DOCUMENT RESUME

ED 152 766 SP 012 568

AUTHOR McDonald, Frederick J.
TITLE Research as Evaluation, Evaluation as Research.
INSTITUTION Educational Testing Service, Princeton, N.J.
PUB DATE Feb 78
NOTE 32p.; Paper presented at the Annual Meeting, American Association of Colleges for Teacher Education (Chicago, Illinois, February 21-24, 1978)

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS *Data Analysis; Data Processing; Educational Research; *Evaluation; *Research; *Research Design; Research Methodology; *Research Skills
IDENTIFIERS *Research as a Field of Study

ABSTRACT

The sharply drawn distinction between research and evaluation, and the equally sharp distinction between the carefully controlled laboratory-type experiment and the study occurring in the field, are seen as arbitrary and not very helpful. In illustrating this thesis, the author develops a design generally regarded as a "research" design and shows the evaluative purposes for which the design may be used. A Teacher Corps/Trenton (New Jersey) State College/Trenton Public School inservice program for the improvement of reading instruction skills was chosen as the subject of the study. Observational data on teacher change due to training, effectiveness of varying training programs, and the relationship between change/level of teacher skills and pupil gains in reading are presented. The data, essentially of an evaluative nature, is extensively interpreted to illustrate the contention that researchers are also learning about the nature of teaching and the structure of teaching performances. The evaluative data also yield information on the relationship between kinds of teacher performances and pupil learning. Through this interpretation, the author illustrates that by carefully designing a study, one can achieve two goals: first, one can develop an understanding about teaching performances related to pupil learning; and two, the effectiveness of ideas about what constitutes the appropriate substance of training and whether or not the chosen methods are effective can be evaluated. (MJB)

RESEARCH AS EVALUATION, EVALUATION AS RESEARCH

Frederick J. McDonald
Educational Testing Service
Princeton, New Jersey

Presentation to Strand 8:
Evaluation of Teaching and Careers
AACTE Conference
Chicago, Illinois
February 21-24, 1978

Introduction

Sometimes a distinction is made between research and evaluation. In some people's minds research is a set of activities designed to develop theories and to test hypotheses. Evaluation, in contrast, is thought of as a set of activities designed to make a judgment about the effectiveness of an educational practice or program.

The two activities comprise the same kinds of activities; for example, they include measurements of various kinds and statistical analyses of these measurements. The distinction between them, however, is not necessarily a distinction between the use of an experimental design and the use of naturalistic field settings. Insofar as a distinction is made, it seems to be in terms of the purpose of the activity. The similarities in the activities, however, suggest that in some cases, for example, research on teaching, the distinction is not helpful.

An evaluation of a program necessarily includes theoretical considerations; for example, the designer of the program usually has some theoretical conception on which the program is based, so that the evaluation of the program is a more or less direct test of the theory. Suppose that a teacher educator designs two different kinds of teacher-training programs. They may be different applications of a particular theory. If they are, evaluation estimates the effectiveness of one application in contrast to the other. Or they may be applications of two different theories, in which case the respective theories as applied are being tested.

Research studies in teacher education, usually are designed to estimate the effectiveness of a particular training program or teaching practice. If a particular method of training a teaching skill, for example, is found to be more effective than another method, then that method ought to be used in practice.

Another distinction that is made is that research, particularly when it is experimental in character, is designed for hypothesis-testing, whereas, evaluation is designed for decision-making. But in an applied field, it seems that this distinction can be too sharply drawn. In practice the research is usually intended to lead to a practical application more or less directly, and as suggested above, an evaluation in some sense of the word tests a theoretical conception either of training or of effective teaching.

Can the two styles be combined? The major purpose of this paper is to describe the use of experimental research as an evaluation technique. I will report one attempt to conduct such a study, and point out what it is that can be learned by conducting such a study about a theory of effective teaching and about training, and about effective teaching.

Experimental Evaluation Designs

The Teacher Corps program supported a project at Trenton State College conducted in conjunction with the Trenton Public Schools which will illustrate the use of an experimental design as an evaluation design. I will point out later how this particular design leads both to decision-making about training and to enlarged understanding of theories of effective teaching.

In this particular study a program of in-service training was developed to improve teachers' skills in the teaching of reading. The skills on which the teachers were to be trained were derived from two sources: (1) from existing research on effective teaching; and (2) conceptions of effective teaching derived from theory and practical experience.

Data from the Phase II of the Beginning Teacher Evaluation Study was the research-source used to develop a conception of the skills required for the effective teaching of reading (McDonald and Elias, 1976). In that study it was found that teachers who applied a model of direct instruction more frequently were teachers who produced greater gains in pupil learning. The model of direct instruction included such teaching practices and strategies as maintaining on-task, productive pupil behavior; interacting more frequently and for longer periods of time with each pupil; increasing the amount of instructional time available for each child, and using a form of social control which essentially redirected the child to the work at hand. The model of direct instruction was the tutorial model applied in the context of teaching 20 to 30 children.

Ideas on effective teaching practice were also derived from the literature on the teaching of reading and from practical experience. These ideas, developed by faculty members of Trenton State College, were concepts of teaching skills in assessing and diagnosing reading difficulties, and in developing reading comprehension.

These ideas were used in the first year of the Trenton program to develop a series of training modules. The first three modules were designed to increase teachers' ability to recognize when pupils were off-task and nonproductively engaged and to increase their ability to redirect pupils to on-task work. The

fourth and fifth modules in this series were devoted to training teachers on the skills of assessing and diagnosing reading difficulties. In Phase II of the project, the second year, three modules were designed to teach teachers skills to improve pupils' reading comprehension abilities.

Each module consisted of three components: (1) a knowledge component which presented the concept of the skill and the rationale for it; (2) classroom practice on the use of the skill; and (3) feedback on classroom performance.

An experimental design was used in this training program. The teachers of the staff of one school were assigned randomly to three different training modalities in the knowledge-component of the training plan. The practice and feedback component were essentially the same for all teachers. The three variations on the presentation of the knowledge-component consisted of presenting the same substance in three different forms, through a picture book, through slides, and through a videotape presentation. In the second year only the latter two of these modes of presentation were used.

The teachers were observed regularly and daily throughout the program. They were observed before the actual training began. They were observed during each module, that is, during the part in which the teachers received the formal training and for a period of time after the feedback sessions..

The reading skills of the pupils of the teachers were measured at three times, first in the fall of the year, again in the winter and in the spring.

This design enabled us to answer three kinds of questions. First, did the teachers change as a result of the training? Second, was any one of the training procedures more effective than any of the others? Third, was there a relation between the change in teachers' skills or the level of the teachers' skills and pupil gains in reading?

The observational data were used to determine whether or not there was an effect on teacher performance. I present here a number of tables which demonstrate the degree to which a change has occurred. As I read the tables, I will simply note some kinds of changes that occurred and did not occur. Refer to Table 1. Notice how the table is constructed. First of all the table is divided by particular types of skills. These skills were derived by an analysis of what was to be learned and from the information provided by the observational system. The observational system gathered information on more aspects of teaching behavior than were the objects of training; but we considered among these, those that would be expected to be changed by the training, and other behaviors that might be indirectly or directly affected by the training.

Notice that the first set of variables refers to the size of the group being taught. This variable was studied because we had found in previous research that, if pupils worked by themselves, there was an increase in the amount of off-task, nonproductive behavior. So the training was designed to increase the extent to which the teacher worked directly with children either by rearranging them in groups with which she or he worked or by moving around the room so that he or she maintained contact with children working by themselves.

In the left-hand column under each variable are some designations, such as B, B+1, and so forth. These designations refer to various starting points; for example, B, refers to the calculations made from the baseline, B+1 is the baseline data plus the data from the first module. It is possible to tell by reading across in any line how much of an increase in a particular skill had been made at the end of each module.

Table 1

The Effects of the Training Modules on Teaching Performance: Frequency by Module

		Pre	1	2	3	4	Post 5
Group Size: N (Individual Student)							
	B	.07	.07	.00	.00	.00	.02
	B+1	.02		.00	.04	.03	.02
	B+1+2	.02			.04	.03	.02
	B+1+2+3	.06				.03	.02
	B+1+2+3+4	.05					.02
Group Size: L+S (Small Group)							
	B	.66	.68	.72	.68	.74	.70
	B+1	.71		.71	.64	.73	.68
	B+1+2	.68			.64	.73	.68
	B+1+2+3	.63				.73	.70
	B+1+2+3+4	.66					.70
Group Size: W+M (Large Group)							
	B	.26	.26	.27	.32	.24	.28
	B+1	.26		.27	.32	.23	.30
	B+1+2	.29			.32	.23	.30
	B+1+2+3	.30				.23	.28
	B+1+2+3+4	.29					.28

Table 1 (Continued)

		Pre	1	2	3	4	5
					Post		
Role: A (Assess/Diagnose)							
	B	.05	.08	.14	.09	.16	.20
	B+1	.05		.13	.11	.15	.21
	B+1+2	.08			.11	.15	.21
	B+1+2+3	.10				.14	.20
	B+1+2+3+4	.11					.20
Role: D (Discipline)							
	B	.02	.01	.03	.00	.00	.00
	B+1	.01		.03	.00	.00	.00
	B+1+2	.01			.00	.00	.00
	B+1+2+3	.02				.00	.00
	B+1+2+3+4	.01					.00
Role: N (Instruction)							
	B	.55	.45	.24*	.46	.36	.32
	B+1	.46		.25	.45	.36	.30
	B+1+2	.40			.45	.36	.30
	B+1+2+3	.41				.36	.32
	B+1+2+3+4	.40					.32
Role: F (Facilitate)							
	B	.16	.24	.24	.25	.22	.25
	B+1	.24		.25	.24	.24	.27
	B+1+2	.25			.24	.24	.27
	B+1+2+3	.23				.23	.27
	B+1+2+3+4	.23					.27

Table 1 (Continued)

		Pre	1	2	3	4	5
				Post			
Role: M (Manage)							
	B	.17	.21	.24	.13	.11*	.12*
	B+1	.20		.23	.13	.11*	.12*
	B+1+2	.20			.13	.11*	.12*
	B+1+2+3	.18				.11	.12
	B+1+2+3+4	.16					.12
Role: T+S (Independent and Supervise Staff)							
	B	.00	.00	.02	.03	.03	.00
	B+1	.00		.02	.02	.04	.00
	B+1+2	.01			.02	.04	.00
	B+1+2+3	.01			.02	.04	.00*
	B+1+2+3+4	.02					.00
Role: A+N (Assess/Diagnose and Instruction)							
	B	.59	.53	.38*	.55	.51	.52
	B+1	.51		.38	.56	.50	.51
	B+1+2	.47			.56	.50	.51
	B+1+2+3	.50				.51	.52
	B+1+2+3+4	.51					.52
Role: A+N+F (Assess/Diagnose and Instruction and Facilitate)							
	B	.78	.76	.62	.80	.74	.78
	B+1	.75		.63	.80*	.74	.78
	B+1+2	.72			.80	.74	.78
	B+1+2+3	.74				.74	.79
	B+1+2+3+4	.74					.79

Table 1 (Continued)

		Pre	Post 1	2	3	4	5
Mobility: S (Stationary)							
	B	.62	.70	.52	.48	.71	.75*
	B+1	.58		.53	.49	.73	.71*
	B+1+2	.58			.49	.73*	.71*
	B+1+2+3	.56				.74*	.73*
	B+1+2+3+4	.60					.73
Mobility: L+M (Moving)							
	B	.38	.30	.45	.51	.29	.24*
	B+1	.41		.44	.50	.27	.29
	B+1+2	.41			.50	.27	.29*
	B+1+2+3	.43				.26	.27
	B+1+2+3+4	.39					.27
Mobility: Number of Moves (Number of time teacher changes groups)							
	B	.23	.16	.24	.18	.12	.18
	B+1	.20		.23	.18	.12	.19
	B+1+2	.22			.18	.12*	.19
	B+1+2+3	.21				.12	.18
	B+1+2+3+4	.19					.18

Table 1 (Continued)

		Pre		Post 1	2	3	4	5
Feedback Sign: BT (Both Positive and Negative Task Specific)								
	B	.28	.29	.28	.25	.31	.23	
	B+1	.24		.27	.26	.30	.22	
	B+1+2	.26			.26	.30	.22	
	B+1+2+3	.25				.28	.25	
	B+1+2+3+4	.26					.25	
Feedback Sign: +T (Positive Task Specific)								
	B	.26	.33	.27	.30	.24	.17	
	B+1	.28		.30	.28	.22	.17	
	B+1+2	.30			.28	.22	.17	
	B+1+2+3	.26				.25	.17	
	B+1+2+3+4	.26					.17	
Feedback Sign: BC (Both Positive and Negative Corrective)								
	B	.12	.03	.05	.07	.00	.11	
	B+1	.08		.05	.09	.02	.10	
	B+1+2	.06			.09	.02	.10	
	B+1+2+3	.07				.02	.10	
	B+1+2+3+4	.06					.10	
Feedback Sign: +C (Positive Corrective)								
	B	.00	.05	.01	.06	.00	.01	
	B+1	.07		.01	.07	.00	.01	
	B+1+2	.03			.07	.00*	.01	
	B+1+2+3	.04				.00	.01	
	B+1+2+3+4	.03					.01	

Table 1 (Continued)

		Pre	Post 1	2	3	4	5
General Skills							
	B	.59	.43	.68	.64	.48	.44
	B+1	.56		.64	.62	.48	.45
	B+1+2	.57			.62	.48	.45
	B+1+2+3	.56				.47	.46
	B+1+2+3+4	.54					.46
Phonics Skills							
	B	.21	.20	.21	.09	.32	.24
	B+1	.16		.19	.10	.31	.26
	B+1+2	.18			.10	.31	.26
	B+1+2+3	.15				.31	.24
	B+1+2+3+4	.20					.24
Vocabulary Skills							
	B	.20	.33	.21	.22	.20	.11*
	B+1	.29		.19	.22	.18	.13
	B+1+2	.24			.22	.18	.13
	B+1+2+3	.22				.19	.13
	B+1+2+3+4	.22					.13

Table 1 (Continued)

		Pre	Post 1	2	3	4	5
Grammar Skills							
	B	.08	.11	.04	.05	.04	.08
	B+1	.10		.03	.05	.04	.07
	B+1+2	.07			.05	.04	.08
	B+1+2+3	.07				.03	.10
	B+1+2+3+4	.06					.10
Comprehension Skills							
	B	.27	.29	.23	.40	.25	.49
	B+1	.25		.23	.38	.24	.46
	B+1+2	.25			.38	.24	.46
	B+1+2+3	.29				.25	.43
	B+1+2+3+4	.28					.43
Interpreting Skills							
	B	.20	.20	.11	.17	.21	.17
	B+1	.17		.14	.16	.23	.17
	B+1+2	.18			.16	.23	.17
	B+1+2+3	.18				.22	.18
	B+1+2+3+4	.18					.18

Table 1 (Continued)

		Pre	1	2	Post 3	4	5
Critical Judgment Skills (Not used in Phase I)							
	B						
	B+1						
	B+1+2						
	B+1+2+3						
Material: BR (Basal Reader)							
	B	.03	.00	.03	.00	.01	.04
	B+1	.01		.03	.00	.02	.05
	B+1+2	.02			.00	.02	.05
	B+1+2+3	.02				.02	.05
	B+1+2+3+4	.02					.05
Material: BW (Basal Workbook)							
	B	.12	.09	.05	.01*	.03	.00*
	B+1	.08		.04	.01*	.02	.00
	B+1+2	.05			.01	.02	.00
	B+1+2+3	.03				.02	.00
	B+1+2+3+4	.03					.00
Material: BT (Basal Test)							
	B	.28	.29	.28	.25	.32	.23
	B+1	.24		.27	.26	.32	.22
	B+1+2	.26			.26	.32	.22
	B+1+2+3	.25				.30	.25
	B+1+2+3+4	.26					.25

Table 1 (Continued)

		Post				
	Pre	1	2	3	4	5
Material: QT (Teacher Questions and Explanation)						
B	.20	.31	.27	.44*	.42*	.51*
B+1	.24		.27	.42*	.41*	.49*
B+1+2	.26			.42*	.41*	.49*
B+1+2+3	.30				.44	.48*
B+1+2+3+4	.35					.48
Material: AM (Art Materials)						
B	.01	.02	.00	.00	.01	.00
B+1	.01		.00	.00	.01	.00
B+1+2	.01			.00	.01	.00
B+1+2+3	.01				.01	.00
B+1+2+3+4	.01					.00
Material: GA (Game)						
B						
B+1						
B+1+2						
B+1+2+3						
B+1+2+3+4						

Across the top are designations "Pre," 1, 2, and so forth. "Pre" stands for the data gathered during the period prior to actual training. The numbers 1, 2, etc., refer to the numbers of the modules.

The table is read as follows. Under Group Size: N if you read across from "B" and under "Pre," you will see the number .07. This number is the percentage of time that the teacher had children working by themselves during the period prior to immediate training. If you now move across the same line to under 1, you will notice the number .07, or 7 percent. There has been no change in the grouping of pupils by themselves from the period of the pretraining to the end of the first module. The first module did not affect grouping practices. But notice from this point on that there is a rapid decrease in the use of this particular technique. The change here is not statistically significant, but the original percentages are so small that these differences are worth noting.

If you read down the first column under "Pre," you can see the degree to which the teachers use any one of the practices; in other words, you obtain a picture of typical practice prior to training. We have noted that 7 percent of the time the students were working by themselves; 66 percent of the time they were working in small groups, and 26 percent of the time they were working in a large group or with the class-as-a-whole. These relative proportions do not change across the training period except that there is an increase in the use of small groups, a decrease in students working alone, and relatively little change in the frequency with which large-group instruction is used.

Let us consider another example. Move over now to the portion of the table in which the various teaching roles are described. Notice that by reading down under "Pre" that the two roles, Assess/Diagnose (A) and Discipline (D)

are not roles in which the teachers engage frequently. In contrast most of the teachers' role activity is directed to Instruction 55 percent of the time, 16 percent to Facilitating, and 17 percent to Managing.

The object of the training was not to affect these roles directly, but the nature of the training was such that one would expect some changes in the various roles. What is of interest in this research and evaluation is to determine the extent to which these roles change in some significant way.

One would not expect Disciplining to change markedly; however, we did expect a change in the Assess/Diagnose role since instruction was given on assessing and diagnosing in Modules 4 and 5. We note an increase, not statistically significant but substantial, in the context of the size of these figures, from 5 percent of the time to 16 and 20 percent of the time, respectively, after Modules 4 and 5, the two modules in which training was given on assessing and diagnosing.

The nature of teaching is such that a change in one role necessarily affects another role. The change here has occurred at the expense of direct instruction. Notice that there has been a drop from 55 percent to 32 percent by the end of Module 5.

These data illustrate facts relevant to understanding the nature of teaching activity. It is obvious that a modification in one aspect of teaching necessarily requires a modification in some other aspect of teaching. Here we see an increase in Assessing/Diagnosing and a decrease in Direct Instruction. Teaching behavior is so constituted structurally that changes of these kinds are inevitable, and the goal of research ought to be to determine how to bring about desirable changes in one aspect of teaching without necessarily sacrificing other desirable aspects of teaching.

You may question whether an increase in Assessing/Diagnosing at the expense of Instructing is necessarily undesirable. From the data available we cannot tell. The teachers may have improved their Assessing/Diagnosing, which in turn may have improved the quality of their instruction. But it is difficult to see how a decrease in the amount of instructing would necessarily improve learning, and in fact we have data to indicate that a decrease in instruction does affect pupil learning negatively.

I said earlier that research of the kind that I am describing here, which is also evaluative research, enables us both to theorize about the nature of teaching as well as to evaluate the effect of the training program. The data that I am showing you is, strictly speaking, evaluative data; it is data that tells us whether the modules have had the effect that was desired. The fact that some of the modules do and some of the modules do not leads us to think about how the program might be redesigned to be more effective.

But at the same time, we are learning something about the nature of teaching which enables us to think more precisely and more comprehensively about its character. We see the intercorrelation among these behaviors and recognize that we are working with a structure of teaching performances. It does not seem possible simply to add a new skill or increase the use of existing skills without affecting this structure in some substantial way. This fact leads me to believe that our approach to understanding teaching activity must be one in which we attempt to identify these different structures. If we wish to modify these structures, we have to think about processes for modifying a structure, rather than processes for adding or subtracting performances in a repertoire of performances.

I present now another set of data which is directly related to research goals. The purpose of research on teaching is to determine the relation between various teaching performances and pupil learning. In this particular study an attempt was made to increase the amount of performance of certain kinds of skills, and these performances necessarily will be related more or less directly to increases in pupil learning.

For the analysis which I am about to describe we selected nine variables which we thought to be critical. In Table 2 you have a list of these variables. Note that the variables which are designated by letters L+S, A, and so on, occur three times in the upper, middle, and lower portions of the table. In the upper portion, the variable designation is preceded by an M and the middle portion by an S, and in the lower portion by a B.

The M represents the mean of the teacher's performance over time. We are asking, is there any relation between the mean level of the teacher's performance on these variables and pupil learning? The S represents the variance; the question being asked is, is there any relation between the variance in the teacher's performance and pupil gain? In the lower portion, the designation B stands for the slope of a regression line calculated to estimate how much the teachers changed over time in the acquisition of the skill. The question being asked is, is there a relation between the rate of change in acquiring these teaching skills and pupil learning?

I present the data for two levels of instruction; Level 1 is the primary grades and Level 3 is the upper grades. Notice that the numbers are given for two aspects of reading skill: comprehension and decoding. As is apparent, at Level 1 there is relatively little relation between any performance and increases in pupil learning. One variable is significantly related to learning, "No Feedback" but the relation is negative; that is, the more variance there is in teachers' use of No Feedback, the lower the pupils' gain.

Table 2

F-Values for the Regression Analysis on RAMOS Variables

Level 1

RAMOS Variables	Comprehension df = (1, 4) F	Decoding df = (1, 4) F
M-L+S	.0188	.0698
M-A	.1197	1.0653
M-N	.2044	.2367
M-F	.8175	.9172
M-M	.4368	.7264
M-L+M	1.0262	.6558
M-XX	1.8695	6.1653
M-CIVCJ	1.8024	2.0962
M-QT	.1701	.2542
S-L+S	1.7045	4.3588
S-A	.4222	1.2698
S-N	.3988	.7640
S-F	.6418	1.1089
S-M	.1353	.6696
S-L+M	.4472	.5330
S-XX	1.6562	7.0413**
S-CIVCJ	.1841	.4367
S-QT	.1238	.4219
B-L+S	.0116	.8599
B-A	.0000	.8395
B-N	.9054	.0623
B-F	.1712	.1274
B-M	.0060	1.4526
B-L+M	.3273	2.0032
B-XX	.6309	.0491
B-CIVCJ	.3126	.5908
B-QT	.3356	.0119

*$p_{.05} = 7.71$

**Approaching significance

Table 2 (Continued)

Level 3

RAMOS Variables	Comprehension df = (1, 8) F	Decoding df = (1, 8) F
M-L+S	.0250	1.1284
M-A	1.1484	.1256
M-N	.3413	3.3839
M-F	.0441	1.7119
M-M	2.2265	.6307
M-L+M	1.7356	.4107
M-XX	.0336	.4719
M-CIVCJ	4.6384**	.5418
M-QT	.0022	1.5817
S-L+S	.2989	1.3615
S-A	.3505	.8948
S-N	1.8104	.0761
S-F	2.0024	1.4472
S-M	.9363	.0395
S-L+M	3.2636	.0001
S-XX	.0074	.1106
S-CIVCJ	.0925	.6374
S-QT	15.8450*	.1120
B-L+S	.2833	.5707
B-A	.4851	.0053
B-N	.1780	1.4547
B-F	.1031	5.0617**
B-M	.2181	.2821
B-L+M	2.5481	.3299
B-XX	.0657	.0315
B-CIVCJ	1.1449	.7359
B-QT	.0812	5.4050*

*$p_{.05} = 5.32$

**Approaching significance

If we now look at Level 3 we see a number of significant results. Notice that under Comprehension a variable labelled CIVCJ is directly related to increases in pupil learning. This variable is a score derived from codes representing performances in which the teachers attempted to increase pupils' comprehension skills. An increase in these performances would be a direct measure of the effect of the training program on pupil learning. We see here that the mean value of the teachers' performances does predict differences in pupil learning; that is, teachers who increased in this skill did in fact produce greater pupil gains.

Note also that the variance of the teachers' scores over time on the variable designated QT, "Teacher Questions" was significantly related to pupils' gains in Comprehension.

We should also note that gains on Decoding skills are related to the rate of change in a teacher's use of the Facilitating role and, again, to the use of Teachers' Questions.

These results give us some direct information on the relation between certain kinds of performances and pupil learning. The teacher sample is small, so that one could not make substantial claims for advancing our understanding of effective teaching practices. But the results are consistent with what is to be expected and with other research. The major hypothesis of the study, for example, was that if teachers developed skills that facilitated pupils' reading comprehension, pupils' comprehension skills would improve. We have direct evidence for this result in the data that have been presented. Similarly, the result on the relation between the rate of change with which teachers gain skill in Facilitating and the acquisition of pupils' Decoding skills is also significant. The Facilitating role is one in which the teacher moves around the room and works with the pupils to help them on the instructional

tasks. We would expect this kind of teaching performance to be more effective in increasing decoding skills where the tasks are discrete, concrete, and linear in character. We would also expect an increase in questions to be related to gains in pupil learning for much the same reasons.

In other words, we have data which are directly related to understanding the relation between teaching performance and pupil learning. At the same time these data also provide us an evaluation of the substance of the training program. A training program may be ineffective in either of two ways. It may be ineffective because we are training on variables that are unrelated to pupil learning; or we may be using ineffective training procedures. Now let us consider this latter question.

One of the major goals of training was to increase the use of teaching skills related to reading comprehension. Now, from the data presented in these tables, we see that for the pupils of teachers teaching at the upper level, the mean level of the teachers' performance in these skills was directly related to gains in pupil learning. It was not, however, related at the lower level probably because there is in this particular school less work on comprehension activity in the lower grades and more work on decoding skills.

Now let me introduce a table that shows the effect of the training modules on the teaching of comprehension skills. Notice in this particular table that we have listed Vocabulary skills, which are usually taught by all teachers, and Grammar skills, which are not very frequently taught. The three categories of interest are Comprehension skills, Interpreting skills, and Critical Judgment skills. We see that teachers during the pretraining period devoted about 35 percent of their time to Comprehension skills, about

Table 3

The Effects of the Training Modules on Teaching Performance: Frequency by Module

			Post		
		Pre	6	7	8
Vocabulary Skills					
	B	.21	.23	.23	.27
	B+6	.22		.23	.27
	B+6+7	.22			.27
Grammar Skills	B	.06	.04	.07	.05
	B+6	.04		.07	.05
	B+6+7	.06			.05
Comprehension Skills					
	B	.35	.37	.38	.36
	B+6	.37		.37	.36
	B+6+7	.37			.36
Interpreting Skills					
	B	.15	.18	.24*	.30*
	B+6	.17		.25*	.31*
	B+6+7	.21			.31*
Critical Judgment Skills					
	B	.00	.01	.01*	.01*
	B+6	.00		.01*	.01
	B+6+7	.01			.01

15 percent to Interpreting skills, and 0 percent to Critical Judgment skills. As a consequence of training, we find that the use of Comprehension skills has hardly changed at all, but there has been a substantial and significant increase in Interpreting skills. There has also been a significant but very small change in the use of Critical Judgment skills. Therefore, we may conclude: (1) that the substance of the program is appropriately conceived; that is, the program is training on the appropriate skills because when teachers acquire these skills there is an increase in pupil learning; (2) from the data on training, it is clear that Modules 7 and 8 do produce a significant change in the use of Interpreting and Critical Judgment skills by these teachers. But the modules do not affect Comprehension skills at all.

These data are essentially evaluative data; that is, we may conclude from them that the modules are effectively designed for producing the use of two kinds of teaching skills but not for teaching of the third kind, Comprehension skills. There is also obviously a problem in increasing the absolute amount of Critical Judgment skills used by the teachers. The problem may be that the modules are not well designed for teaching these skills, or it may be that the reading material itself does not require extensive Critical Judgment skills.

What I have presented are data which illustrate that by carefully designing a study we can achieve two goals. First, we can develop our understanding about teaching performances related to pupil learning; and two, we can evaluate the effectiveness of our ideas about what constitutes the appropriate substance of training, and whether or not the training methods that we used were particularly effective.

I should add one piece of information that I have not presented. I stated that there were three training methods in the first year and two in the second. Generally, the differences in training methods made very little difference in the degree to which the skills were learned. It appears to me that what has been achieved in the program is that appropriate substance has been selected but that the particular methods of presentation were not sufficiently fine-tuned to make a difference in how much the teachers learned about and used this substance. In the first year the picture book was particularly effective for increasing on-task productive behavior of pupils, and we suspected the reason for this was that it was easier to recognize off-task behavior in still photographs, and more difficult to grasp the idea through the videotape. But on variables such as Assessing and Diagnosing and Comprehension skills the training methods had no real effect.

These data raise for us a question about the most effective way to portray complex skills of teaching. When the skill is relatively simple, a straightforward programmed approach appears to be effective as it was in the case of the modules designed to train teachers in increasing on-task productive behavior. But as the skills become more complex, involving considerable cognitive activity on the part of teachers, the use of videotapes probably does not portray sufficiently the processes involved to increase substantially the learning of the teachers of the required skills.

Some teachers are simply more responsive to some modules than to others, a fact which raises another kind of training question. In the next table, I present data for each teacher in the study on the amount of change that occurred in each teaching variable for each teacher. Across the top of the table are the designations of the variables at which we have been looking and

Table 4

Teacher Regressions on
RAMOS Variables Over Time
Modules 6, 7, 8

Teacher	Size: L+S	Role: A	Role: N	Role: F	Role: M	Mobility: L+M	Feedback Sign XX	Skills CIVCJ	Material: QT
01	.0001(-)	.8085(-)	.3397(-)	.0155(+)	.2747(-)	.7919(-)	.6270(+)	.4454(-)	.2251(+)
02	.0001(+)	.3342(-)	.6001(+)	.8263(+)	.9946(-)	.1450(-)	.8321(-)	.6500(+)	.2840(+)
03	.1142(+)	.4328(+)	.0135(-)	.0132(+)	.0061(-)	.1666(+)	.2075(+)	.3927(+)	.0718(-)*
05	.8654(+)	.0462(-)	.8787(+)	.2545(+)	.5267(-)	.0634(+)*	.3866(-)	.0047(+)	.7501(+)
06	.0214(+)	.5146(-)	.1451(-)	.0134(+)	.4867(-)	.4289(-)	.4303(-)	.2864(+)	.1190(-)
07	.1977(+)	.8519(-)	.5809(-)	.0709(+)*	.8862(-)	.6216(+)	.8983(+)	.0147(+)	.5864(-)
08	.2458(-)	.1238(+)	.0269(-)	.0022(+)	.8842(-)	.0664(+)*	.0607(+)*	.5712(-)	.8271(-)
09	.7440(+)	.4720(-)	.4032(+)	.8428(+)	.0296(-)	.4777(-)	.1271(-)	.3277(+)	.0502(+)
10	.0477(+)	.8625(-)	.6894(-)	.2692(+)	.0385(-)	.6771(-)	.8547(+)	.2749(+)	.9730(-)
11	.0206(-)	.1502(-)	.1740(+)	.0776(+)*	.0593(-)*	.7155(-)	.2282(+)	.6171(+)	.6502(+)
12	.8311(-)	.4800(+)	.0136(-)	.2569(+)	.8171(-)	.3875(-)	.2463(+)	.3132(-)	.1910(-)
13	.1027(+)	.2707(+)	.0195(-)	.0871(+)*	.5417(-)	.4374(+)	.1676(+)	.0013(+)	.6321(-)
14	.9229(+)	.1586(+)	.4355(-)	.9795(-)	.7510(-)	.8469(-)	.1182(+)	.7107(+)	.7047(+)
15	.0048(+)	.7719(-)	.9318(+)	.4937(+)	.1821(-)	.1687(-)	.7557(+)	.8640(-)	.3614(-)
16	.5389(+)	.4141(+)	.7804(-)	.1222(+)	.3230(-)	.5206(-)	.1349(+)	.1709(+)	.4993(-)
18	.9495(-)	.0275(-)	.9089(+)	.7433(-)	.6917(-)	.1183(+)	.6320(+)	.3775(+)	.1989(+)
19	.9095(-)	.2984(+)	.5995(-)	.8658(-)	.4085(-)	.5152(-)	.3324(+)	.7092(+)	.5757(+)
23	.0473(-)	.2457(+)	.9653(-)	.6659(+)	.1516(-)	.2076(-)	.7931(+)	.8982(+)	.7955(-)

Significant p-value

*Approaching significance

down the left-hand side is the number of the teacher. The number in the columns is the size of the p value estimating the statistical significance of how much the teachers changed over time on that particular variable. The significant variables are indicated as are those approaching significance.

As you read this table, notice that there are some teachers who did not change at all over time; for example, Teacher 14. There are some teachers who changed in some respects more frequently than other teachers. Notice also that some of the changes were in a positive direction and some were in a negative direction.

By reading down the columns you can find the kinds of changes that occurred most frequently in a particular variable; for example, the variable L+S (use of small groups) shows the greatest change, but four of these changes are in the positive direction and three in the negative direction. Data of this kind raises a question about why there is a differential effect of training for a particular teaching activity. Notice also that a role such as Assess/Diagnose has been changed significantly only for two teachers, and this change was in a negative direction. The role, Facilitating, changed in four cases. The directions were all positive, and there were three changes approaching significance, again all positive. We have already seen that an increase in this role is significantly related to changes in gains in pupils' decoding skills.

Conclusion

In summary: we have developed a study which enables us to answer different kinds of questions. One kind of question is what is usually regarded as a research question: what differences in teacher performances are directly related to pupil learning? A second kind of question inquires about the extent to which a particular training program is producing changes in the teaching behavior of teachers.

The study was designed as part of a research and development process. Data of this kind can now be used to think through what may be occurring as the training program is used. We need to ask such questions as why are some teachers apparently completely unresponsive to any of the training; that is, why is their style totally unaffected by the training? We need to ask also why the training has positive effects in some cases and negative effects in others. We need to ask why an increase in one desirable kind of performance occurs at the expense of another performance, equally desirable. We need to determine whether an increase in Assessing/Diagnosing, for example, which is desirable, actually improves the quality of instruction even though the absolute amount of instruction decreases.

The general theme of this paper is that by using what is generally regarded as a research design, one can also accomplish significant evaluative purposes. The sharply drawn distinction between research and evaluation, and the equally sharp distinction between the carefully controlled laboratory-type experiment and the study occurring "in the field" seem to be arbitrary and not very helpful distinctions. I suggest to you that this design, which is one of many kinds that can be used fruitfully, combines the purposes of what is usually conceived of as a research study and what is usually conceived of as the evaluation study.

References

McDonald, F. J., & Elias, P. The Beginning Teacher Evaluation Study, Phase II: Final Report. Princeton: Educational Testing Service, 1976.