

Anatomy of a Major Evaluation:

Context, Procedures, and Results of a Two-year Evaluation of a School District's Reading Instruction

Leslie Hendrickson and Charles Stephens
Research, Development and Evaluation Division
Eugene Public Schools
Eugene, Oregon

Despite widespread public concern with the cost and quality of services provided by school districts, comprehensive evaluations of reading, writing, or mathematics are not frequently done. In 1976 the Eugene, Oregon, School Board appointed a seven-person Reading Evaluation Committee to evaluate the reading instruction provided the district's 20,000 students. Their three-volume, 474-page report was presented to the board in January 1978.

Reading and mathematics achievement scores in the district's elementary grades have risen up to twenty-five percentile points in the last four years¹. In 1974-1975 the district's achievement test scores in reading were at or below the fiftieth percentile. (The district had average test scores.) The evaluation described in this article was part of a broad and successful effort to improve basic-skills instruction.

The case study of a large, complicated evaluation is illustrative because it shows

the wide range of political and procedural problems that evaluators typically encounter. This article introduces the chronology of events, describes the evaluation process, and discusses the results and lessons learned from it.

An understanding of the background within which the evaluation occurred is useful in understanding the evaluation's specifics.

In 1974 a new superintendent was hired by the Eugene School Board. The superintendent's reorganization plan, among other actions, created a decentralized administration and a centralized research, planning, and evaluation unit. In October 1975 the school board requested that a comprehensive plan for assessing reading instruction be developed. The first phase of work required principals to prepare descriptions of their school's reading programs. These descriptions were collected and reported on in March 1976. Discussions of these reports within the school district's top administrative group raised credibility questions about the self-reports.

In April 1976 a comprehensive evaluation was begun in the district's thirty-one elementary schools.

All but one of the district's high school principals were transferred to junior high

¹ The comparison of reading scores was based on Gates MacGinitie scores in 1974-1975 and Metropolitan Achievement Test Form F 1970 and for later years. Norm tables from the Anchor Test Study were used for equating these data. Analysts found aggravating mistakes in the Anchor norm procedures. Mathematic scores were all based on the Metropolitan Achievement Test.

schools in April 1976². A five-person evaluation unit was formed and its first tasks were to review the literature on reading evaluations and contact other school districts. A review of literature published since 1960 was made to find articles or books reporting on actual reading evaluations carried out by schools districts. A hand search of the 5,300 titles in the 1960-1968 Education Index listed 82 references relevant to the evaluation of reading programs. An ERIC computer search of published literature since 1968 listed 49 references to reading programs. All published references were read. No articles or books describing actual evaluations were found. The five largest school districts and the state department of education in eleven western states were called and asked if they knew of any school districts which had carried out an evaluation of their reading instruction. One school district was found which had carried out an evaluation similar to the planned evaluation and a copy of its evaluation report was obtained.

Approximately 25 interviews with school district staff were carried out to obtain background data prior to making an evaluation design.

In August 1976 the school board appointed a seven-person evaluation committee to provide policy advice during the evaluation. The committee consisted of one elementary principal, one teacher, a parent, a school board member, a senior administrator, and a faculty member from the College of Education and the Univer-

² Although a recall election of the school board narrowly failed in the fall of 1976, aggressive newspaper reporting and a political struggle for control of the school board were background factors that affected the evaluation processes used. The Oregon Supreme Court is now reviewing the lawsuits that emerged from the transfers. In April 1978 new school-board members were elected who were critical of the superintendent. Perceiving the loss of a once supportive board, in December 1978 the superintendent announced his acceptance of a job as head of the Oklahoma City, Oklahoma, schools.

Many observers of collective bargaining in Oregon school and college districts agree that management-labor lines are becoming clearer. The Eugene Educational Association, which represents teachers in collective bargaining, closely watched the progress of the evaluation.

sity of Oregon. The Eugene Educational Association, which represents the district's 1,100 teachers, was asked to appoint two of the seven committee members. They appointed a reading specialist, who was the immediate past president of the association, and a fifth-grade teacher.

As its first meeting, evaluators told the committee that there were no standards available which specified of what a good reading evaluation should consist. Evaluators further stated they were unable to obtain agreement from their sources as to which behaviors constituted the teaching of reading. A list of approximately 100 possible topics that could be studied were given to the committee.

In October 1976 the committee decided that the purpose of the evaluation was threefold, and evaluators were requested to determine (1) whether the district's elementary reading programs were effective according to the district's curriculum documents which specify what subjects should be studied at what grade level; (2) what makes effective programs effective; and (3) what could be done to make ineffective programs effective.

By January 1, 1977, the committee had reviewed the 100 topics, ranked them as to priority, made cost estimates for each topic, and decided to study 26 questions within the threefold purposes of the evaluation.

The committee agreed to evaluate the following kinds of reading activities: comprehension, vocabulary, oral reading, word analysis, study skills, and silent reading. It was not possible to systematically evaluate all six activities and in May 1977 the committee decided to concentrate on the first two.

Data collection took place from January to July 1977. Table 1 shows the numbers and kinds of people who were contacted. Previous evaluations in the district had shown there were reliability problems in the district's computerized personnel files. Accordingly, a list of all certified staff in the elementary schools was drawn up by evaluators and checked person by person to determine how many and which people taught "a regular scheduled reading period."

As Te
staff w
instruct
100 per
all cate
All s
written
percent
cialists
membe
ber of c
These
whethe
what
should
reading
policy
should
assign
chang
teache
ing m:
organ
In a
ers an
ing re
Over

N

Quest
Pe
Ck
Ce
bu
To
wr
Lit
Sp
Ce
To
Pers
Pr
To
R
To
lit
Obs
C
C

As Table 1 shows, almost all of the 464 staff who were concerned with reading instruction were sent a questionnaire. A 100 percent response was obtained from all categories of staff.

All staff reading specialists received a written questionnaire. Approximately 50 percent of the staff, including reading specialists, were personally interviewed. Staff members were asked to respond to a number of questions about reading instruction. These questions sought to determine whether the district had expectations of what teachers and reading specialists should be capable of doing when teaching reading, whether a school had a written policy stating how the teaching of reading should be organized, who (if anyone) is assigned responsibility for reviewing and changing reading programs, whether teachers have agreed to use the same reading materials, and just how each school is organized for reading instruction.

In addition, 66 randomly selected teachers and 1,100 students were observed during regularly scheduled reading periods. Over 68,000 observations of students and

over 67,000 observations of teachers were recorded.

All principals were asked to fill out a cost questionnaire. In July 1976 the school district adopted the United States Office of Education program-budgeting system, known as Handbook II in Oregon. The conversion of account structures made cost finding difficult.

The study was also made more difficult because of 1) the lack of comparable information about reading performance from one school in the system to the next and from one year to the next; 2) the difference in materials and techniques used to test reading skills from one year to the next; and 3) the lack of consistency in identifying students as they progressed through the school system.

Data on the actual readability levels of the Houghton Mifflin 1974 and 1976 reading series and the Ginn 360 and 720 series were obtained. Substantial discrepancies between the publisher's designated reading level and the actual reading level of reading materials were found.

A three-to-four page report was made

TABLE I
Number of People Who Received Questionnaires, Were Personally Interviewed, and/or Were Observed Teaching Reading, January 1, 1977-June 30, 1977

Means of Contact	A Total Number in Category	B Number Contacted	C Percentage Contacted
Questionnaire			
<i>Persons who were sent a questionnaire</i>			
Classroom teachers who teach reading	397	397	100%
Certified staff members who are not classroom teachers, but teach reading	5	5	100%
Total classroom teachers and other certified staff members who teach reading (Total of two categories above)	402	402	100%
Librarians	22	22	100%
Special education teachers	14	14	100%
Certified staff who did not teach a reading period	270	00	000%
Total certified staff	708	438	62%
Personal Interview			
<i>Persons who were interviewed</i>			
Total certified staff members who teach reading	402	193	45%
Reading specialists	26	26	100%
Total certified staff concerned with reading (Total teachers, librarians, special education teachers, reading specialists)	464	219	47%
Observations			
<i>Category of persons observed</i>			
Classroom teachers who teach reading	397	66	17%

on each school. This report described the program, summarized the advantages and disadvantages, and presented data on 20 variables describing the school's program.

In summer of 1977 the evaluation unit met repeatedly to deal with the now-documented facts that less than one out of six schools had a program for teaching reading. Evaluators were unable to fulfill the committee's threefold purpose because there were no programs to evaluate. The evaluators concluded they were being asked to evaluate objects which did not exist.

This impasse was resolved with the construction of an "Organizational Continuity Index". All data elements pertinent to program planning, coordination, and staff communication were identified. A 32-item index was created. The index yielded a score for each school summarizing how well organized it was for teaching reading.

Approximately 100 variables in all were collected measuring characteristics of the 34 elementary schools studied.

Results

An analysis of scores on nationally used reading tests indicated there had been dramatic improvement in reading performance at most schools in the district over the past two years. On the average, students in the second-grade classes scored in the top one-third of the nationwide student population that took these annual reading tests, and fourth- and sixth-grade scores were above national averages. Analysis of spring 1978 scores shows these upward trends continued.

Results of standardized tests and teacher judgment of the situation in their schools agreed that 15 to 20 percent of the children enrolled in the schools were experiencing reading difficulties. One out of six schools had more than 25 percent of their sixth-grade students scoring in the bottom one-third of the student population taking the test each year nationally. A multiple regression analysis of income, education, and withdrawal rates was used to develop "expected" scores for each school; that is, levels of performance at which students in those schools might be expected to perform if reading instruction

were effective. These expectations were contrasted with actual scores.

A longitudinal analysis revealed that students who stayed in the same district school are likely to learn more than those who left or entered the reading programs during the year or who change schools after being enrolled in a particular reading program for some time. Average scores for students who stayed in the same school during the test period (1974-1977) were dramatically higher than the test average scores for all other students at their grade level in the district.

Two major basic reading series were used in the Eugene School District, Ginn and Houghton Mifflin. Thirty-nine different main materials were used by the four hundred teachers. An analysis of how hard it is to read Ginn and Houghton Mifflin books shows that these publishing companies do not provide reliable information about how hard it is to read their books. The misleading information provided lowers the effectiveness of the district's reading instruction.

The study of record-keeping systems showed that existing record-keeping systems became inefficient when student entry or withdrawal rates were higher than 30 percent. Test scores on standardized tests are the most frequently transferred student records, and reading-skill-development records were maintained the most.

A review of data on classroom instructional patterns showed that schools with more reading- and media-specialist time assigned to them had smaller reading groups, met less frequently, and had shorter meetings. The more reading- and media-specialist time schools had, on the average, the less time classroom teachers spent teaching reading. These arguably unintended results fit into the findings about problems in the administration of district reading and media specialists. Teachers in 14 schools reported that children with reading problems received help only from a classroom teacher.

The proportion of students identified by teachers in each school as having reading problems was negatively and significantly correlated with nine variables measuring achievement on reading tests. The higher

the l
lowe

So
readi
school
clude
to tea
ness
readi
cially
consi
of the
distri
that s

suppe

Cor

staff

ment

no or

instru

strati

not er

enoug

in rea

ruptiv

ing th

not in

school

At t

conclu

not ha

expect

ing spe

der to

agreen

who h

for rev

grams,

ment.

sponsil

there v

stood d

ing in t

Neith

nation

ment,

school

The rep

job title

tivities,

adminis

school p

Appr

were id

were the proportion of students identified, the lower the average test scores.

Some strengths and weaknesses in the reading programs were common to several schools. Commonly observed strengths included 1) development of a staff agreement to teach from similar materials; 2) willingness of staff to work on improvement of reading skills; 3) help from persons specially trained in reading instruction; 4) consistency and continuity between levels of the reading program; and 5) use of the district's list of reading skills to make sure that students learned the skills they were supposed to learn.

Common weaknesses included 1) lack of staff commitment to carry out an agreement about the use of similar materials; 2) no organized school program for reading instruction because of ineffective administrative leadership; 3) lack of materials; 4) not enough time spent on reading; 5) not enough help from those specially trained in reading instruction; 6) a high and disruptive rate of students leaving and entering the program through the year; and 7) not integrating library resources into the school's instruction.

At the district-wide level, the committee concluded that the district, as a whole, did not have a written or agreed-upon set of expectations as to what teachers and reading specialists should be capable of in order to teach reading. There was greater agreement among administrators as to who had responsibility at the school level for reviewing and changing reading programs, but there was no uniform agreement. Methods for carrying out such responsibilities were not specified. In short, there was no written or commonly understood district-wide plan for teaching reading in the Eugene public schools.

Neither was there district-wide coordination regarding the selection, job assignment, work activities, or work load of school librarians and reading specialists. The report stated that their wide variety of job titles, academic preparation, work activities, and the like were indicative of an administratively disorganized category of school personnel.

Approximately one out of five schools were identified as having ineffective ad-

ministrative leadership in its reading instruction.

In an attempt to determine exact costs of reading instruction and what effect expenditures might have on reading abilities from school to school, the committee concluded that the district and individual school budgeting and accounting systems are not detailed enough to provide data for an accurate and uniform record of the amount of money expended for reading instruction. Maximum and minimum cost estimates were prepared.

A random sample of second-, fourth-, and sixth-grade teachers was selected, and observations of their regularly scheduled reading periods were made. Students were doing what they were supposed to be doing 90 percent of the time. Teachers were seen to be directly instructing their students 75 percent of the time. Observations from different classrooms within the same school were analyzed and found to vary widely.

The study closed by observing that, in future evaluations, better data were needed on individual classrooms and teachers because much information is lost when data for a single school are summarized. This loss of information made it difficult to identify which classroom instructional patterns were most related to effective reading instruction.

Recommendations

Based on its analysis, the committee recommended

1. The district administration should develop a data collecting system which allows collection of comparable data from each school and students.
2. The district administration should develop a consistent system for student identification to monitor student progress.
3. The district administration should examine the discrepancy between the second- and fourth-grade test scores to determine the explanation for the discrepancy.
4. The district administration should examine the 15 to 20 percent of the students with reading problems to

- determine the nature of and severity of the problems and what is being done for these students.
5. The district administration should initiate an analysis of those schools that have less effective programs and take the necessary steps to make them effective.
 6. The district administration should direct each principal to review the reading evaluation report to determine the areas in which the school's reading program can be improved. The principal should be responsible for developing and implementing plans to correct the reading program deficiencies.
 7. The district administration should assign responsibility for reviewing new reading programs and changing old programs.
 8. The district administration should direct principals to examine the role of media specialists in their schools, in order to provide a library program that is integrated with the reading program.
 9. The district administration should be responsible for specifying competencies required of elementary:
 - a) teachers (for teaching reading),
 - b) principals (for administering a school's reading instruction), and
 - c) reading specialists.
 10. The district administration should be responsible for assessing the elementary teachers', principals', and reading specialists' competencies for teaching and/or administering the reading program.
 11. The district administration should be responsible for developing and implementing an in-service program for elementary teachers, principals, and reading specialists who have not yet met the required competencies for reading instruction.
 12. The board should be responsible for modifying its hiring, promotion, and retention policies to insure that teachers, principals, and reading specialists have the required competencies for reading instruction.
 13. The district administration should

examine the role of the reading specialist in the district's reading program to determine:

- a) the need for district coordination of specialists,
 - b) responsibilities of the specialist in the individual school program, and
 - c) whether each school can benefit from the services of a reading specialist.
14. The district administration should establish criteria for an effective reading program including an evaluation component and require each school to develop a plan for reading instruction which meets those criteria.

In January 1978 the Eugene School Board unanimously adopted the recommendations, and they were implemented in the 1978-1979 school year.

Lessons

This was a unique and successful evaluation. No evaluation like this had ever been done in the school district, nor are such comprehensive evaluations typically done in school districts. What were the advantages of the evaluation?

First, wide support existed for a review of reading instruction. The school board, superintendent, and many administrators and teachers wanted a rigorous, comprehensive look at the ways the district taught reading. Many, and often conflicting, questions were raised by different audiences.

Second, the process had checks and balances. The committee selected the topics and approved the major data-collecting procedures. Draft copies of chapters dealing with the individual schools were sent to schools staffs before going to the committee. The staffs of the school were thus provided opportunities to negotiate the content of what would be said about them publicly. Drafts of pertinent chapters were also sent to reading specialists and librarians.

Almost-final drafts of the evaluation were read by consultants from the National Assessment of Educational Progress (NAEP) in Denver and the Center for the

Study
consult
the su
tional
from th
(NEA)
draft
evaluat
side sci
Othe
by keep
apprais
exampl
data-co
people
three. T
top adr
and tea
two to l
viewed
studies
implicat
data. Ma
cussed a
sial late
could be
tions.

Third,
and the
lightene
of varia
to exami
tives. Fo
requested
school's
one mea
some dis
never dic
Rather, t
indicator
rough-an
participa
enough d
understo

Fourth,
into the
perintenc
to act in
uation ur
tendent. T
omy of b
committe
was mair
and eval

Study of Evaluation at Los Angeles. These consultants were hired by and reported to the superintendent. The Eugene Educational Association brought in a consultant from the National Educational Association (NEA) in Washington to review the final draft. Thus, in the closing weeks of the evaluation, it underwent systematic outside scrutiny.

Other checks and balances were created by keeping major audiences systematically apprised of the evaluation's progress. For example, in August 1977 at the end of the data-collection phase, approximately 30 people were briefed in groups of two and three. These briefings of board members, top administrators, committee members, and teacher representatives lasted from two to four hours and chronologically reviewed the evaluation design, the kinds of studies undertaken, and possible policy implications as indicated by preliminary data. Major preliminary findings were discussed and items that might be controversial later were identified so the audiences could begin to prepare their public positions.

Third, the large amount of data collected and the magnitude of the study enlightened participants. The large number of variables studied allowed participants to examine a topic from different perspectives. For example, evaluators were requested to state the "effectiveness" of each school's reading program. However, any one measure of effectiveness always had some disadvantage. The evaluation report never did define what "effectiveness" was. Rather, the cumulative impact of multiple indicators more accurately reflected the rough-and-tumble complexity that most participants perceived while providing enough data so that most people felt they understood how well schools were doing.

Fourth, ambiguity was deliberately built into the committee's charge from the superintendent. The committee was charged to act in an advisory capacity to the evaluation unit and to report to the superintendent. This charge maximized the autonomy of both the evaluation unit and the committee. A high personal involvement was maintained by committee members and evaluators. Decision-making in con-

troversial situations was done on a consensus basis. The resultant process was that the committee generally decided what to study and the evaluators decided how to study it.

Fifth, because the subject of the evaluation was politically acceptable, all members of a divided school board could agree that reading instruction was both desirable and important to study.

What were the disadvantages of the evaluation? Corresponding to the advantages, there was a parallel series of disadvantages.

First, there was considerable apprehension among principals and teachers because of the evaluation's broad scope and intensive digging. Some organization units reacted sharply to being "criticized" by another school district unit.

Second, the advisory committee structure worked well until the press attended meetings and began quoting committee members. A reporter from the *Eugene Register-Guard*, the largest paper in the country, told the evaluation staff that unless he was allowed to attend committee meetings the paper would take the district to court. A review of the Oregon Open Meeting Law by the district's legal counsel concluded that the committee was an extension of the board and also subject to the provisions of the open-meeting law. The first article about the committee was in November 1977. The headline was "Schools Fail at Reading." Individual committee members were quoted. (For example, the board member on the committee was quoted as saying that the district's personnel office had no standards for hiring elementary teachers.) The candid, often blunt, advice and policy-making from the committee dried up.

Third, more data was collected than should have been. Program-cost data should not have been collected on elementary schools. More were collected than could be analyzed using conventional statistical packages such as the BMDP, Biomedical Computer Programs, and the Statistical Package for the Social Sciences (SPSS). The lack of sophisticated computer programs such as LISREL, COFAMM, or MIMAC made it difficult to sort

out the structure of the 100 school-level-variables data base.

Fourth, the amount and complexity of the data were more than audiences could efficiently absorb. The school board and top administrators spent nine hours in work sessions absorbing the data. Many of the people did not know what a mean or standard deviation was and the use of correlation coefficients, discriminant analysis, multiple regression, factor analysis, and other statistical procedures was incomprehensible to some people. (For example, two full pages of a *Sunday Register-Guard* was used to report school-by-school results. These pages contained errors because of difficulty the reporter encountered in interpreting some results.)

Finally, the stormy political atmosphere

of the district required evaluators to avoid the appearance of favoritism. For example, data collected about the lack of district-wide coordination in reading instruction was used by opponents of the superintendent's 1974 reorganization plan.

Concluding Comment

How much of the district's remarkable improvement in reading scores is due to the evaluation? This question cannot be answered precisely because many factors are operating. However, the long and comprehensive evaluation with its public reporting of results undoubtedly made some contribution. It certainly illuminated significant factors contributing to the success or failure of reading instruction.

Trad
models
treatme
outcom
or "regu
output
ized noi
program
sider ei
school-p
between
come. N
there is
the wor
who use
ence of
is a ma
links in
school
directly.

An ev
preparec
Spanish-
has atten
deficient
been ma
bilingual
Laosa, 19
ner-Goug
1975; Pa
tion-desi
about the