DOCUMENT RESUME

ED 022 715 SP 001 554

By-Hite, Herbert

FOLLOW-UP STUDY: LONG-TERM EFFECTS OF MODIFIED INTERNSHIP FOR BEGINNING ELEMENTARY TEACHERS.

Spons Agency-School Information and Research Service, Seattle, Wash.
Pub Date 11 Jun 68
Note-10p.
EDRS Price MF-$0.25 HC-$0.48
Descriptors-*BEGINNING TEACHERS, EDUCATIONAL EXPERIMENTS, EFFECTIVE TEACHING, ELEMENTARY SCHOOL TEACHERS, *INSERVICE TEACHER EDUCATION, *INTERNSHIP PROGRAMS, OBSERVATION, STATISTICAL ANALYSIS, *TEACHER BEHAVIOR, *TEACHING LOAD
Identifiers-Seattle Performance Appraisal Guide

This study is a followup conducted 1 year after the 1966 study (ED 001 162) which attempted to assess the effects of reduced loads and inservice help on the classroom behavior of 120 beginning teachers. A conclusion of the original study was that the experimental group showed at least 25% higher scores on teaching performances than the control group; the followup study was designed to determine if the relative differences in teaching behavior persisted. Subjects were 10 randomly selected members from each of the original 4 experimental groups. Four observers, 2 of which participated in the original study, were trained by Harry L. Garrison, who had trained the original observers in the use of the same instrument. Each team of 2 observers appraised half of the 40 subjects, and the reliability of observers' ratings was assessed statistically. There were no significant differences when analysis of variance was computed among the 4 groups for each of 10 variables . Results indicated that differences among the 4 groups tended to become smaller after a year; however, most of the variability was accounted for by the atypical scores of 5 subjects. Judging by means of observer ratings, the total group of 40 tended to show small positive gains in teaching performance standards. Included are 3 statistical tables and the reactions of Garrison who designed the observation instruments used. (JS)

at your service SIRS · SCHOOL INFORMATION AND RESEARCH SERVICE

309 KING COUNTY COURTHOUSE • SEATTLE, WASHINGTON 98104 • TELEPHONE MAIN 2-5900, EXT. 435

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

June 11, 1968

1967-68: 11

FOLLOW-UP STUDY: LONG-TERM EFFECTS OF MODIFIED INTERNSHIP FOR BEGINNING ELEMENTARY TEACHERS

In 1966 Professor Herbert Hite, College of Education, Washington State University published an interesting report on an experimental project in teacher training, "Effects of Reduced Loads on Intensive Inservice Training Upon the Classroom Behavior of Beginning Elementary Teachers."

The Staff and Trustees of SIRS were pleased to help, financially, in a follow-up study which Professor Hite organized and administered. And we are happy to publish Professor Hite's report on this second study.

Since Dr. Harry Garrison aided Professor Hite in the project by designing instruments by which performance was appraised, we asked him to read the manuscript and react. Dr. Garrison's background includes serving the Seattle Public Schools as Personnel Assistant, Evaluation, and as Coordinator of Student Teaching, Western Washington State College.

We have included Dr. Garrison's letter for its thought-provoking values.

Morton A. Johnson
Director

FOLLOW-UP STUDY: LONG-TERM EFFECTS OF MODIFIED INTERNSHIP FOR BEGINNING ELEMENTARY TEACHERS

by Herbert Hite, Professor
Washington State University

INTRODUCTION

In this follow-up study, the teaching of 40 second-year, elementary teachers was appraised to determine if the effects of special treatments in the previous year would still be apparent. The 40 teachers were part of a group of approximately 120 elementary beginning teachers who were the subjects of an experiment to assess effects of reduced loads and inservice help on the classroom behavior of beginning teachers.[1]

In the original study, in 1965-66, the beginning elementary teachers were first matched and then assigned to four different groups. Group I was released of 25 per cent of classroom time for preparations and to confer with a supervisor who observed the teacher's performance during the week. Group II was released from 25 per cent of classroom teaching time and used this time partly to visit experienced teachers with similar assignments. Group III teachers were assigned about 25 per cent fewer pupils than the average of the classrooms in that school district. Group IV teachers were a control group and received no special treatment of released time or special inservice help. A conclusion of the original study was that the experimental group showed at least 25 per cent higher scores on teaching performances than the control group. The purpose of the follow-up study was to determine if the relative differences in teaching behavior persisted after a period of one year.

PROCEDURES OF THE FOLLOW-UP STUDY

The project staff corresponded with the elementary schools who participated in the first study to locate all of the original group of subjects who were still teaching in their original assignments. The staff then randomly selected 12 members of each of the original four experimental groups. They sent each of these 48 selected subjects a post-card questionnaire inviting them to participate in the follow-up study. Enough returns were received to form four groups of ten members each.

Four observers were experienced elementary teachers. Two had served as observers in the original study; the other two were substitute teachers in one of the participating school districts. The observers learned to use the Seattle Performance Appraisal Guide, the same instrument used in the original study, under the direction of Dr. Harry L. Garrison. Dr. Garrison, who also trained the observers in the first study, had the four study the Appraisal Guide and practice using it to judge video-taped performances of experienced teachers. The observers could in this way compare their appraisals with each others' and examine the taped performances repeatedly. After this training experience, the four practiced by visiting and appraising cadet teachers.

---

[1] Herbert Hite. "Effects of Reduced Loads and Intensive Inservice Training Upon the Classroom Behavior of Beginning Elementary Teachers." Coop Research Project #2973 U. S. Office of Education, Washington State Superintendent of Public Instruction and WSU, 1966.

For the purpose of this follow-up study, a team of two observers was to appraise each of the 40 subjects, ten members from each original group, one time. The team of observers arrived at a prearranged time at the subject's classroom. They were seated in the rear of the room, and independently observed and rated the teacher on ten different teaching behaviors. Each team of two observers rated one half of the subjects.

The observers' ratings were then analyzed by computer in the office of the State Superintendent of Public Instruction.

RESULTS

The first step in analyzing the data was to assess the reliability of the observers' ratings. Table 1 shows correlation coefficients of each pair of observers for each of the ten performances of the teachers which they rated. In order to compare observer scores, Pearson product moment coefficients, which were obtained for each pair of observers, were converted to Fisher Z scores. As shown in Table 1, one pair of observers agreed more closely than the other pair. The degree of observer agreement was sufficiently high in the judgment of the project staff to warrant further treatment of the observer data.

TABLE 1

OBSERVER CORRELATIONS[a] BY TEAM FOR EACH TEACHING BEHAVIOR

Observer Team	Behaviors										
	1	2	3	4	5	6	7	8	9	10	Overall
1	.95	.92	.84	.88	.89	.86	.88	.89	.78	.86	.88
2	.71	.74	.55	.64	.55	.61	.42	.50	.39	.28	.55

[a] Fisher Z scores: Henry F. Garrett, "Statistics in Psychology and Education," New York: David McKay Co., Inc., 1962, pages 132, 172.

The data for analyzing changes in ten teacher performances on the part of the 40 subjects is in Table 2. The mean ratings for each observer team on the occasion of the last appraisal of the original study appear as scores under "Round 4" in the Table. This mean observer rating is of the scores of only the ten teachers who were members of each group in the follow-up study. The mean ratings for each observer team for each group appears as scores under "Round 5" in the Table. Next to these two scores for each group for each behavior, is the difference between the two mean ratings. The Table, then, shows the difference between appraisals of the four groups of teachers at the end of the original study and appraisals made slightly over one year later.

Analyses of variance were computed for the differences among the four groups of subjects for each variable. There were no significant differences.

DISCUSSION

1. Because only ten members of each of the original four groups were members of the four groups in the follow-up study, the mean ratings for Round 4 were not those of the original groups. The ten members of Groups I, III and IV had lower mean ratings than their respective original groups; Group II in the follow-up study had mean ratings that were higher on Round 4 than their original total group. Thus, the relative rank of the four groups on Round 4 was different from the original ranking in which Group IV was considerably lower than the other three, and Group III was clearly highest of all.

2. The differences among the four treatment groups tended to become smaller after a year. The highest scores on Round 4 (Group II) tended to move in a negative direction; the lowest scores on Round 4 (Group IV) tended to move in a positive direction. This general observation, however, is modified by the following observation.

3. Most of the variability among the groups on Round 5 was accounted for by the scores of five individuals. These five teachers were rated differently on Round 5 compared to Round 4 by at least three points on a seven-point scale. (See Table 3) This amount of change over the year period was not typical of the general group of 40. On the follow-up study the greatest gains were made apparently by Group IV, while the negative-direction changes were scored for Group II. If the scores of three individuals are not used, however, the two groups show almost the same amount of difference between Rounds 4 and 5--.13 and .08 respectively.

The dramatic changes in observer ratings of five individual teachers can not be attributed to the original experimental treatments, because the five were members of all four groups--two members from Group IV were in the group of five. Possibly there were real changes in teaching behavior of this magnitude. There is also the possibility that the observer teams saw an atypical teaching performance on the day they appraised these five individuals. There is also the possibility of a systematic bias on the part of the observers. At any rate, the staff feels that the changes in observer ratings for these five teachers can not be considered an outcome of the special treatments, or of the lapse in time.

4. The general tendency for the total group of 40 teachers was to show small positive gains in teaching performance standards, judging by the rather slim evidence of mean observer ratings. Twenty-four of the 40 made positive gains.

5. The greatest gains of the total group of 40 teachers were on Variable 3, Use of Resources, Variable 5, Organizing the Class, and Variable 8, Student Participation.

SUMMARY

Observers' ratings as used in the original and follow-up studies are gross measurements of teaching behavior. The criteria scale ranges from a possible low score of 1.0 to a high of 7.0. In the two studies, beginning teachers tend to be rated lower than experienced teachers. The possible change in scores on the type of scale used in the Seattle Performance Appraisal Form makes it very unlikely that an Analysis of Variance for groups of ten individuals will show statistically significant differences. In spite of the small differences noted, there appears to be a tendency for beginning teachers to show consistent and small gains from midway in the first year to the end of the second year of teaching. These teachers in this study, in general, appeared to show this tendency to improve regardless of the original experiment treatment. That is, the relative advantages of teachers receiving the original treatment of reduced load and/or inservice help was maintained.

This conclusion is difficult to prove statistically, for the reasons already stated. The conclusion depends somewhat on the interpretation of the fact that a small number of teachers were rated much differently from the original study and these differences from first year to second year were not typical for the total group. If the atypical cases are left out of the calculation the original differences among the four groups remain in about the same relationship after a period of one year.

TABLE 2

MEAN SCORES OF FOUR TREATMENT GROUPS ON FINAL OBSERVATIONS OF ORIGINAL STUDY AND ON OBSERVATIONS OF FOLLOW-UP STUDY

		GROUP I			GROUP II			GROUP III			GROUP IV		
	Variable	Rounds 4	5	Diff.	Rounds 4	5	Diff.	Rounds 4	5	Diff.	Rounds 4	5	Diff.
1.	Suitability of Goals	2.71	3.10	.39	3.70	3.35	-.35	3.15	3.30	.15	2.12	3.00	.88
2.	Student Acceptance of Goals	2.28	2.35	.07	3.50	2.75	-.75	2.91	3.40	.49	2.14	3.15	1.01
3.	Use of Resources	2.67	3.45	.78	3.70	3.70	.00	2.84	3.30	.46	1.53	3.00	1.47
4.	Selecting the Plan	2.52	3.00	.48	3.61	3.15	-.41	3.07	3.05	-.02	1.94	3.00	1.06
5.	Organizing Class	2.25	2.90	.65	3.70	3.70	.00	3.14	3.55	.41	2.09	3.15	1.06
6.	Class Control	2.09	2.70	.61	3.76	3.15	-.61	2.96	3.05	.09	2.25	3.05	.80
7.	Classroom Climate	2.26	2.80	.54	3.77	3.65	-.12	2.99	3.40	.41	2.23	3.30	1.07
8.	Student Participation	2.43	3.05	.62	3.83	3.40	-.43	3.34	3.65	.31	2.07	3.00	.93
9.	Measuring Achievement	2.25	2.65	.40	3.64	3.20	-.44	3.07	2.95	-.12	2.60	3.20	.60
10.	Using Measurement	1.83	2.45	.62	3.37	3.05	-.32	2.85	3.05	.20	2.37	2.95	.58
11.	Performances 1 and 2	2.49	2.59	.09	3.60	3.05	-.55	3.03	3.35	.32	2.13	3.08	.95
12.	Performances 3, 4 and 5	2.51	3.11	.60	3.67	3.53	-.14	3.02	3.30	.28	1.85	3.08	1.23
13.	Performances 6, 7 and 8	2.26	2.85	.59	3.79	3.40	-.39	3.10	3.36	.26	2.18	3.11	.93
14.	Performances 9 and 10	2.04	2.55	.51	3.51	3.13	-.38	2.96	3.00	.04	2.49	3.08	.59
15.	Performances 1 - 10	2.30	2.85	.53	3.66	3.32	-.34	3.03	3.27	.24	2.13	3.08	.95

TABLE 3

DIFFERENCE IN MEAN OBSERVER RATINGS ON ROUNDS 4 AND 5 FOR TOTAL OF ALL TEN BEHAVIORS FOR EACH TEACHER BY GROUP

Teacher	Group	Difference Between Rounds 4 and 5, Variable 15
033	1	- .87
094	1	1.11
047	1	-1.29
036	1	- .12
012	1	.48
007	1	.20
003	1	3.25
058	1	.67
099	1	.14
085	1	1.72
040	2	.42
062	2	-1.39
053	2	-4.18
049	2	- .44
041	2	- .28
110	2	1.16
104	2	- .21
102	2	.04
092	2	- .70
079	2	2.15
056	3	.07
043	3	1.44
032	3	1.20
018	3	.36
015	3	2.88
013	3	.97
004	3	- .41
093	3	- .84
090	3	-3.27
068	3	1.92
052	4	.99
055	4	- .98
030	4	- .80
023	4	.23
014	4	1.24
011	4	4.81
010	4	.12
009	4	- .44
006	4	.69
108	4	3.60

DR. GARRISON'S REACTIONS TO PROFESSOR HITE'S FOLLOW-UP STUDY

I have reviewed the Follow-Up Study, Long Term Effects of Modified Internship for Beginning Elementary Teachers. I can add little to Herb's discussion of the data. His analysis of the limitations of the study as a basis for conclusive proof of the hypothesis that the three experimental treatments in the "modified internship" of 1965-66 do have significant long term effects extending into the second year of teaching is sound.

In paragraph 1, page 3, of his discussion, Herb points out that the overall mean performance score of each of the four groups of ten in January 1966 was different from the original ranking of mean score of the total membership of each group at that time. It was interesting to compare the rankings of the four groups of ten each in January 1966 and in the spring of 1967.

1966 Ranking--Mean Overall Score		1967 Ranking--Mean Overall Score		Difference
Group II	3.66	Group II	3.32	- .34
Group III	3.03	Group III	3.27	+ .24
Group I	2.30	Group IV	3.08	+ .95
Group IV	2.13	Group I	2.85	+ .55

The two bottom groups in 1966 made the greatest comparative gain in the ensuing year. Yet Groups II and III maintained their top ranking, even though Group II dropped in mean performance -.34, and Groups IV and I made the greatest jump in mean performance, .95 and .55 respectively. As Herb points out, if the five atypical teachers, two in Group IV, one in Group I, one in Group II, and one in Group III are not considered, the comparative rankings would be about the same for 1966 and 1967, even though all groups were closer together. It would have helped in the follow-up study if each group of ten had been a representative sample of the January 1966 distribution of scores within each experimental group. The regression phenomena, the tendency of mean scores of subgroups to move in subsequent measures toward the mean of the total population may account in part for the above changes.

The kind of inquiry expressed by Dr. Hite's studies, in which your organization has contributed some needed encouragement, should be continued, if we are to identify and put into practice those treatments which really facilitate the long term individualized career growth of our faculties in the years ahead. This inquiry has to be collaborative, involving cooperative effort by at least these agencies in this state:

(1) The individual teacher.
(2) The colleges which offer pre-service and graduate training.
(3) The employing district.
(4) The professional organizations in various teaching domains.

This collaboration will take time and money. Predictably, the investment will pay rich dividends in the returns (educational productivity) from our investment in classroom leadership. Perhaps the SIRS program will play a key role in the collaboration. The TTT program and the recent State Department of Education studies of teacher certification improvements are also promising.

As we continue the inquiry I have some suggestions which, if implemented, predictably will yield better results than as yet achieved.

The first: Improvement of our criteria, our statement of those specific teaching competences which define the general practitioner expectations of the teacher in classroom leadership and later additional tasks expected in various career specializations (example, the training of a teacher in a specific domain).

The second: Improvement of tools and methods to carefully, systematically appraise criteria performance. For example, the performance appraisal guides used in the two Hite studies, although I feel they were very useful compared with other available measures (I admit I am biased because I designed these tools) have weaknesses. For any specific task we need to develop a scale which yields performance scores considerably more sensitive than a seven point spread, a range in scores from, say, 0 to 50, in which even the master teacher would rarely achieve above 40 points. We need to establish on such scales minimal qualifying scores for beginning certification, which leave a considerable range to recognize future growth in the skill. As we develop such scales, it is highly important that master teachers in various subject matter and grade level domains be active participants.

For each skill or criterion, we need to define much more clearly an adequate sample of the teaching upon which to base an appraisal of the skill. In the Hite studies we used as an adequate sample one short observation of a lesson.

As Dr. Hite has so well pointed out, this one lesson may not be typical. Perhaps four or five lessons would be barely adequate. For some skills, particularly those involving preparation and evaluative tasks, perhaps a conference with the teacher should be part of the appraised sample.

The third: Related to the above, we need to include in our appraisal data the self-appraisal of the individual teacher, his own self-perception of how well he demonstrates a specific criterion. It would have been interesting, indeed, to get such data in a repeat of the Hite studies, or even to use as performance scores data which an observer and the teacher develop together. The use in the Hite studies of experienced elementary teachers as observers and analyzers of teaching is, I believe, a constructive development which should be continued with the collaboration of the professional associations of experienced specialists. I anticipate the possibility in the future that any teacher can call on a team of fellow specialists to assist him in making a careful analysis and diagnosis of his own practice in any given teaching situation.

The fourth: Long term, longitudinal studies, covering periods of three to five years of career development, are highly desirable. Only in this way can we get some answers to such important questions as explaining atypical cases such as Dr. Hite described, and identifying causes of costly teacher turnover and drop out from the profession. In such studies we need to identify and explore the effect of many context factors, beyond the individual teacher's control, which have positive or negative effects on his teaching competences. Among the more important of such context factors are these:

Differences in school buildings, facilities, technical obsolescence.

Differences in district programs and investments in special services, guidance, libraries, instructional materials, teacher aide and clerical staff, inservice training, department leadership patterns.

Differences across building faculties in teacher load, daily preparation, faculty teaming and communication systems, innovative climate, variation in experiences of staff, turnover.

Differences in student groupings, socio-cultural variations in the communication served by school buildings, parent-teacher, and teacher-community agency communication patterns.

Differences in teachers' outside-of-school roles and responsibilities.

Differences between teaching role expectations held by the teachers themselves and those held by parents and students served by the schools.

Differences between teaching role expectations accepted by college faculties, academic or departments of education, and role expectations of the district and practicing specialists at various grade levels.

Differences in salary schedules and compensation patterns for added responsibilities, extra-curricular, or non-teaching roles assumed by individual teachers.

Differences in the outside-of-school communication experiences of learners which increasingly influence their educational development inside the school, particularly variations in recreational and mass media habits, unique socio-cultural characteristics of their environment.

One additional need, as we look to the future of continuous career development of our teachers, focuses on the importance of obtaining, preferably from individual teachers, information as to what action on their part was taken, if any, to use or apply an appraisal of a specific teaching competence in subsequent career development, whether it was an appraisal of the teacher's own performance or of another's. Such information might help to identify useful career development strategies, might help to explain atypical cases of career growth.

The above verbiage has probably been more confusing than helpful. I do believe we are making definite progress in this state toward solving a very complex problem of nation-wide significance. In spite of the lack of conclusive evidence, Dr. Hite's studies point in the right direction. Hopefully we can continue this exploration.