ABSTRACT
        An inservice program, Project Alert, was conducted
during the summer of 1973 in 50 school districts in New York State.
Project Alert had a two-fold purpose: to increase the
diagnostic/prescriptive skills of selected classroom teachers and to
increase the leadership capabilities of the selected teachers so that
they might function as inservice facilitators during the following
school year. Evaluation of the inservice program was accomplished
through the use of teachers' self-ratings, directors' ratings, and
monitors' ratings. Analysis of the pre- and post-self-ratings
suggests that teacher competency increased between the start of the
inservice program and its termination. (TO)

THOMAS P. FITZGERALD
New York State Education Department

RICHARD M. CLARK
State University of New York at Albany

BEST COPY AVAILABLE

PROCESS EVALUATION FOR INSERVICE TRAINING*

Many school districts have increased their inservice activities

for teachers and have explored a variety of methods in presenting pro-

grams. Both product and process evaluation are necessary to provide

the necessary information on which to judge the effectiveness of such

training programs. The ultimate goal of most inservice programs may

be viewed as a change in the product: in this case, the improved read-

ing achievement of children. The more immediate objective may be

viewed as a process change in which the teacher demonstrates a cog-

nitive, affective or behavioral development as a result of the inser-

vice training.

Little effort has been made to evaluate many inservice programs.

When an evaluation has been made, it has often taken the form of a

series of questions asking the participants to indicate the perceived

---

*Paper presented at the National Reading Conference, Kansas City,
Dec. 1974.

degree of enjoyment, interest, or value in the inservice activity.
Few programs have used a product evaluation in an effort to link stu-
dent achievement to teacher training. When inservice programs have
been subjected to either a product or process evaluation in reading,
the results have tended to receive criticism based on their reported
procedures and instrumentation (Mobury, 1972; Herrick, 1957). Gen-
erally, training programs have not included judgements concerning
changes in the classroom performance of teachers. More frequently,
the evaluation of reading inservice has focused on attitudes (Schirmer
and Navarre, 1968; Brown and Dutton, 1972), cognitive growth (Wall,
1965) and simulation techniques (Kasden and Kelly, 1969).

Another approach to assessing process change is the rating of
classroom performance estimates made by supervisors (Sawyer and Taylor,
1968) and by teacher self-rating (Heilman, 1966; DeCarlo and Cleland,
1968). Such self-rating by teachers was designed to reflect changes
in professional knowledge about the reading process or changes in at-
titudes toward reading techniques. However, previous studies have
not employed teacher self-rating to assess classroom applications of
the topics discussed during the inservice.

Self-ratings of teacher classroom behavior have not been used
in conjunction with external ratings of the teacher's classroom
behavior.

Purpose

The purpose of this study was to compare three methods of assess-
ing teacher classroom behavior in the teaching of reading. A rating
scale was constructed based on behavioral definitions of steps in a
diagnostic-prescriptive approach to reading instruction. The behavior

assessed was then related to the objectives of an in-service program. Pre and post workshop ratings were made by teachers on a self-assessment basis and by workshop directors who were working closely with the teachers involved. Post workshop ratings were made by external observers on the basis of classroom visits and teacher interviews.

The questions considered in this study are:

(1) To what degree do ratings of teaching skills in reading change from pre-workshop to post-workshop?

(2) What are the relationships of the three methods of assessing teacher skill?

(3) What are the inter-relationships of the four skills in teaching reading which are being assessed for each procedures used?

(4) What relationships exist between a teacher's academic preparation in reading and the various assessments of reading competency?

Procedure

An inservice program, Project Alert, was conducted during the Summer of 1973 in 50 school districts in New York State. The program was initiated and coordinated by the Reading Bureau of the State Education Department. Each of the fifty district programs involved was administered by a project director who had received assistance from the Reading Bureau's staff and had received a six package kit of inservice material which was used as the basic inservice tool. Project Alert had a twofold purpose:

(1) to increase the diagnostic/prescriptive skills of selected

classroom teachers; and

(2) to increase the leadership capabilities of the selected

teachers so that they might function as inservice facili-

tators during the 1973-74 academic year.

Data for this study were compiled from 383 teachers participating

in the inservice program. Each teacher recorded a rating of their

classroom skills during the first (pre-inservice) and last (post-in-

service) weeks of the program. Each project director used the same

rating scale to estimate the classroom performance of individual tea-

chers during the second and fourth weeks of training. An outside

monitor observed randomly selected teachers and estimated their per-

formance using the same rating scale during the fourth week.

Assessment Instrument

The assessment instrument (Appendix A) was constructed on a five

point scale ranging from little behavioral evidence (1 on the scale)

to considerable behavioral evidence (5) of skill demonstrated within

a classroom. The set of five definitions was organized to reflect

reasonably discrete steps leading from an absence of a skill to a

highly developed skill display within the classroom.

Such a set of five definitions was constructed in each of four

areas: diagnosis, prescription, classroom management (execution)

and evaluation.

A panel of reading administrators suggested stages in the develop-

ment of each of the four areas. The project evaluators then formu-

lated behavioral statements to reflect these stages. The scales were

modified in language for the self-rating and the observation. The

scales also were organized so that teachers and administrators could have a common understanding of the described operations.

Results

To determine the consistency of the ratings across rating groups of teachers, directors and monitors and the consistency of ratings across the time span of the inservice program, the mean ratings of each group were computed. These data are displayed in Table 1. The teacher self-ratings before the inservice were consistently higher than the estimate made by the directors during the second week. The lowest ratings from teachers occurred in the category of diagnosing while the directors rated the evaluation skills of the teachers as the lowest demonstrated skill. Both groups rated the classroom management (execution) skill as the most highly developed.

(Insert Table 1 about here)

The second series of ratings during the final weeks of the inservice program, revealed considerable variations among rating groups. The directors reported the highest ratings while the monitors recorded the lowest rating. The monitors' estimates were considerably lower than either the teacher's self-rating or the directors rating.

The data presented in Table 1 indicate that changes in the ratings occurred during the inservice program in teachers' self-ratings and, more dramatically, in the ratings made by the directors. For each

group the greatest rating change occurred in the diagnosis category

with the smallest change occurring in the execution category.

(Insert Table 2 about here)

The inter-correlational matrix presented in Table 2 shows the

inter-relationships between the various ratings of teacher behavior.

In their pre-workshop assessments teachers clearly do not make inde-

pendent ratings of the four skills which they are asked to rate.

Correlations range from .49 to .82. (Correlation cluster 1-4). Post

workshop ratings of teachers are also inter-related, but to a lesser

degree, with correlations ranging from .43 to .57. (Cluster 5-3).

While positive correlations continue, there is a relatively small

relationship between the ratings teachers give themselves pre-work-

shop and post-workshop. Correlations range from .07 to .38.

(Cluster 1-4 related to 5-8).

Directors show a consistent pattern of interrelationship in

their ratings of the four skills both _within_ the two periods when

they made ratings and between the periods. (Cluster 9-16). However,

the relationship between the ratings of the Director and the ratings

of the teachers are generally positive but small. For example, the

inter-correlations of items 5-8 with 13.-16 represents the relation-

ships between the teachers' post workshop rating and the Directors'

post-workshop rating. Correlations range from .09 to .28.

Inter-correlations of monitors' ratings in the four areas suggest

that monitors made almost completely a global rating. Correlations range from .91 to .94 (Cluster 17-20). Further, extremely low relationships are seen between teacher self-ratings and monitor ratings. (Cluster 1-8 and 17-20). However, modest but significant positive ratings were obtained between Directors' and monitors ratings. (Cluster 9-16 and 17-20).

The fourth question in this study was the relationship between the teachers' self-rating and their academic background in reading. It might be expected that teachers with more advanced preparation in reading would evaluate their skills higher than those with little formal course work in reading.

(Insert Table 3 about here)

Table 3 summarizes the ratings of two groups of teachers: one with advanced degrees in reading and one with a single undergraduate course in reading or no formal preparation. As expected, the group with advanced training in reading consistently indicated higher self-ratings when compared to the group with little course work. Teachers with little training recorded positive changes in their ratings from the pre- to post assessment. Unexpectedly, the teachers with advanced degrees tended to lower their post-assessment of the classroom behavior possible indicating that their graduate program did not concentrate on classroom applications of the four skill areas that were rated.

Conclusions

The use of teacher self-ratings coupled with ratings from other sources appears to be a reasonable approach to determining behavioral changes generated through an inservice program. The rating from teachers, inservice directors and outside monitors produced different estimates of classroom behaviors, and these ratings did differ between the start of the inservice program and its termination.

The two pre-post ratings suggest that teacher competency increased and a small positive relationship was found between the two means of assessment. In this study it was assumed that the four specific skills to be assessed might be somewhat related, but that they were also to some degree independent. For example, teachers very good in diagnosis might generally be above average in evaluation, but the two skills would not be highly correlated. Self assessments of teachers' and monitors' ratings inter-relate to a moderate degree, suggesting that some independence might exist for the four skills rates. Monitors', however, failed to make any meaningful distinction among the four areas. Further, monitors' ratings had almost a zero correlation with teacher self-ratings, and relatively small relationships with directors' ratings.

Supposedly, the three measures used in this study, self-rating, directors' rating and monitors ratings were all measures of the same behaviors. Empirically, the three measures are not the same. Further work needs to be done in order to determine which assessment technique may be the most valid for a specific purpose.

The teacher's rating pattern appeared to be related to academic background especially for the pre-inservice assessment. Surprisingly,

during the post assessment, a smaller number of teachers with the
higher academic background in reading reported estimates of their
skills to be at the highest ranking (5), while the number of teachers
with low academic background in reading generally increased in re-
porting higher self estimates. Perhaps an effect of the inservice
program was to reveal to teachers more of what could be known about
the diagnostic-prescriptive procedure. If this explanation is cor-
rect, then teachers with considerable academic background in an
area, would enter a training program thinking that they had little
need for further work with techniques for improving reading in-
struction. During the inservice program they might have discovered
a need for additional ways to apply their knowledge within their
classroom. This explanation is being tested in further research.

Appendix A

Descriptive Categories for Ratings

Diagnosis

1. No real diagnostic techniques in evidence except in such general terms as "children lack reading skills."

2. Diagnostic techniques very general. Teacher has some data indicating general reading weaknesses of some members of the group. However, it is quire possible that some individuals don't need the program which is specified.

3. Diagnostic techniques seem to be rather general. Data are available to indicate the general needs of a group, but it is not clear that every individual in the group has the need specified.

4. Evidence of some individualized diagnostic procedures. Teacher is able to show specific procedures, but perhaps for only part of the group. For some children, a rather general diagnostic approach has been used.

5. Clear evidence of an individualized diagnostic procedure. Teacher is able to show specific procedures which have been used which seem to pinpoint the needs of each individual in the group.

Prescription

1. Materials and activities used seem not to be especially relevant to the needs of many individuals in the group.

2. No clear rationale seems to exist which relates the materials and activities of the program to the needs of the learners. The material and activities may well be appropriate, but such appropriateness is assumed rather than reasoned.

3. Materials and activities used seem to be based on a group assessment of need, not an individual diagnostic approach. Materials and activities appear to be well suited to this group diagnosis.

4. The teacher seems to have a clear rationale for the selection of materials and activities for some children, but some members of the group appear to be working on material which is not based on a specific prescription of need.

5. Teacher seems to have a clear rationale which she has used to match each child to the materials and activities with which he is engaged. Thus, the teacher explains the work group. Materials selected to be appropriate for the purposes which the teacher

has in mind.

Execution

1. Many problems in carrying out program as planned. Activities are often not carried out by children, because of material bottle-necks, lack of effective organization, or similar reasons.

2. Some problems encountered in carrying out program plans. Some children seem not to know what they are expected to do. At times, they disrupt other children. Frequent delays interrupt the flow of the day's activities.

3. A rather large amount of time seems to be taken by the mechanics of the operation. Children often have to wait for material to be ready. There may be use of "busy work" material primarily to keep children occupied. However, teacher-pupil rapport generally satisfactory, no general disruption of program.

4. A generally good learning environment. Children are usually occupied, and appropriate material is readily available. At times, children are forced to wait for teacher to be available so that they may move to new activity, but usually children move smoothly from one activity to another.

5. Teacher exhibits a great deal of skill in having appropriate material is readily available for children, in arranging a good learning environment, in moving children from one activity to another, in maintaining rapport with children.

Evaluation Process

1. A fixed program is offered the children, perhaps on the basis of diagnostic techniques established before the program began. However, no effort is apparent within the program, to determine whether the children, even as a group, are meeting the objectives set for them. Even though pre- post-testing might be used, the data are so general that little or no evaluation of specific program materials is possible.

2. Evaluation of the program seems to be largely on a pre- post basis and largely on a group basis. Thus, although children may be programmed on the basis of assessed needs, little or no evaluation activity goes on within the group. After the program is ended, one might know how children have benefited, but such information would not be used during the program.

3. Some evaluation activities occur within the instructional period. Some evidence that at least groups of children are making progress, within the program, and that such evidence is being used to make program decisions.

4. Clear evidence that evaluation of individual children is taking

place on a regular basis and that these evaluations are used to determine the next set of activities for some individuals. At times evaluations might be somewhat delayed or not up to date for individuals.

5. Clear evidence that frequent checks are made of the progress of almost every child. Evidence from such evaluations is promptly used to determine the next set of activities in which the individual should be engaged.

References

Brown, E. and Dutton, D. The Summer of 72: An Assessment of the adult education institutes in Tennessee. Office of Education Bulletin, Washington, D.D.: December, 1972.

DeCarlo, M. and Cleland, D. A Reading Inservice Education Program for Teachers. Reading Teacher, 1968, 22, 163-64.

Heilman, A. Effects of an intensive inservice reading program on teacher classroom behavior and pupil reading achievement. Reading Teacher, 1966, 19, 622-26.

Herrick, V. The evaluation of change in programs of inservice education in Inservice Education, N. Henry (ed.), Yearbook of National Society for the Study of Education, Chicago, Illinois, 1957, 56, 311-338.

Kasden, L. and Kelly, D. Simulation: Inservice education for teachers of reading. Journal of Experimental Education, 1969, 38, 79-86.

Moburg, L. Inservice teacher training in reading. Reading Information Series. International Reading Association, Newark, Delaware, 1972.

Sawyer, R. and Taylor, L. Evaluating teacher effectiveness in reading instruction. Journal of Reading, 1968, 11, 415-18, 483-88.

Schirmer, G. and Navarre, G. Evaluating a summer seminar. Journal of Reading, 1968, 12, 234-36.

Wall, R. Teacher attitudes toward reading educational books and periodicals. Illinois School Research, 1969, 1, 14-19.

Table 1

Mean Ratings of Teachers' Diagnostic
Prescriptive Abilities

	ABILITIES			
Ratings By	Diagnosis	Prescription	Execution	Evaluation
Teacher - Pre summer	3.48	3.52	3.95	3.69
Director - 2 weeks	3.37	3.43	3.71	3.23
Teacher - Post summer	3.79	3.80	3.98	3.88
Director - 4 weeks	4.24	4.11	4.25	3.94
Monitor - 4 weeks	2.22	2.10	2.14	1.89

Table 2

Correlations Between Various Assessments
of Diagnostic-Prescriptive Ability

Assessments By	1	2	3	4	5	6	7	8	9	1C
Teachers' Pre										
1. Diagnosis	1.00	.82	.49	.64	.38	.35	.08	.26	.26	.2:
2. Prescription		1.00	.49	.66	.37	.38	.15	.29	.20	.1:
3. Execution			1.00	.68	.18	.24	.15	.18	.0⁹	.06
4. Evaluation				1.00	.07	.30	.32	.17	.13	.1:
Teachers' Post										
5. Diagnosis					1.00	.55	.35	.52	.31	.3:
6. Prescription						1.00	.43	.57	.23	.2:
7. Execution							1.00	.51	.11	.C:
8. Evaluation								1.00	.20	.1
Directors' two week										
9. Diagnosis									1.00	.7
10. Prescription										1.C
11. Execution										
12. Evaluation										
Directors' four week										
13. Diagnosis										
14. Prescription										
15. Execution										
16. Evaluation										
Monitors' rating										
17. Diagnosis										
18. Prescription										
19. Execution										
20. Evaluation										

Table 2 continued

Assessment By	11	12	13	14	15	16	17	18	19	20
Teachers' Pre										
1. Diagnosis	.23	.19	.16	.10	.05	.04	.11	.12	.14	.09
2. Prescription	.16	.12	.12	.09	.04	.00	.05	.06	.08	.03
3. Execution	.11	.09	.02	.03	.00	.02	.01	.02	.01	-.02
4. Evaluation	.17	.11	.04	.00	-.01	-.03	-.01	.00	.00	-.04
Teachers' Post										
5. Diagnosis	.27	.25	.28	.20	.17	.16	.01	.00	.03	.00
6. Prescription	.18	.11	.21	.13	.11	.09	.08	.08	.09	.07
7. Execution	.14	.12	.16	.11	.12	.11	.08	.08	.08	.09
8. Evaluation	.14	.12	.16	.09	.12	.06	.06	.04	.05	.05
Directors' two week										
9. Diagnosis	.63	.67	.68	.54	.45	.49	.19	21	.22	.19
10. Prescription	.71	.76	.62	.63	.51	.55	.19	.22	.19	.18
11. Execution	1.00	.72	.55	.60	.65	.55	.20	.23	.22	.21
12. Evaluation		1.00	.57	.63	.56	.68	.19	.21	.17	.16
Directors' four week										
13. Diagnosis			1.00	.73	.61	.67	.19	.18	.19	18
14. Prescription				1.00	.79	.78	.19	.16	.17	.19
15. Execution					1.00	.71	.22	.21	.22	.24
16. Evaluation						1.00	.19	.18	.16	.18
Monitors' rating										
17. Diagnosis							1.00	.91	.92	.92
18. Prescription								1.00	.94	.92
19. Execution									1.00	.93
20. Evaluation										1.00

TABLE 3

Contrast of Pre- and Post- Self-Ratings of Teachers
with the Most and Least Academic Background
in Reading

Teachers with Advanced Degrees in Reading							Teachers with the Least Academic Background in Reading						
Skill		1	2	3	4	5	1	2	3	4	5	Skill	
Diagnosis	Pre	0	0	2	17	33	15	46	72	83	2	Diagnosis	Pr
	Post		1	1	25	25	3	7	80	109	19		Po
Prescription	Pre			4	17	31	11	37	81	82	6	Prescription	Pr
	Post		1	4	24	23	4	9	68	113	21		Po
Execution	Pre		1	9	15	27	4	11	42	124	37	Execution	Pr
	Post			4	30	18	1	9	21	152	35		Po
Evaluation	Pre			5	20	27	5	22	73	104	14	Evaluation	Pr
	Post		2	4	26	20	1	10	55	127	25		Po

Note: Columns 1–5 under each section represent Rating.