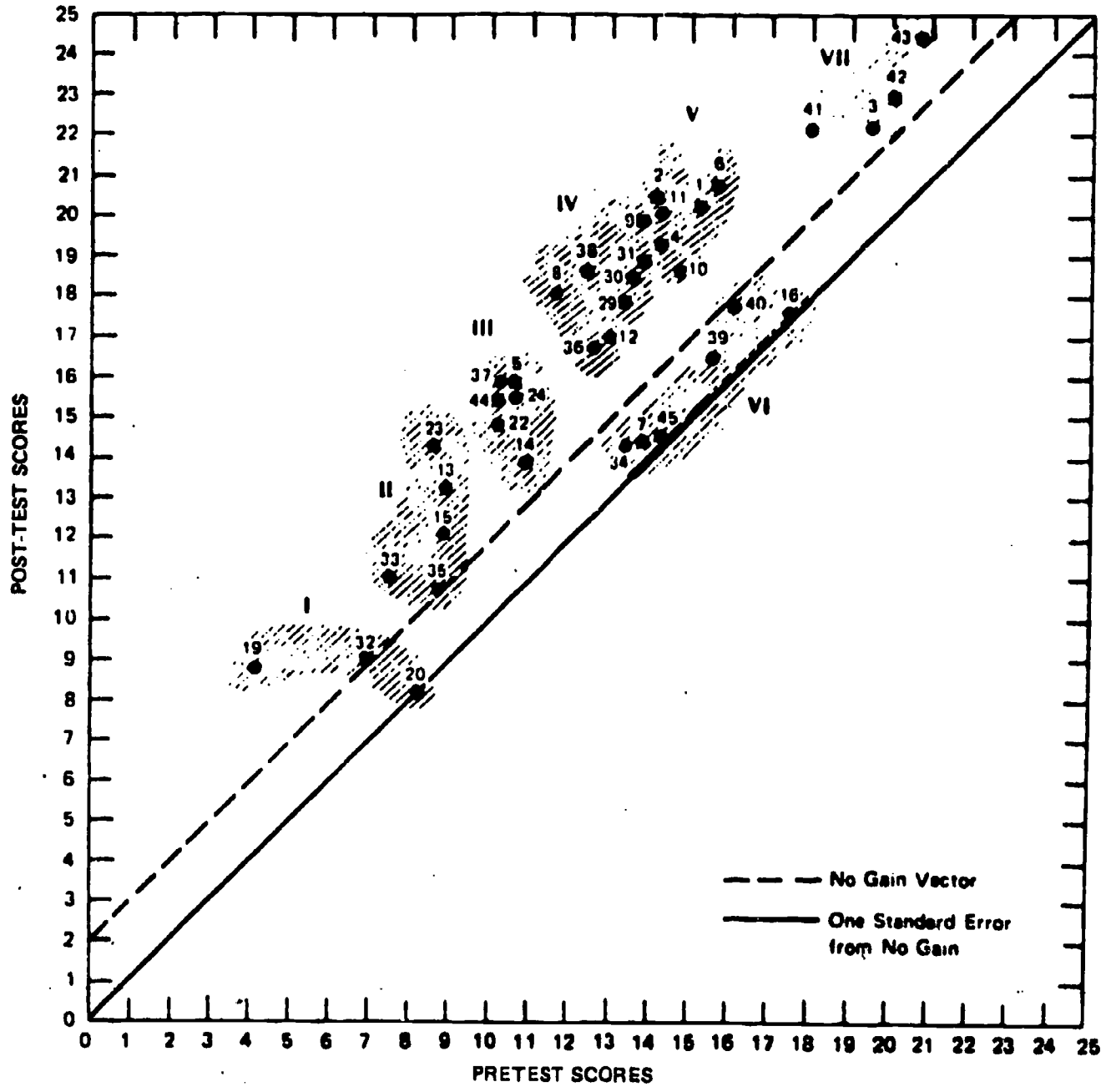

Fig. 1. Scattergram of ECE Third Grade Pretest and Post-Test Scores

Group V: $\bar{X}$ raw score = 15.04

Group VI: $\bar{X}$ raw score = 15.30

t = 5.60; p = .12 (no statistical difference)

In addition to the high-medium pretest gain/no gain comparison, SRI attempted to find an analogous situation at the lower pretest level. Groups I and II, then, were identified and an F test was conducted to examine the difference in pretest score level. Unfortunately, the average pretest level for the two groups was significantly different, so that a gain/no gain comparison controlling for pretest level was not possible:

Table 12

Number of Classes in Each Subgroup

Group	Number of classes	Pretest score*	Percent gain
I Low pretest/no gain	3	37%	8%
II Low pretest/gain	5	37%	12%
III Low-medium pretest/gain	6	41%	15%
IV Medium pretest/gain	8	47%	16%
V High-medium pretest/gain	6	57%	14%
VI High-medium pretest/no gain	6	54%	7%
VII High pretest/gain	4	73%	10%
Total	38		

*Percent of total correct.

Group I: $\bar{X}$ raw score = 6.44

Group II: $\bar{X}$ raw score = 8.63

F = 5.51; p = .06 (significantly different at the .10 level)

From the seven groups formed, only groups V and VI were selected for comparison. Groups V and VI both started with scores of approximately 60 percent correct, but only Group V gained. SRI thought it would be of interest to see how these groups differed on the process variables being used in the classrooms. In order to see whether these two sets of subgroups differed initially, several analyses were conducted.

Discriminant analysis. In order to ascertain whether the groups defined by achievement and prescore level differed in structure, three discriminant analyses were conducted. The objective was to find out whether a discriminant analysis

would confirm the legitimacy of the a priori groups. A stepwise discriminant procedure was selected, using an F-to-enter criterion of approximately .10 level of significance. Based upon prior research findings and the utility of the variables in describing classroom procedures, 100 variables were selected for the analyses by the principal researcher. The first analysis attempted to differentiate among all seven groups simultaneously. It was hypothesized that such a comparison would be inconclusive, given the arbitrary separation of groups IV and V and the relatively close pretest levels of groups I, II, and III. The discriminant analysis failed to derive a set of functions that would correctly categorize the seven groups (51.35 percent of the classes were assigned to the correct a priori group, a low success ratio).

Of more interest, then, was the comparison of the appropriate gain/no gain groups from the scattergram (i.e., high-medium/gain versus high-medium/no gain). groups V and VI also produced two variables and one discriminant function that was sufficient to classify all the classes into the appropriate a priori groups (Table 13). The gain group had a lower ratio of children to adults (more adults per student) and more teacher acknowledgment in academic tasks.

In conclusion, the second discriminant analyses presented a statistical confirmation of the gain/no gain groups derived from the scattergram, providing the researchers with a measure of confidence in proceeding with analyses to determine within-group composition and intergroup differences. The remaining groups were kept primarily for descriptive purposes.

Table 13

Classification of Groups V and VI
by Discriminant Analysis

Actual group	Number of cases	Predicted Group V	Group membership Group VI
Group V	6	6 100.0%	0 0.0%
Group VI	6	0 0.0%	6 100.0%

Note: Percent of "grouped" cases correctly classified = 100 percent.

Analysis of variance. In order to examine intergroup differences, two sets of one-way analyses of variance, using the same variables selected for the discriminant analyses, were conducted. That is, all seven groups were compared in addition to Group V versus Group VI. See Table14 for descriptive statistics of the seven groups.

Medium high pretest scores. Twelve classrooms had medium high pretest scores. These pretest scores ranged from 54.7 to 70 percent. During the third grade year, six of these classrooms (Group V) made gains of between 14.9 and 23 percent. Three

Table 14

Descriptive Statistics of Seven Groups Examined by SRI International

Group	Number of classes	Mean number of students	Ratio		Absence		Pretest raw score		Post-test raw score		Gain raw score	
			X	S.D.	X	S.D.	X	S.D.	X	S.D.	X	S.D.
I Low pretest/no gain	3	31.0	13.03	2.00	10.70	3.65	6.44	2.07	8.63	0.42	2.19	2.29
II Low pretest/gain	5	29.6	15.00	5.72	12.58	4.06	8.63	0.55	12.30	1.52	3.67	1.36
III Low-medium pretest/gain	6	28.7	13.82	5.20	13.23	4.84	10.43	0.30	15.26	0.73	4.62	0.89
IV Medium pretest/gain	8	40.3	11.66	2.67	9.54	2.43	13.16	0.79	18.16	0.98	5.00	0.87
V High-medium pretest/gain	6	22.4	5.85	2.30	11.50	3.73	15.04	0.65	19.84	0.90	4.80	0.79
VI High-medium pretest/no gain	6	28.5	13.31	2.83	10.45	2.50	15.30	1.54	15.86	1.58	0.56	0.48
VII High pretest/gain	4	25.5	11.66	1.68	6.25	6.33	19.83	1.15	22.93	1.01	3.10	0.62
State average							16.92		20.42		3.25	

63

of these were located in schools identified as "increaser" schools, and three were in schools identified as "decreaser" schools. Six other classrooms (Group VI) made gains of only 0.9 to 6.9 percent. Two of these classrooms were in schools identified as "increaser" schools, and four were in schools identified as "decreaser" schools.

Twice as many students were present in the classroom for Group VI, the no-gain group, and the ratio of students to teacher was much lower in Group V, the gain group. The ratio was approximately six children to one adult in the gain classrooms, whereas there were approximately 13 children to one adult in the no-gain group. The difference in this ratio could account for much of the test score difference, but how the aides and teachers work with the children is of primary importance.

The two groups differed in the instructional processes used in several important ways. (See Table 15.) The gain group, for example, used more instructional materials and equipment, achievement charts, story books, and audiovisual equipment (variables 40, 48, 59, 237). The group also had a wider variety of activities occurring during the day (variables 82, 83).

The no-gain group had more arts and crafts and dramatic play occurring (variables 211, 281, 434) while the children in the gain group were more often involved in reading (variable 163). Even though the ratio of children to adults was higher in the no-gain group (18:1), the adults in the no-gain classrooms worked more frequently with one child at a time (variables 104, 124, 132). It follows, then, that other children in the no-gain classrooms worked alone more frequently than did children in the gain group (variables 117, 139, 165). This organization does not seem efficient for children, because many may not have the supervision, guidance, or reinforcement they need as they wait for their turn to work with the teacher.

The no-gain classrooms may, in general, provide less supervision. The teacher is more often out of the classroom and is less often directly involved with the children (variables 81, 272). This lack of supervision may be reflected in the fact that more negative behavior was exhibited in no-gain classrooms (variables 428, 434, 463).

In the gain group, the adults provide instruction more frequently and more often in a small group (variables 373, 376). They asked more direct academic questions; the children responded to these questions and received feedback from adults (variables 412). Much of the feedback was in the form of acknowledgment during these academic interactions (variable 395).

The teachers in gain classrooms also asked more open-ended questions, both academic and nonacademic (variables 355, 452). The children responded to these questions and more often with extended replies (variables 368, 369). Overall, the teachers in gain classrooms provided more acknowledgment and praise to children (variables 394, 404).

Summary of gain and no-gain analysis of variance. In the six classrooms of Group V, students started the third grade with test scores lower than the average for the state. In all six classrooms, the students gained as much or more during the year as the average gain for the state. In four of these six classrooms, the achievement was so great that their scores were as high or higher than the state

Table 15

Process Variables on Which Groups V and VI Differed

Variable number and description	Group V (n=6) $\bar{X}$	Group VI (n=6) $\bar{X}$	p
010 Number of children enrolled	22.40	40.33	.05
011 Number of children present	15.90	32.03	.04
015 Child/teacher and aide ratio	6.65	14.33	.01
016. Child/adult ratio (teachers, aides, and volunteers)	5.85	13.31	.001
040 General equipment, materials used	3.80	1.83	.07
048 Achievement charts presen.	1.00	0.50	.07
059 Children's storybooks used	0.80	0.17	.04
081 Teacher out of room	0.91	6.95	.04
082 Wide variety of activities, concurrent	1.60	1.09	.03
083 Wide variety of activities over one day	6.80	4.64	.07
104 Adult with one child	8.85	20.12	.08
117 Large group of children independent	10.12	22.23	.06
124 One child with aide--math	0.33	1.21	.04
132 One child with any adults--math	0.78	2.17	.09
139 Large group of children independent--math	17.31	48.33	.04
163 Approximate number of times children were recorded in reading over one day	100.00	85.00	.10
165 All children independent--reading	35.52	62.05	.06
211 Large group of children with teacher--arts and crafts	0.00	42.70	.08
237 Audiovisual equipment, academic activities	8.02	1.04	.08
272 Teacher not involved with children	0.00	6.52	.08
281 Number of FMOs, with arts and crafts as beginning activity	0.20	1.17	.07
355 Adult open-ended questions to children, nonacademic	0.08	0.01	.04
368 Child responds to adult open-ended question	0.11		.01
369 Child extended response, nonacademic	0.17	0.07	.05
373 Adult instructs all children	1.39	0.43	.09
376 Adult instructs a group	3.68	1.41	.05
394 All adul. acknowledgment to children	3.23	1.69	.02
395 Adult acknowledgment to children, academic	2.23	1.13	.04
397 Adult acknowledgment to children other than task-related	0.97	0.51	.09
404 Adult praise, other than task-related	0.26	0.08	.05
412 Adult feedback for child response to adult academic command, request, or direct question	1.55	0.68	.05
428 Negative behavior, adults to children	0.05	0.15	.09
434 Dramatic play, pretending	1.61	3.21	.02
452 All adult open-ended questions to children	0.18	0.02	.02
454 Child's extended response	0.64	0.22	.03
463 All negative behavior	0.05	0.15	.05

average by spring, 1977. (See Table 14.) In the classrooms that showed such gains, the classes were smaller and the ratio of children to adults was lower than in classrooms that did not show the gains. The teachers instructed small groups more often and used a direct approach to teaching reading. They used educational equipment and materials more often, including audiovisual equipment. During nonacademic activities they asked open-ended questions more often, and children made extended responses more often.

In Group VI, the no-gain classrooms, the teachers worked more often with one child at a time, and the other children worked independently. More arts and crafts and more dramatic play occurred. Teachers were less often directly involved with children, and there was more negative behavior displayed.

To promote better achievement in reading, the implications here for program planning would be, it seems, to encourage teachers to work with small groups, keep engaged with the children, allocate time for reading throughout the day, and use a direct method to teach reading (question-response-feedback).

These findings are very similar to those reported in the partial correlation analysis in the preceding section. The exception is that in this narrower band of ability levels, some important variables did emerge (open-ended questions, child-extended responses) that did not have an effect over the whole sample.

Analysis of Absence Rate Across All Classrooms

A high absence rate of students is a problem for teachers and students. The average absence rates for the 45 classrooms ranged from 1.0 day average per student to 22.5 days average per student. The burden on the teacher was great when students were frequently absent, because the teacher had to take precious time to help the student catch up. Also, there was a financial loss to schools in their average daily attendance dollars when students were absent.

In the Follow Through Classroom Observation Evaluation, 1972-73, the researchers for SRI International found that what teachers did in classrooms was related to how frequently children were absent from school. SRI identified process variables that accounted for 62 percent of the variance in absence rate in third grade classrooms. In classrooms where the students had more choice of materials, activities, and work groups, and when they could ask questions and socialize a little more, they were present more often. Since so much variance was explained by classroom processes, SRI considered absence rate as an indicator of the students' attitude toward school; i.e., even young children seem to control to some extent how frequently they attend school. SRI suggests that when children like their experience in their classrooms, they are absent less frequently; and when they do not like that experience, they are absent more frequently.

Based on this prior research, SRI wanted to see whether some classroom variables could be identified that were related to student absence rate in ECE classrooms. Correlations were computed using 150 selected classroom process variables with absence rate. Out of the 150 correlations, 54 were significant at $p < .05$.

Test scores. Classrooms with higher pretest and post-test scores had lower absence rates. SRI cannot, of course, interpret whether classrooms with lower absence rates are likely to have higher test scores or whether classrooms with

higher test scores are likely to have lower absence rates. Regardless of the direction, a significant relationship between test scores and absence rate was identified in this study. Since classrooms are within districts, there is also a correlation between districts and absence rate. Some districts had much higher test scores and also lower absence rates.

Classroom organization. The number of aides in the classroom and the ratio of students to adults were related to students' absence rate. The lower the ratio of students to adults, the lower the absence rate was. Students were absent more often where there were stationary desks in rows and where they were assigned to seats. They were absent less often where they could select their own seats and groups part of the time.

Group size. Several variables describing the organization of the groups and the responsibility of the adults were related to absence rate. In classrooms where the adults worked with one child at a time, the children were absent less often. Students were also absent less often in classrooms where adults often observed or attended to small groups of children. Although it sounds contradictory, in classrooms where children often worked independently either in reading or math, they were also absent less frequently. This suggests that children enjoyed the personalized attention but also enjoyed working independently.

Activities and materials. Three variables describing less task-oriented and more social activities--group time, transition, snack time, and lunchtime--had a positive relationship with absence rate, i.e., children were absent more frequently in classrooms where these activities occurred more frequently.[6] Wherever adults were less frequently involved with children or the children were not involved in any activity, the absence rate was higher.

Reading and math activities. The occurrence of reading and math activities had a negative relationship with absence. This suggested that children were absent less often in classrooms where more rigorous academic activities were occurring. This on-task pattern continued as SRI found the use of five academic-type materials related to lower absence rate.

Interactions. The students seemed to attend school more often in classrooms where there was a businesslike structured approach to education. The following variables describe a direct teaching method, and all of them had a negative relationship to absenteeism: teachers provide academic instruction, ask academic questions, child responds to question, and receives feedback or acknowledgment. When more of the total verbal interactions were academic in content, children were absent less often. Interestingly, several variables which describe a more relaxed interaction pattern where children initiate and adults respond were also negatively related to absence rate. As other studies have also found, when adults have to exert more behavior control and when the children express more unhappiness, the absence rate is higher.

Stepwise regressions. A stepwise regression was computed to see if any set of variables could explain the variance in the absence rate of classrooms. Only one variable was identified with a regression coefficient significant at $p < .05$. This variable describes the percent of time any academic activity is occurring, and it alone explained 26 percent of the variance in absence rate. This supported

[6] This finding was counter to the follow through findings that children were absent less frequently in classrooms that had more social interactions.

the findings in the correlations that children's attendance is better in classrooms which are more academically oriented.

Summary of absence rate. The process variables recorded in ECE classrooms related differently to absence rate than did process variables recorded in the follow through study. Follow through had many more open-classroom types of variables that related to lower absenteeism; ECE had a consistent pattern of having less absenteeism in more structured classrooms that had a lower ratio of children to adults.

Summary of classroom analysis. The variables which describe a more structured teacher-directed approach to education were also correlated positively with reading post-test scores. It follows then that in classrooms where the absence rate was lower, the pretest and post-test scores were higher.

Other ECE variables such as the presence of aides in the classrooms were positively associated with reading post-test scores and a lower absence rate. More individualized attention to children was associated with a lower absence rate but not with higher or lower reading test scores. A variety of instructional materials encouraged by the ECE program were associated with reading achievement and lower absence rate.

CONCLUSIONS AND RECOMMENDATIONS

The in-depth study of selected ECE classrooms indicates that the instructional methods or processes used in some classrooms are more efficient in teaching reading than those used in other classrooms. Of particular importance is the finding that when there are fewer students per adult, the classroom gain is greater and the absence rate is lower.

However, it is not enough to have more adults in the classroom; how the groups are organized and managed is of critical importance. Teachers in the classrooms that gained very little instructed one child at a time more often than teachers in classrooms that had good gains; however, the ratio of adults to children was quite high in classrooms that had poor gains. Of course, when the teacher was instructing one child at a time, the other students had to work for longer periods of time on their own. Misbehavior and negative effects occurred more frequently in classrooms that operated in this manner. This does not imply that the children should not be tutored or taught on an individual basis. It does imply that allocating one staff person to work with one student at a time within the class reading period when the total class is trying to work appears not to be an efficient strategy.

In the group of classrooms where students made excellent gains in reading, the adults worked often with small groups of students. They most often used directive teaching methods, providing instruction and asking questions or asking students to read aloud. The students' responses were given immediate feedback. If the response was correct or satisfactory, they received acknowledgment or praise, and if it was incorrect, they were guided to a correct response. These teachers were very successful in keeping students on specific tasks and, consequently, very little misbehavior was recorded in these classrooms.

Interestingly, three of the six classrooms that used the more efficient methods cited above were in schools identified as "increaser" schools, and three were in schools identified as "decreaser" schools. Stanford Research (SRI) International's

data do suggest that some of the schools in its sample and the classrooms therein do not fit neatly into the classifications of "increaser" or "decreaser" schools.

In 16 classrooms of "decreaser" schools, the students started the third grade year with test scores (16.60 to 50 percent) lower than the average for the state (67.7 percent). Nine of these classrooms whose students had low scores gained as much or more (14.9 to 24.8 percent gain) during the third grade as did the average classrooms in the state (14 percent average gain). Two of the classrooms in "decreaser" schools not only showed striking improvement but also achieved scores by the end of third grade as high as the state average for third grade (81.7 percent).

In the 20 classrooms of "increaser" schools, four started the third grade with scores above the average for the state. Twelve of the 20 gained more than the average gain for the state. Six of the 20 had scores at or above the state average for third grades at the end of the third grade year.

These data suggest that, over time, early childhood education (ECE) is effective. When the students in the 38 sample classrooms entered school in the first grade, they were in the lowest 20th percent percentile in the state. Now students in eight classrooms can no longer be defined as low achieving; they are at or above the state average. Twenty of the 38 classrooms of students are progressing nicely; they gained more than the average classroom of students gain between the end of second and third grade.

In the past, student achievement in California has most often been expressed in terms of the average percent of correct answers on the California assessment test, by grade level, for each school. School level scores are reported each year, as in the accompanying example.

School and grade	Scores, in percents	
	Spring, 1976	Spring, 1977
School X		
Second grade	42	41.5
Third grade	60.5	59
School Y		
Second grade	69	69.5
Third grade	80	81
State average		
Second grade	67.7	68.4
Third grade	81.4	81.7

The conclusion of such reporting could be that "School X's second and third grade classes have declining test scores (from 42 to 41.5 percent; from 60.5 to 59 percent), and they are failing in their efforts to teach children since they are below the average for the state." Actually, however, the gain in percent correct from second grade to third grade was greater in School X; between second

grade in 1976 and third grade in 1977, there was an 18 percent change, whereas the change in School Y was only 12 percent between second grade and third grade, and the average change for the state was 13 percent.

Local papers normally report the school scores by percentile, and rank schools accordingly; thus, School X would probably be placed at the bottom of the list. The fact that the teachers in School X have been able to improve the reading of low-achieving children almost a third more than the average second grade to third grade change for the state is ignored in this type of reporting.

The practice of reporting achievement in terms of school scores also ignores differences in the performance of teachers and students within the schools. Another problem in this type of reporting is in using schools as the level of analysis. Not long ago this practice received sharp criticism. In the American Sociological Review, Alexander and McDill summarized the school-level studies of Coleman (1966), Mostellen and Moynihan (1972), Jencks (1972), McDill and Riply (1973), and Alexander and Eckard (1975); the authors conclude that "educational outcomes are largely independent of all school-to-school differences. These results of the research on schools have been reproduced over a wide range of school outcomes on numerous samples of students with various analytic strategies. The acceptance of the proposition that school-level analyses are unproductive is reflected in the redirection of recent research on teaching and learning."

Many requests for proposals issued by the National Institute of Education and the U.S. Office of Education specify their preference for classroom or student levels of analysis and request that classroom processes be documented. Several such research efforts by Wiley and Harnischfeger (1976), Brophy (1975), Gage (1975), and Stallings (1975) report that quality of instruction within classrooms is related to student achievement. These researchers have noted that differences in student outcomes are associated with differences in teachers' classroom practices. Their findings, like those of this report, suggest that teachers who organize classrooms so that students are kept on task and not distracted have students who perform better on achievement tests. Such differences in teaching processes and student achievement are not detected in research that uses the school as the unit of analysis. The influence upon student achievement by teachers within schools is systematically neglected in traditional research on school effects. The assumption that education within schools is homogeneous is not a reasonable one.

The research for the SRI study was dedicated to the study of classroom instructional processes within ECE schools identified as performing in the lowest quintile on state assessment tests. SRI found that teachers within schools vary considerably in their instructional methods, and these differences have an impact upon the quantity and quality of schooling children receive.

SRI recommended that the findings from its study be used in program development efforts to assist teachers in organizing classrooms and interacting with students in those ways found to be most efficient.

SRI recognized that its study was limited in its scope. It focused on classroom instructional processes related to reading; it did not address several important outcomes for children expected as a result of the implementation of ECE. A more thorough analysis should assess classroom processes related to such student competences as the ability to work together cooperatively, to ask task-related questions, to identify problems (both academic and social), to consider alternatives and probable consequences, and to generate solutions. Attitudes toward

oneself and the school should also be examined. The data available did not permit those kinds of analyses, but such data could be collected through child-focused observation and selected affective and cognitive tests. SRI recommended that a further study of ECE should include such measures in order to examine more fully and fairly the impact of the program upon child growth and development.

79-19 DE11694 2-80 1M

Other Publications Available from the Department of Education

Report on the Special Studies of Selected ECE Schools with Increasing and Decreasing Reading Scores is one of approximately 450 publications that are available from the California State Department of Education. Some of the more recent publications or those most widely used are the following:

Accounting Procedures for Student Organizations (1979)	$ 1.50
An Assessment of the Writing Performance of California High School Seniors (1977)	2.75
Bicycle Rules of the Road in California (1977)	1.50
California Guide to Parent Participation in Driver Education (1978)	3.15
California Master Plan for Special Education (1974)	1.00†
California Private School Directory, 1979	5.00
California Public School Directory, 1980	11.00
California Public Schools Selected Statistics, 1977-78 (1979)	1.00
California School Accounting Manual (1978)	1.65
California School Energy Concepts (1978)	.85
California School Lighting Design and Evaluation (1978)	.85
California Schools Beyond Serrano (1979)	.85
Child Care and Development Services: Report of the Commission to Formulate a State Plan (1978)	2.50
Computers for Learning (1977)	1.25
Directory of Private Postsecondary Institutions in California (1978)	1.50
Discussion Guide for the California School Improvement Program (1978)	1.50*†
District Master Plan for School Improvement (1979)	1.50*
English Language Framework for California Public Schools (1976)	1.50
Establishing School Site Councils: The California School Improvement Program (1977)	1.50*†
Evaluation Report of Consolidated Application Programs (1979)	2.25
Genetic Conditions: A Resource Book and Instructional Guide (1977)	1.30
Guidance Services in Adult Education (1979)	2.25
Guide for Multicultural Education: Content and Context (1977)	1.25
Guide for Ongoing Planning (1977)	1.10
Handbook for Assessing an Elementary School Program (1978)	1.50*
Handbook for Assessing a Secondary School Program (1979)	1.50*
Handbook for Instruction on Aging (1978)	1.75
Handbook for Planning an Effective Reading Program (1979)	1.50*
Handbook for Reporting and Using Test Results (1976)	8.50
A Handbook Regarding the Privacy and Disclosure of Pupil Records (1978)	.85
Health Instruction Framework for California Public Schools (1978)	1.35
Improving the Human Environment of Schools (1979)	2.50
Liability Insurance in California Public Schools (1978)	2.00
A New Era in Special Education: California Master Plan in Action (1979)	2.00
Parents Can Be Partners (1978)	1.35*
Pedestrian Rules of the Road in California (1979)	1.50
Physical Education for Children, Ages Four Through Nine (1978)	2.50
Planning for Multicultural Education as a Part of School Improvement (1979)	1.25*
Planning Handbook (1978)	1.50*
Publicizing Adult Education Programs (1978)	2.00
Putting It Together With Parents (1979)	.85†
Report of the Ad Hoc Committee on Integrated Educational Programs (1978)	2.60
Science Framework for California Public Schools (1978)	1.65
Site Management (1977)	1.50
Social Sciences Education Framework for California Public Schools (1975)	1.10
State Guidelines for School Athletic Programs (1978)	2.20
Student Achievement in California Schools (1979)	1.25
Students' Rights and Responsibilities Handbook (1978)	1.50†
Teaching About Sexually Transmitted Diseases (1979)	1.65
A Unified Approach to Occupational Education: Report of the Commission on Vocational Education (1979)	2.00

Orders should be directed to:

California State Department of Education
P.O. Box 271
Sacramento, CA 95802

Remittance or purchase order must accompany order. Purchase orders without checks are accepted only from government agencies in California. Sales tax should be added to all orders from California purchasers.

A complete list of publications available from the Department may be obtained by writing to the address listed above.

†Also available in Spanish, at the price indicated.
*Developed for implementation of AB 65.