DOCUMENT RESUME

ED 198 176 TM 810 144

AUTHOR Leinhardt, Gaea; Engel, Mary
TITLE Iterative Evaluaton: NRS. An Example.
INSTITUTION Pittsburgh Univ., Pa. Learning Research and Development Center.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
REPORT NO LRDC-1980/12
PUB DATE 79
NOTE 36p.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Evaluation Methods; Grade 1; *Instructional Innovation; Primary Education; *Program Evaluation; Program Implementation; *Reading Instruction
IDENTIFIERS *Iterative Methods; *New Primary Grades Reading System

ABSTRACT

The use of increasingly more informative evaluations of a single innovation are documented. The innovation, The New Primary Grades Reading System (NRS), was implemented in a variety of settings, and the evaluations track the implementation from the early stages of pilot testing through large-scale adoption of the program. NRS is characterized as an eclectic code-breaking approach to beginning reading instruction as opposed to a whole-word approach. It employs letter-sound correspondences, synthetic phonics, analytic phonics and linguistic techniques such as displaying differences and similarities in spelling patterns. NRS is both individualized and adaptive. Once all the students are at or above Level III and receiving instructions from audio cassettes, the teacher is free to "travel" around the room to diagnose, prescribe, teach and monitor student progress and to assess student performance. Many studies of the effectiveness of NRS have been conducted. A sample of seven of the 20 or more studies are reported here. There is one within-school study, five within-district across-school studies, and one across-district study. In all cases there is pre/posttest information on grade 1 students. Cumulation of evidence across settings allows for greater confidence in conclusions drawn from the evaluation. (Author/RL)

LEARNING RESEARCH AND DEVELOPMENT CENTER

ITERATIVE EVALUATION: NRS, AN EXAMPLE

1980/12

GAEA LEINHARDT AND MARY ENGEL

University of Pittsburgh

ITERATIVE EVALUATION: NRS, AN EXAMPLE

Gaea Leinhardt and Mary Engel

Learning Research and Development Center

University of Pittsburgh

1979

The research reported herein was supported by the Learning Research and Development Center, supported in part as a research and development center by funds from the National Institute of Education (NIE), United States Department of Health, Education, and Welfare. The opinions expressed do not necessarily reflect the position or policy of NIE, and no official endorsement should be inferred.

Abstract

This paper documents the use of increasingly more informative evaluations of a single innovation. The innovation, The New Primary Grades Reading System (NRS), was implemented in a variety of settings, and the evaluations track the implementation from the early stages of pilot testing through large-scale adoption of the program. Accompanying contrasts were made within a single school, within a school district, and across districts. This approach to evaluation followed the naturally occurring process of implementation of the innovation and provides convergent evidence as to its effects. The cumulation of evidence regarding the success of an innovation across settings allows for greater confidence in the conclusions drawn from the evaluation.

ITERATIVE EVALUATION: NRS, AN EXAMPLE

Gaea Leinhardt and Mary Engel

Learning Research and Development Center
University of Pittsburgh

Since the mid-1960's educational evaluation has become a prominent aspect of educational reform and innovation. Evaluations of both major educational changes and minor curriculum revisions tend to occur after the design and implementation of the innovation rather than during it. There are rational and defensible reasons for this. First, when an innovation of any type is supported, it is supported on the assumption that some change or improvement is desirable. Whether the change represents an improvement is another question. Second, the individuals involved in designing, implementing, and revising an innovation are not necessarily the same people who design and implement an evaluation; therefore, especially in curriculum design, evaluation efforts are tacked on after development.

Two important approaches to evaluation of curricular innovations are often proposed and rarely followed. One approach is the design and implementation of an experiment with random assignment of units (students, classes, or schools) to treatment and tight control to avoid deterioration of the design. The results of such a study should be clear cut (Boruch & Rindskopf, 1977). The process of true randomization and selection of the correct unit for assignment has proven difficult but by no means impossible in education. Homogeneity of treatment within groups and separation of treatment between groups has been a more difficult problem. A second approach is the design and implementation of an explanatory observational study (Cooley,

1978). Such a study should start with a model of how the innovation may work and include measures of the independent variables. While this approach is undoubtedly the one that will be followed in most large scale field-based evaluations, the process of delineating a model and measuring its components requires large numbers of observational units (students, classrooms, schools), and it is therefore unlikely that during early stages of curriculum evaluation this approach will be followed because of cost and complexity of administration.

The purpose of this paper is to examine the feasibility of a third, somewhat more pragmatic, approach to evaluating curriculum innovations. To use this approach, the innovation is followed through its implementation in different educational settings, data are gathered at whatever level possible on its effects, and then the total picture of program effects is constructed by analyzing the convergent and divergent evidence. This approach follows the natural route of program development; that is, programs start at some level, positive "results" at that level lead to expansion, and then further assessment is followed by further expansion. One can visualize this approach as a set of concentric rings moving out from a single limited contrast to larger and more complex ones. The rings are both the levels of expansion for the innovation and the levels of contrast for the evaluation.

This paper draws together a set of small scale evaluations of an educational innovation, the New Primary Grades Reading System (NRS; Beck, 1977). NRS was implemented in a variety of settings starting in 1973. The settings were of increasing physical distance from the developers and tended to have decreasing intervention from the implementation staff. At some point in each new implementation situation, some evaluative data were collected. This paper describes the NRS innovation and presents a chronological overview of its implementation and the type of contrasting information available. The

evaluations are described as moving with the implementations from historical within-school contrasts to more taxing concurrent across-district contrasts. The set of studies is followed by a synthesis of the impact of NRS on classroom processes and student achievement.

Background

Before exploring the details of the individual studies that evolved from the development and implementation of NRS, a description of the reading system, how it operates in the classroom, and a brief history of its development are reviewed.

The NRS Approach

NRS is characterized by its developers as an eclectic code-breaking approach to beginning reading instruction as opposed to a whole-word approach (Beck, 1977; Beck, Note 1). Children are taught by receiving explicit instruction about letter-sound correspondences that can be used in attempting to read the word. One way in which children are taught to read unfamiliar words is by using synthetic phonics, or putting sounds together to form words. The strategy developed for use in putting the sounds together is called the blending chain. A second way children are taught to approach unfamiliar words in NRS is to explore the word for its parts, or, more generally, to use analytic phonics. NRS also teaches children about linguistic techniques that can be used in learning to read words. Words and text within NRS are displayed so that differences and similarities in spelling patterns are evident to the children.

The developers of NRS emphasize its individualized, adaptive structure. The reading system is individualized in that students, after completing the first two levels of the curriculum, work through the curriculum at different rates. The curriculum is adaptive in that

teachers are provided with multiple teaching strategies for use with different students. The curriculum also provides the students with a wide variety of instructional resources which can be used to adapt to their individual needs and interests. NRS is also designed to teach children self-management skills, permitting students to work on an activity or to move between activities without the supervision of the teacher.

The first two levels of the curriculum are designed to teach self-management skills, letter-sound correspondences, and techniques for blending those sounds into words. Most of the academic and managerial instruction in these early levels of NRS occurs in teacher-led small groups. As the first student enters Level III of the curriculum, the teacher's role changes: The students receive initial instruction in new content or skills from audio cassettes, and, once all of the students are at or above Level III, the teacher is free to "travel" around the room to diagnose, prescribe, teach, and monitor student progress and to assess student performance.

Prescriptive activities assigned by the teacher must be completed by the student. Each child knows his/her prescription for the day by looking at a prescription sheet that is sequenced within the child's own consumable workbook. Prescriptive activities also include teacher-led instruction, cassette-led instruction, and tests or progress checks. In addition to prescriptive activities, children are assigned or select reinforcement and fluency-building activities. These activities may include reading a story written to be appropriate for the level in which the child is working, playing specially designed games, reading a story in a teacher-led group situation, reading books other than those in NRS, or writing stories.

Chronology of NRS

Shortly after its inception in 1964, the Learning Research and Development Center (LRDC) became involved with two public elementary schools: one located in a major urban area, and the other in a suburb. LRDC was attempting to individualize elementary school instruction in reading, mathematics, and science. The research and development of individualized reading instruction began with two basic restrictions in mind: that the reading curriculum should emphasize code-breaking, and that the LRDC should not attempt to develop its own reading curriculum, but should try to modify existing ones (Beck, 1977).

A published, programmed text was in use in reading classrooms at both schools in 1970. The text was designed so that the teacher presented new content to students; students, in turn, practiced skills in their individual workbooks after content was presented. It became apparent that a teacher in a classroom with a large number of students could not present content to each student, one at a time. The LRDC developed, from the instructional strategies available in the teacher's manual, audio lessons on cassette tapes to take some of the introduction of new material off the teachers' hands. The addition of the tapes to the curriculum increased the amount of time the teacher had at his/her disposal for tutoring students and monitoring their performance.

A later modification of the published program made by the researchers was the introduction of the Early Reading Program (ERP), which incorporated new instructional strategies for the introduction of letter-sound correspondence and the blending of those sounds into words. As these and other changes of the program were being tried out in the developmental schools, the reading project staff began to examine the results of standardized reading achievement tests. The results were encouraging, but no statement could have been made linking the changes to

modifications in the curriculum. Observations of children made by the staff indicated that many were still having difficulty in acquiring reading skills, and the staff continued to be dissatisfied with the curriculum itself for a number of other reasons--such as the inappropriate use of "fill in the blanks" and the inappropriate introduction and practice of similar graphemes (e.g., b and d). For these reasons, a decision to develop a new reading curriculum was made.

The reading staff gradually began implementing NRS in the two developmental schools in 1972-1973. Data from the 1973-1974 school year at both schools for first grade classrooms revealed that achievement for NRS students was good as compared to first grade students in the previous year who had been taught with the patched-up curriculum (Beck, 1977; Buehler & Eichelberger, Note 3; Eichelberger & Lee, Note 2). These data were difficult to interpret for several reasons. Any first grade student who was taught reading using NRS probably received some beginning reading instruction in kindergarten using the old system. The reading staff also provided assistance in the classrooms, for those children who needed it, before 1973-1974. During and after that year, the assistance was removed.

As the LRDC reading staff and curriculum developers changed their approach from rewriting and revising a beginning reading curriculum to developing and implementing a new one, the need for further evaluation of the curriculum became more pressing. The study of the implementation and its effects on student achievement become more structured as finer and more numerous changes were being made in NRS. Strengths and weaknesses of the program were being brought to light. Since all evaluations of NRS had been conducted in schools with long-term, ongoing relationships with LRDC (i.e., they were using other individualized curricula at the same time), and because of the relatively positive results of NRS, it was decided to implement

and evaluate the reading program in a setting that had had no previous ties to LRDC.

The 1974-75 school year represented a major change for NRS, the revised program was implemented in three first grade classrooms at School 1 in District A (Beck, 1977; Wilkinsburg Gazette, 1974; McCaslin, Note 4).[1] This represents a second ring out from the original developmental site. All first graders at School 1 used NRS that year; 40 students remained in the program for the entire school year. This group of 40 students was compared with a group of 43 students who were in first grade in 1973-1974 at School 1 and who had used a basal reading program. The LRDC reading staff was responsible for preservice/inservice training as well as extensive support activities during that first year of implementation at School 1. In addition, the reading staff was available to the teachers and staff at the school to answer questions, solve problems, collect data on student progress in NRS, supervise the distribution and quality of curricular materials, and supervise the administration of standardized tests at the end of the school year (McCaslin, Note 4) to the 1974-1975 first graders and to the 1974-1975 second graders (1973-1974 first graders). Results of this first year were positive.

District administrators, principals, teachers, and students had favorable reactions to the first attempts at implementation of NRS in District A (Wilkinsburg Gazette, 1975). Administrators and curriculum specialists expressed interest in continuing to use the NRS reading curriculum during the 1975-1976 school year in four first grade

---

[1] Where available, local press and TV reactions are cited. This is in part to demonstrate the dissemination of results and to document the obvious interplay between innovation, evaluation, and use of evaluative data.

classrooms in two District A schools, thus adding another ring to the implementation and evaluation from one school to two schools. The first concurrent contrast (as opposed to historical) was carried out the following year: Four NRS first grades were compared with seven basal first grades within the same school district. The NRS students again performed better than the students in the basal programs. Local response to the program was very enthusiastic (Wilkinsburg Gazette, 1976).

In the spring of 1976, District A personnel decided to use NRS as the main reading curriculum in their primary grades. The program was introduced in all first grade classrooms beginning in September, 1976. The program was also used in four second grade and three third grade classrooms the same year.

Accumulations of positive evidence about the benefits of NRS resulted in the adoption of the reading program by an LRDC-affiliated District B Follow Through site. During 1976-1977 NRS was implemented in two kindergarten Follow Through classrooms and in six first grade Follow Through classrooms. Evaluators also collected data from six non-NRS, non-Follow Through classrooms for comparison study, adding still another ring (multiple within-district comparisons) to the implementation and evaluation.

The developers of NRS saw a need for further evaluation of the program that would make comparisons outside of District A. Another district (District C) agreed to permit data to be gathered for use in such a contrast. Three first grade classrooms in District C acted as the control group for eight first grade classrooms in District A, and another ring, an across-district contrast, was added to the series of evaluations.

Another level of contrast was added to the evaluation of NRS in District B in 1977-1978. In addition to collecting data on six NRS

and six non-NRS first grade classrooms, data was gathered on six NRS and eight non-NRS second grade classrooms. Most second graders who had NRS in 1977-1978 had had NRS in 1976-1977 as first graders.

Specific Studies

Many studies of the effectiveness of NRS have been conducted over the years. A sample of seven of the twenty or more studies are reported here. All the studies chosen had reasonably clean data, pre- and posttests, and verifiable test administrations. No study was eliminated from this discussion due to directionality of the findings. This section presents the details of each selected contrast and shows how each study added to information about NRS. There is one within-school study, five within-district across-school studies, and one across-district study. In all cases there is pre/posttest information on students; in some cases there is descriptive classroom information from interviews or questionnaires.

Within-School Contrast: Study 1

After the pilot test of NRS described in the preceding section, the first of the series of evaluations, the within-school contrast, was carried out. Data collection and analysis for this study were under the control of the implementors (McCaslin, Note 4); raw data are no longer available for reanalysis.

Population. In 1974-1975, the NRS was implemented in three first grade classrooms in School 1 of District A. The population of the school was approximately 50% white. In 1973-1974, first graders at the school had received reading instruction with Scott-Foresman or Harper & Row, both basal reading series. These students constituted the contrast group. There were 40 students in the NRS group and 43 students in the basal group.

Method. Both groups received the Pintner-Cunningham Primary Test (Pintner, Cunningham, & Durost, 1966) in the spring of their kindergarten year. This test is one of a series of Pintner General Abilities Tests and yields an IQ score. In the spring of their first grade year, both groups were tested on four subtests of the Stanford Achievement Test (SAT; Madden, Gardner, Rudman, Karlsen, & Merwin, 1964): Word Reading, Paragraph Meaning, Vocabulary, and Word Study Skills.

Analysis and results. Scores for the Pintner-Cunningham did not differ significantly between the two groups. T-tests were conducted to determine if the NRS group performed significantly better than the basal group on the criterion measure. Scores on the SAT were significantly higher for the NRS group on two of the four subtests, Word Reading and Paragraph Meaning. Scores on the Word Study Skills and Vocabulary subtests were higher for the NRS group, but not significantly so.

Within-District Contrast: Study 2

In response to positive reading achievement test results from the within-school across-year contrast and to positive reactions by the administrators in District A, a study was designed to test the effectiveness of NRS in a more challenging, demanding set of circumstances--a concurrent within-district across-school contrast. This would help to control but not eliminate the challenge to results arising from possible historical differences from year 1 to year 2. This contrast represents a second ring in the evaluation of the effectiveness of NRS and its implementation.

Population. In 1975-1976, NRS was implemented in four first grade classrooms in two District A schools. The seven remaining first grade classrooms within the district that received basal instruc-

tion (Scott-Foresman or Harper and Row) were used as the contrast group. These seven classrooms were divided among three schools in the district. There were 66 students in the NRS group and 121 students in the non-NRS group who had pretest and posttest scores. The percentage of minority students enrolled in the first grade had increased slightly from the preceding year.

Method, analysis, and results. All students were tested in the fall with the Stanford Early School Achievement Test (SESAT; Madden & Gardner, 1969) and in the spring with the newly-modified Stanford Achievement Test (Madden, et al., 1973). The NRS group scored significantly better on the combination of two subtests of the SESAT measuring reading skills: Letters and Sounds, and Aural Comprehension. Subtests of the SAT used for the analysis were Reading: Part A, Reading: Part B, and Word Study Skills, the sum of which yields a Total Read score.

Table 1

Study 2: 1975-1976 Within-District Contrast

Achievement Means and Standard Deviations of NRS and non-NRS Groups: Matched Cases			
	NRS (66)	Non-NRS (121)	
Pretest: SESAT			
Sum of Letters and Sounds	45.82	41.24	$\bar{X}$
and Aural Comprehension	7.40	7.55	S.D.
Posttest: SAT			
Total Read: Sum of	108.42	84.79	$\bar{X}$
Reading Part A, Part B	32.19	30.27	S.D.
and Word Study Skills			

Correlation and Regression ($N = 187$)

		1	2	3
1)	SESAT Pretest	X		
2)	NRS (1)	.28*	X	
3)	SAT Total Read	.71*	.24*	X

Mult $R = .73^*$ SAT = $2.8^*X_1 + 10.7^*X_2$

*$p < .01$

Table 1 presents the means and standard deviations of the SESAT pretest and the SAT posttest for both the NRS and the non-NRS groups. Pretest scores indicate that the NRS group started out slightly ahead of the non-NRS group on the combination of the Letters & Sounds and Aural Comprehension subtests of the SESAT (4.6 points) and finished the year considerably higher on the posttest (23.6 points).

End-of-year achievement (SAT Total Read) was regressed on pretest (SESAT Letters and Sounds + Aural Comprehension) and a dummy variable code for NRS (1). Table 1 shows the correlation matrix of the variables in the regression, the multiple R, and the raw B weights for the regression solution. Both B weights (pretest and NRS) are significant (2.8 and 10.7), F (1,184) = 163.09 and 8.77 respectively.

One aid in interpreting the difference between the two groups in end-of-year achievement comes from the information about instructional processes used in reading. In January and February, 1976, an interview designed to measure instructional practices in reading was administered to the first grade teachers involved in the study. Data from the interview can be described using three general descriptors or constructs: the opportunity the students have to learn materials sampled on the end-of-year performance measures; the structure of the instruction for both the teacher and curriculum; and the content of reading, or how often students engaged in various types of reading activities. Only a selected number of variables from the constructs will be discussed here. A summary of these variables is presented in Table 2.

The first set of variables reflects the opportunity students had to learn the material on which they were tested. Important measures in this construct are: the number of students enrolled; the number of adults in the room during reading instruction; student attendance; the number of minutes allotted for reading per week; and the overlap

Table 2

Means and Standard Deviations of NRS and Non-NRS on Selected Measures

	Study 2		Study 3		Study 4		Study 5		Study 6	
	NRS (4)	SF* H&R** (7)	NRS (6)	HM*** (6)	NRS (6)	HM*** (6)	NRS (6)	HM*** (8)	NRS (8)	SF* Economy (3)
Opportunity										
Number of students enrolled	17.75	21.14	22.33	29.67	25.33	27.60	24.17	25.00	25.63	26.67
	(3.86)	(4.26)	(4.13)	(4.50)	(1.51)	(1.14)	(1.17)	(2.38)	(2.83)	(.58)
Number of adults in reading	1.25	1.00	2.00	1.00	2.00	1.00	2.17	1.00	1.13	1.00
	(.50)	(0)	(0)	(0)	(0)	(0)	(.41)	(0)	(.35)	(0)
Percent attendance	89.34	91.41	92.71	93.79	84.09	84.54	87.73	90.00	93.40	94.25
	(8.89)	(3.20)	(4.63)	(4.21)	(9.39)	(13.04)	(6.92)	(13.47)	(4.99)	(7.00)
Number of minutes/wk. in reading	472.50	369.71	500.00	575.00	450.00	362.50	453.33	523.29	516.88	750.00
	(133.32)	(162.54)	(77.46)	(73.49)	(0)	(274.24)	(53.17)	(282.78)	(180.46)	(150.00)
Percent overlap	75.66	48.92	64.23	42.41	NA	NA	NA	NA	57.36	60.18
	(5.91)	(10.61)	(17.76)	(24.03)					(16.42)	(26.40)
Structure										
Self-pacing	1.00	.29	1.00	.33	.83	.40	1.00	.29	1.00	0.00
	(0)	(.49)	(0)	(.52)	(.41)	(.55)	(0)	(.49)	(0)	(0)
Use of games	4.75	2.96	4.17	2.53	4.33	2.40	4.00	1.86	5.00	3.17
	(.50)	(2.12)	(1.33)	(1.96)	(1.63)	(1.56)	(1.67)	(1.95)	(0)	(2.02)
Speed of feedback	.75	.71	3.75	.50	3.92	2.80	3.42	2.00	3.75	3.33
	(.50)	(.49)	(.42)	(.84)	(.20)	(.91)	(.80)	(.41)	(.46)	(.29)
Percent unique assignments	91.98	16.55	85.98	14.23	80.95	16.51	98.61	13.65	89.87	10.83
	(5.92)	(8.77)	(14.42)	(6.71)	(10.31)	(2.01)	(3.40)	(3.89)	(5.06)	(.88)
Number of days since last test	5.03	86.36	4.05	22.74	9.05	12.43	6.57	50.51	2.29	46.76
	(1.48)	(60.57)	(2.61)	(17.79)	(3.28)	NA	(7.00)	(33.33)	(1.25)	(23.70)
Content										
Letter-sound correspondence	4.75	4.43	4.67	4.17	5.00	4.60	4.00	4.29	4.63	4.67
	(.50)	(1.13)	(.82)	(1.17)	(0)	(.55)	(1.10)	(.95)	(1.06)	(.58)
Letter name identification	5.00	4.43	4.67	3.83	4.83	3.20	3.33	3.43	4.13	4.67
	(0)	(.98)	(.82)	(1.17)	(.41)	(1.10)	(1.51)	(1.90)	(1.81)	(.58)
Blending	5.00	4.43	4.50	4.00	4.50	4.80	4.50	4.71	4.63	4.67
	(0)	(.98)	(1.23)	(1.10)	(.84)	(.45)	(.84)	(.49)	(1.06)	(.58)
Match between curriculum & objectives	4.50	3.57	4.92	4.50	4.67	4.20	4.50	3.50	4.50	2.50
	(.58)	(.98)	(.20)	(.84)	(.82)	(1.10)	(.84)	(1.05)	(.76)	(.50)

*Scott-Foresman **Harper & Row ***Houghton-Mifflin

between the curriculum and the criterion test. The average number of students enrolled in the NRS classrooms in District A during 1975-1976 was lower than the number enrolled in non-NRS classrooms (17.8 vs. 21.1). There were also more adults present for reading instruction in a few of the NRS classrooms. Attendance for the two groups was very similar (89.3% for NRS and 91.1% for non-NRS). NRS teachers assigned approximately 100 minutes per week more reading instruction to their students than non-NRS teachers, and they reported a higher overlap between what was presented in the curriculum and what was tested on the SAT (75.6% overlap reported by NRS teachers and 48.9% by non-NRS teachers). This means that, in general, the NRS students had a greater chance to learn reading than the non-NRS students because they had fewer students, more adults, more time, and a larger amount of criterion-relevant material.

The structure construct is represented by five variables: self-pacing of students, frequency of use of games, speed with which assignments are corrected and returned to students, the percentage of unique assignments, and the number of days since the last test. Without exception, in the NRS classrooms students worked at their own rate, while in non-NRS classrooms most students did not work at their own rate. Reading games were used more frequently in NRS classrooms. The rate of correction tended to be the same for NRS and non-NRS students. There were large differences between groups in the percentage of unique assignments and the number of days since the last test: percent unique was 92.0% for NRS and 16.6% for non-NRS, number of days since last test was 5.0 for NRS and 86.4 for non-NRS. Thus, NRS provides for greater individualization, more structure and more feedback.

The content construct is made up of four variables: frequency of student engagement in letter-sound correspondence, frequency of

letter identification, frequency of blending, and an estimate of the match between the curriculum and what the teacher considered to be important first grade reading objectives. NRS students were reported to have spent slightly more time in letter-sound correspondence, letter identification, and blending than non-NRS students. NRS teachers reported a closer match between NRS and important reading objectives, 4.5 on a 5-point scale; non-NRS teachers rated the match between their curriculum and objectives at 3.6. Thus, the content of the NRS curriculum seems to focus on acquiring basic reading skills more so than the basal sequence.

Within-District Contrast: Study 3

Population. A second within-district contrast was conducted in 1976-1977 when an urban school district (District B) from a different state than District A adopted NRS for use in six of its first grade Follow Through classrooms. The control group consisted of six first grade classrooms that received reading instruction with Houghton-Mifflin, a basal series. There were 119 students in the NRS group of six classes and 144 students in the non-NRS group for whom pre-test and posttest data were available. This study replicated Study 2 in a more economically depressed urban area. In addition, the pre/posttests were selected to more closely reflect the objectives of traditional basal series.

Method. Students in both groups were tested in October, 1976 with the Murphy-Durrell Reading Readiness Analysis (Murphy & Durrell, 1965). The test contains three subtests. Only the Letter Names subtest was used for analysis because performance on the Phonemes subtest was strongly influenced by the presence of NRS in the classroom prior to pretesting in October. The Learning Rate subtest was not used because of inconsistency in administration.

End-of-year achievement was measured in May, 1977 by the SAT (Madden, et al., 1973). Subtests used in the analysis were: Reading: Part A, Reading: Part B, and Word Study Skills. The three subtest scores yield a score for Total Read.

Analysis and results. Table 3 shows the means and standard deviations of the Murphy-Durrell pretest and SAT posttest for both the NRS and basal groups. The NRS group started out slightly ahead of the control group, possibly because testing of entering abilities did not occur until well over a month into the school year.

A regression analysis indicates that end-of-year achievement test scores for the NRS group were significantly higher than for the children who used the basal series. End-of-year achievement (SAT Total Read) for all students was regressed on initial ability and a dummy variable code for NRS (1). Table 3 shows the correlation matrix of the variables in the regression, with the raw B weights for the regression solution. Both B weights, pretest and NRS, were significant (2.1 and 10.6), F (1, 260) = 136.0, and 12.84.

Table 3

Study 3: 1976-1977 Within-District Contrast

Achievement Means and Standard Deviations of NRS and non-NRS Groups: Matched Cases			
	NRS (119)	Non-NRS (144)	
Pretest: Murphy-Durrell Letter Names Total	46.73	42.81	$\bar{X}$
	5.75	9.15	S.D.
Posttest: SAT Total Read: Sum of Reading Part A, Part B and Word Study Skills	102.79	83.81	$\bar{X}$
	27.78	29.09	S.D.

Correlation and Regression (N = 263)			
	1	2	3
1) Murphy-Durrell Pretest	X		
2) NRS (1)	.24*	X	
3) SAT Total Read	.62*	.32*	X
Mult R = .64*	SAT = $2.1^{*}X_1 + 10.6^{*}X_2$		

*$p < .01$

Process data for this contrast were collected in the first two months of 1977, again through the use of a teacher interview. These data are summarized in the second column of Table 2. The opportunity construct suggests that the average number of students enrolled in NRS classrooms in District B during 1976-1977 was lower ($\underline{X}$ = 22.3) than the number enrolled in the non-NRS classrooms ($\underline{X}$ = 29.7). NRS classrooms had twice as many adults present during reading instruction. The percentage of students present in both settings was equivalent. Teachers in NRS classrooms reported that they allocated 500 minutes per week for reading instruction; non-NRS teachers allocated 85 minutes per week more. NRS teachers reported a 64% overlap between what the SAT measured and what was presented in the NRS curriculum; non-NRS teachers reported a 42% overlap. Thus, in the second within-district contrast there is again a greater opportunity for students to learn to read in NRS classrooms because there are more teachers, fewer students, and greater overlap. However, less time is allocated to reading indicating that perhaps NRS is more efficient in teaching reading than the basal series.

Fundamental differences between NRS and non-NRS groups appear in the structure of the curriculum. Again the NRS group reported a higher level of student self-pacing, and that they use games about twice as frequently as non-NRS teachers. NRS teachers also reported that they correct student work and return it to the students much sooner than their basal counterparts, though neither group reported correcting or returning work on-line. The percentage of unique assignments on a given day is a measure of individualization in the classroom, and NRS classrooms have a much higher percentage of unique assignments: 86% for the NRS group and 14.2% for the non-NRS group. NRS students tend to be tested more frequently than non-NRS students: The average number of days between tests for

the NRS group is 4.1 days, and 22.7 days for the non-NRS group. NRS students also tend to score well on those tests. The percentage of correct items on a selected test for the NRS group was 95%; for the non-NRS group, it was about 75% correct. The results of the structure construct are consistent with the other within-district contrast: NRS individualizes instruction and gives more rapid and frequent feedback than non-NRS curricula.

Again NRS teachers report a somewhat greater focus in beginning reading content than non-NRS teachers. In general, NRS teachers report that they teach letter-sound correspondence more frequently than do non-NRS teachers (4.7 for NRS; 4.2 for non-NRS). The same pattern holds true for identifying letters and blending sounds. The rating of the match between the curriculum and what teachers considered to be important first grade reading objectives was high for both groups (4.9 for NRS, 4.5 for non-NRS).

Within-District Contrast: Study 4

Population. Data collection in District B during 1977-1978 formed the basis for the third within-district contrast. Six first grade classrooms received reading instruction with NRS, and six classrooms were taught with Houghton-Mifflin, a basal series. Pretest and posttest data were collected for 123 NRS students and 133 non-NRS students. This study replicates the previous one with the exception of a minor change of testing instruments.

Method. The Comprehensive Tests of Basic Skills (CTBS Level B, Form S; CTB/McGraw-Hill, 1973) was administered in October, 1977 in all 12 first grade classrooms. Five subtests were administered: Letter Sounds, Word Recognition I, Reading Comprehension, Word Recognition II, and Language I. A total read score was obtained by adding the first four subtest scores. The Language I subtest was not

included in the analyses because it is a test of listening and pictorial interpretation, not reading. Students were posttested in May, 1978 with the same form and subtests of the CTBS. Process data were also collected during the first months of 1978, this time through the use of a questionnaire mailed to the teachers.

Analysis and results. NRS and non-NRS group means and standard deviations for the CTBS are presented in Table 4. Again, the NRS group scored higher on the pretest than the non-NRS group. There is a 3-point difference between groups on the pretest and a 7-point difference on the posttest. End-of-year achievement (CTBS Total Read) for all students was regressed on initial ability and a dummy variable code for NRS (1). Table 4 contains correlations of the variables and raw B weights (pretest and NRS), both of which are significant (.77 and 4.7), F (1, 253) = 93.0 and 7.11.

Table 4

Study 4: 1977-1978 Within-District Contrast

Achievement Means and Standard Deviations of NRS and non-NRS Groups: Matched Cases

	NRS (123)	Non-NRS (133)	
Pretest: CTBS Level B			
Total Read	30.49	27.35	$\bar{X}$
	12.17	9.62	S.D.
Posttest: CTBS Level B			
Total Read	64.11	57.05	$\bar{X}$
	14.15	17.78	S.D.

Correlation and Regression (.V = 256)

	1	2	3
1) CTBS Level B Pretest	X		
2) NRS (1)	.14	X	
3) CTBS Level B Posttest	.53*	.21*	X

Mult R = .55* CTBS = $.77^{*}X_1 + 4.7^{*}X_2$

*$p < .01$

Classroom process data (in Table 2) for opportunity show that the NRS group again had fewer students enrolled per classroom than the non-NRS group and twice as many adults in the room during reading instruction. The number of students present in the classrooms was equivalent. First grade NRS teachers during 1977-1978 uniformly allotted 450 minutes per week for reading instruction, while non-NRS teachers allotted about 362 minutes per week. There exists a rather large standard deviation for the non-NRS group on this variable, indicating that, as a group, non-NRS teachers allocated reading time very differently. Unfortunately, the overlap estimate was not collected during 1977-1978 for this study.

The structure construct reveals that the NRS group had more classrooms in which students were self-paced. NRS teachers continued to report more frequent use of games and that they corrected and returned student work more quickly than teachers who used the basal series. The percentage of unique assignments was much higher for NRS classrooms (81% unique for NRS classrooms; 16.5% unique for non-NRS classrooms); students in NRS classrooms tended to be tested more frequently (calculations of the number of days between tests was possible for only two of the non-NRS classrooms); and they performed better on a selected test than did the non-NRS students.

Regarding the content construct, NRS students spent more time engaged in letter-sound correspondence and letter identification exercises, but non-NRS students spent more time in sound blending exercises. Both groups of teachers rated the curricula as having a close match with what they considered to be important first grade reading objectives.

Within-District Contrast: Study 5

Population. The fourth within-district contrast in this set of evaluations was a contrast at the second grade level in District B.

in 1977-1978, NRS was being implemented in six second grade classrooms. The control group was made up of eight second grade classrooms using Houghton-Mifflin, a basal reading series. There were 113 NRS students and 109 non-NRS students. About 70 percent of the students who were in NRS classrooms in second grade in District B participated as subjects in first grade classrooms in District B during 1976-1977. For the non-NRS group, only about 2% of those second grade students during 1977-1978 participated as first grade subjects during 1976-1977 (however, the curriculum, Houghton-Mifflin is implemented district-wide). This wide variation is due to the closing of one of the non-Follow Through schools. This study extends the previous two studies to the second grade--adding another ring to implementation and evaluation.

Method. Second grade students were pretested in October and posttested in May with Level C, Form S of the Comprehensive Tests of Basic Skills. Subtests administered were Vocabulary, Reading Comprehension: Sentences, and Reading Comprehension: Passages. The sum of the three subtests yields a score for Total Read. Classroom process data, obtained through a teacher questionnaire, were collected during the first few months of 1978.

Analysis and results. Pretest and posttest means and standard deviations from Level C of the CTBS are shown in Table 5. The NRS group scored slightly higher on both pretest and posttest.

A regression analysis was conducted to determine if the difference between groups on the dependent variable was significant. CTBS Total Read for both groups was regressed on pretest and a dummy variable code for NRS (1). Table 5 contains the correlation matrix of the variables in the regression, the multiple R, and the raw B weights for the regression solution. The B weight of .54 for NRS is significant, $F(1, 219) = .135$; however, the pretest .79 B weight, $F(1, 219) = 319$, is significant, as is the overall Mult R.

Table 5

Study 5: 1977-1978 Within-District Contrast

Achievement Means and Standard Deviations of NRS and non-NRS Groups: Matched Cases			
	NRS (113)	Non-NRS (109)	
Pretest: CTBS Level C			
Total Read	34.42	32.92	$\bar{X}$
	16.56	16.71	S.D.
Posttest: CTBS Level C			
Total Read	50.82	49.09	$\bar{X}$
	16.40	17.89	S.D.
Correlation and Regression (N = 222)			
	1	2	3
1) CTBS Level C Pretest	X		
2) NRS (1)	.045	X	
3) CTBS Level C Posttest	.77*	.051	X
Mult R .77*	CTBS Posttest = $.79^{*}X_1 + .54X_2$		

*$p < .01$

Within-District Across-Grade Contrast: Study 5a

Population. Another contrast arose from the second grade study. This analysis was done to determine if there were cumulative effects that were undetected by the simple second grade contrast. This contrast involves tracking NRS students over two years. By spring, 1978, there were 112 students who had been in the District B studies since the fall, 1976: 83 students in NRS and 29 students in the basal program. Clearly the contrast group must be viewed with caution, however, because the means for the full first grade group and the two-year sample are quite similar as shown in Table 6.

Method. Achievement test scores for this contrast were compiled from the first grade fall 1976 administration of the Letter Names subtest of the Murphy-Durrell and the spring 1978 administration of the CTBS Level C to second graders.

Table 6

Study 5a: 1976-1977 and 1977-1978 Within-District and Across-Grade Contrast

Achievement Means and Standard Deviations of NRS and non-NRS Groups: Matched Cases

	NRS (83)	Non-NRS (29)	
Pretest: Murphy-Durrell			
Letter Names Total	47.78	46.00	$\bar{X}$
	4.07	3.79	S.D.
Posttest: CTBS Level C			
Total Read	52.33	49.83	$\bar{X}$
	15.72	17.23	S.D.

Achievement Means and Standard Deviations of NRS and non-NRS Groups: Unmatched Cases

	NRS	Non-NRS	
Pretest: Murphy-Durrell			
Letter Names Total	45.97 (n = 140)	42.02 (n = 172)	$\bar{X}$
	6.97	9.44	S.D.
Posttest: CTBS Level C			
Total Read	50.82 (n = 113)	49.09 (n = 109)	$\bar{X}$
	16.40	17.89	S.D.

Correlation and Regression (N = 112)

		1	2	3
1)	Murphy-Durrell Pretest	X		
2)	NRS (1)	.19	X	
3)	CTBS Level C Posttest	.29*	.07	X

Mult R = .29* CTBS Posttest = $1.1^{*}X_1 + .50X_2$

*$p < .01$

Analysis and results. Table 6 shows group means and standard deviations of the pretest and posttest measures. The NRS group started out slightly ahead of the basal group in first grade and ended up only slightly ahead at the end of second grade. The regression analysis used a dummy variable for NRS (1). Table 6 presents the correlation coefficients, the multiple R, and the raw B weights for the regression solution. Again the B weight for NRS is not significant but significance was obtained for pretest, F (1, 109) = 9.16, as well as the overall Mult R.

Process data for the grade two, District B classrooms were collected during 1977-1978 via a questionnaire completed by the teachers. These data are identical for Studies 5 and 5a. The opportunity and structure information is similar to previous years. In terms of opportunity, NRS had fewer students, more adults, equivalent attendance, and lower allocation of reading time. Variables within the structure construct reveal that in NRS classrooms, students were more often self-paced, used more games, had faster feedback, had individualized assignments, were tested more frequently, and did better on teacher tests.

By second grade, some variables in the content construct have taken on a different look as would be expected. For example, second grade NRS students engaged in letter-sound correspondence, letter identification, and sound blending less often than the basal students and less often than they did in first grade. This can probably be explained by the fact that, in NRS, most of those kinds of exercises are taught in the first few levels, usually covered in the first grade. Second grade NRS teachers rate the curriculum as closely matching what they consider to be important second grade reading objectives (4.5), while non-NRS teachers rate their curriculum less well (3.5).

Across-District Contrast: Study 6

The final study to be presented here involves a contrast between two school districts: District A, which was using NRS in first grade, and District C, using Scott-Foresman or Economy, both of which are basal reading curricula. The study took place during 1976-1977 and represents the third ring in the set of contrasts. District A was implementing NRS in all its first grade classrooms that year; as described earlier, a second district was identified for use as a comparable contrast group. This contrast provides a more rigorous control for the effect of the innovation. Within District A, NRS was clearly identified as the innovation, and it was possible that the non-NRS teachers were distressed at not having it. District C teachers, in contrast, did not know of the existence of NRS.

Population. Census tract data from 1970 revealed similarities between District C and District A in terms of SES and race. By early 1976, when plans were being drawn up to establish District C as a control, the population in District A had begun to change. Three classrooms in two schools in District C closely resembled those in District A with regard to SES and pretest information, so they were used to represent the contrast group for the eight District A classrooms.

Method. Both groups were tested in the fall, 1976 with the Murphy-Durrell Reading Readiness Analysis, and in the spring, 1977 with the Stanford Achievement Test. The Letter Names subtest from the Murphy-Durrell was used as the pretest measure and the Total Read score from the SAT was the criterion measure.

Analysis and results. Table 7 shows means and standard deviations for the groups on both pre- and posttests. NRS students start out slightly below the basal students, but end up well ahead of them. The NRS group scored 2 points below the non-NRS group on the pretest, but 12.3 points above the control group on the posttest. End-

of-year achievement for both groups was regressed on pretest and a dummy variable code for NRS (1). Table 7 shows the correlation matrix of the variables in the regression, the multiple R, and the raw B weights for the regression solution. Both B weights (pretest and NRS, 2.1 and 16.5) are significant, F (1, 238) = 77.42 and 25.04 respectively.

Data regarding classroom practices were collected with an interview of the eight NRS District A teachers and the three comparison District C non-NRS teachers. These data are presented in the last column of Table 2. Variables in the opportunity construct show that District A classrooms had an equivalent number of students enrolled, adults in reading, and percent of students present. District A had fewer minutes allocated to reading (200 minutes less per week), and slightly lower percentage of overlap.

Table 7

Study 6: 1976-1977 Across-District Contrast

Achievement Means and Standard Deviations of NRS and non-NRS Groups: Matched Cases			
	NRS (170)	Non-NRS (71)	
Pretest: Murphy-Durrell Letter Names Total	44.57	46.61	$\bar{X}$
	6.49	6.12	S.D.
Posttest: SAT Total Read	100.89	88.61	$\bar{X}$
	26.60	26.28	S.D.
Correlation and Regression (N = 241)			
	1	2	3
1) Murphy-Durrell pretest	X		
2) NRS (1)	−.14	X	
3) SAT Posttest	.45*	.21*	X
Mult R = .53*	SAT = 2.1*X_1 + 16.5*X_2		

*$p < .01$

In the structure construct, NRS teachers unanimously reported that their classrooms were self-paced and that they used more games. NRS students got work corrected and returned to them more quickly, went for far fewer days between tests, and performed slightly better on the tests, though the difference between groups was not dramatic. Most of the students in the District C group were assigned the same lessons as their classmates, while 90 percent of the students in District A were assigned unique lessons.

The content construct reveals that the NRS group spent the same amount of time engaging in letter-sound correspondence, slightly less time identifying letters, and the same amount of time blending. District A teachers rated the match between NRS and what they thought to be important first grade reading objectives as 4.5, while District C teachers rated the match as 2.5.

Discussion and Implications

The effects of NRS were assessed through several related evaluations. Each new study contributed more information. The contrasts can be considered concentric rings expanding out to more distant settings from the central starting point. Initial settings were geographically close, involved considerable support during the implementation process, and used criterion measures that had a tight relationship to what was taught. Later contrasts used settings that were further away, involved less support, and used more general criterion measures.

After the initial field test of NRS in LRDC-affiliated developmental schools, negotiations to implement the new curriculum in a school unrelated to LRDC were completed, and the first within-school contrast was established. Concerns of developers and school administrators led the program into further implementation within that school

district, and the series of within-district contrasts was begun. The success of NRS in those two settings led to the more complex contrast across districts. The positive results of most of the studies conducted under very different, increasingly demanding circumstances, with less and less implementor assistance, speak well of the iterative, concentric approach to implementation and evaluation.

The data support NRS as a successful first grade reading program. Four studies show NRS students performing considerably better than their counterparts using a variety of basal series (Scott-Foresman, Harper and Row, Houghton-Mifflin, and Economy). One study of second grade students in one district did not show NRS students to have gained more than non-NRS students. We analyzed the second grade data in two ways: first, a simple pre/post contrast; second, a longitudinal study of those students that were in either program for two years. This second contrast, while obviously suffering from selection bias, was important in case we were observing students that had had only one year of NRS. One possibility is that many of the students in NRS in the second grade had moved out of NRS and were in a different curriculum. However, this is only speculation. Further research is clearly needed to evaluate the effectiveness of NRS at the second grade. Five sets of process data were collected over the years. General trends noted from the data are summarized.

Opportunity. NRS classrooms tend to have fewer students enrolled and more adults present during reading instruction. However, in Study 2, the number of adults is almost equal. The percentage of students present tends to be equivalent for NRS and comparison groups. The number of minutes per week allocated for reading instruction varies. It appears that even when NRS students have less time available, they usually perform better. When the overlap estimate between what was taught and what was tested on the end-of-year achievement

measure was collected, the NRS group had a higher percentage of overlap in two out of three cases.

Structure. It is evident that NRS students were self-paced and used games as a part of reading instruction far more frequently than their non-NRS counterparts. NRS teachers corrected and returned work to students more quickly and made assignments on a more individualized basis. NRS students were tested more frequently and performed better on those tests than did the basal students, indicating that they were tested on material they had learned.

Content. First grade NRS students spent more time than basal students and second grade NRS students engaging in letter-sound correspondence, letter identification, and blending. This trend is most probably a function of the fact that the NRS curriculum presents those kinds of activities early in the first grade. NRS teachers, across all of the studies, rated the match between the curriculum and what they considered to be important first or second grade reading objectives as a close match.

This paper used a set of sequential studies on the effectiveness of NRS as an example of a strategy for curriculum evaluation. These studies follow the natural development and expansion of a curriculum program and provide convergent evidence for the value of a particular innovation, NRS. None of the studies alone could provide convincing evidence for the value of the curriculum, and even this collection of studies still leaves questions unanswered. We do know that the program can be implemented successfully in several very different settings with decreasing amounts of developer support. We can also see that NRS is very successful in first grade classrooms on several different dependent measures (two versions of the SAT and the CTBS). However, the one study of NRS at the second grade did not provide evidence for its success there. The strategy of cumulating several small studies

in multiple settings is clearly feasible and follows the natural sequence of development and implementation.

The strategy of iterative evaluation is not offered as a replacement for the two more prominent approaches currently available, experiments and explanatory observational studies. Rather, iterative evaluation is seen as an alternative to waiting for the perfect evaluation conditions to prevail or to ignoring the natural production of evaluative information that accompanies program development. Whether it is a desirable alternative remains to be seen.

Reference Notes

1. Beck, [illegible] The New Primary Grades Reading System: Preview of an adaptive reading system. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, 1973.

2. Eichelberger, R. T., & Lee, J. Formative evaluation design for the New Reading System (NRS): 1973-74. Unpublished manuscript, University of Pittsburgh, Learning Research and Development Center, 1974.

3. Buehler, J. F., & Eichelberger, R. T. Evaluation of an adaptive reading curriculum from a longitudinal perspective. Paper presented at the annual meeting of the American Educational Research Association, Washington, D. C., 1975.

4. McCaslin, E. S. Report of the first year implementation of the New Primary Grades Reading System (NRS) in Kelly School (1974-1975 school year). University of Pittsburgh, undated draft.

References

Beck, I. L. Comprehension during the acquisition of decoding skills. In J. T. Guthrie (Ed.), Cognition, curriculum, and comprehension. Newark, DE: International Reading Association, 1977.

Boruch, R. F., & Rindskopf, D. On randomized experiments, approximation to experiments, and data analysis. In L. Rutman (Ed.), Evaluation research methods: A basic guide. Beverly Hills, CA: Sage, 1977.

Cooley, W. W. Explanatory observational studies. Educational Researcher, 1978, 7 (9), 9-15.

CTB/McGraw-Hill. Comprehensive Tests of Basic Skills. Monterey, CA: McGraw-Hill, 1973.

Madden, R., & Gardner, E. F. Stanford Early School Achievement Test. Level I. New York: Harcourt, Brace, Jovanovich, 1969.

Madden, R., Gardner, E. F., Rudman, H. C., Karlsen, B., & Merwin, J. C. Stanford Achievement Test. Primary I Form A. New York: Harcourt, Brace, Jovanovich, 1964; 1973.

Murphy, H. A., & Durrell, D. D. Murphy-Durrell Reading Readiness Analysis. New York: Harcourt, Brace, & World, 1965.

Pintner, R., Cunningham, B. V., & Durost, W. Pintner-Cunningham Primary Test. New York: Harcourt, Brace, & World, 1966.

Wilkinsburg Gazette, August 28, 1974.

Wilkinsburg Gazette, October 15, 1975.

Wilkinsburg Gazette, March 17, 1976.