DOCUMENT RESUME

ED 039 200 SP 003 885

AUTHOR Overing, R. L. R.
TITLE Research Related to the Evaluation of Graduates of Teacher Preparation Programs.
PUB DATE 70
NOTE 19p.; Paper presented at annual meeting, AERA, Minneapolis, 1970

EDRS PRICE EDRS Price MF-$0.25 HC-$1.05
DESCRIPTORS *Elementary School Teachers, Evaluation Criteria, Evaluation Needs, Experimental Programs, *Program Evaluation, *Teacher Evaluation

ABSTRACT

This review makes use of studies evaluating teacher education graduates against internal criteria, i.e., objectives specified in the program, and external criteria or evidence of pupil change. The Recommended Standards (see SP 003 720) of the AACTE indicate that such studies are necessary for meaningful evaluation. The literature search, principally through ERIC and "Psychological Abstracts," produced some 200 references. There appear to be no large-scale studies of the extent to which graduates acquired the characteristics intended by the program, but this may be remedied by the USOE-sponsored Elementary Teacher Education Programs. The University of Missouri published a report in 1967 devoted largely to evaluation, but this gave no evidence that graduates reflected the objective criteria of the program in their teaching. An experimental program by Sandefur et al (1967) showed significant behavioral differences, while a similar study by Corle (1967) of inservice training by means of a 15-week ETV program showed little significant difference between the experimental and control groups. No studies could be found evaluating the teacher preparation program against pupil achievement. The question of whether we have the means and techniques to evaluate teacher preparation programs needs to be answered, and the parameters of teacher effectiveness must be defined, possibly by means of numerous small studies which would increase the fund of information needed for a major survey. (MBM)

Research Related to the Evaluation of Graduates of Teacher Preparation Programs

R.L.R. Overing
University of British Columbia

The Recommended Standards for Teacher Education (1969) includes five categories of standards. The fifth of these categories has to do with standards for evaluation of graduates, program review, and long-range planning. The review which follows deals only with the topic "evaluation of graduates" and does not touch upon either the topics of program review or long-range planning.

In conducting this review I have searched for studies of the following sorts:

1. Studies evaluating graduates of teacher preparation programs against internal criteria (i.e., using as criteria outcomes of the program specified in terms of teacher behaviors and characteristics)
2. Studies evaluating graduates of programs against external criteria (i.e., using pupil change as a criterion for the evaluation of the graduates' effectiveness)

I have not included in this review studies which attempt to determine experimentally which characteristics of a teacher or a teaching situation interacted with particular learner characteristics to facilitate or inhibit learning. Nor have I included studies which concern themselves with the development and validation of instruments designed to record

---

A paper presented as part of a symposium entitled "An Examination of the Research Base for the Standards for the Accreditation of Teacher Education: Basic Programs," sponsored by the Special Interest Group on Teacher Education, A.E.R.A., Minneapolis, 1970.

teacher behavior or characteristics. Finally, I have not included studies that are concerned solely with developing predictors of success in the training program or in subsequent teaching. I have not excluded studies of these sorts because they are of no concern or of secondary importance. On the contrary, I consider such studies to be of fundamental importance in that, ultimately, they must provide both the empirical basis on which we build programs of teacher education, and the instrumentation for selecting and for evaluating our graduates. Such studies, however, must be omitted from this review for it is not my charge at this time to consider <u>directly</u> the vast area of research on teacher effectiveness. My concern, as I have attempted to make clear above, is to search for studies where 1) either specific objectives were formulated and a serious attempt was made to evaluate the program and/or the graduates, using the objectives as criteria, or 2) the graduates of a program were evaluated using the achievement of their pupils as a criterion. Using criteria such as these it is obvious that I will also omit from this review any mention of evaluation studies which employ as their principal source of data the <u>opinions</u> of the graduates of the program. Studies of this sort seem frequently to be conducted by institutions engaged in training teachers. The ones that I have encountered strike me as being of little use as a main source of data for future decision making.

I have restricted the range of studies reviewed to the two classes noted above for two principal reasons. First, I believe that evaluation

studies of these sorts, rigorously pursued, are the ones most likely to advance both our understanding of the nature of an effective training program and our knowledge of the technology necessary to design, to describe and to evaluate improved training programs. That such studies are likely to be uncommon is suggested by Stiles and Parker (1969) who state, "Evaluation of entire teacher education programs, or even of segments of programs, is spotty and inadequate". (p. 1418). Second, the Recommended Standards (A.A.C.T.E., 1969), themselves, (see Sec. 5, p. 12) identify evaluation studies of these sorts as the ones which they recommend and hope to promote. The authors state, "The ultimate criterion for judging a teacher education program is whether it produces competent graduates who enter the profession and perform effectively" (p. 12). And a few lines further, they state, "Any effort to assess the quality of the graduates requires that evaluations be made in relation to the objectives sought. Therefore, institutions use the stated objectives of their teacher education programs as a basis for evaluating the teachers they prepare." (p. 12).

Consideration of these two statements makes it apparent that two quite different criteria are being advocated and we know that it is quite possible that these two criteria may be independent of one another. That is to say, the stated objectives of the teacher education program may bear no relationship to effective teaching. Hopefully, this is not so. Nevertheless, we must always ask of any program that specifies its

objectives, "What are the grounds for these objectives? Which objectives have a hypothetical basis, which have an analytic basis, and which have an empirical basis?" Questions of this sort relate to problems of criteria. The fact that I have restricted my remarks to two kinds of criteria, namely a) specified teacher behaviors and characteristics and b) pupil change, does not mean that there are not, or cannot be, other criteria. This whole problem of criteria is obviously of fundamental concern in any attempt to evaluate graduates of teacher training programs. Nobody should embark on such a venture without being thoroughly familiar with at least, the reports of AERA (1952,1953); Rabinowitz and Travers (1953); Morsh and Wilder (1954); Mitzel (1960); Ryans (1960); Barr (1961); Ryans (1967), all of which attempt to cast some light on this ~~interior~~ criteria problem.

In passing it may be of interest to note that my search of the literature was made principally within the ERIC Indexes, 1965-68 and within Psychological Abstracts, 1960-68. The rubrics used within ERIC were Evaluation, Evaluation Criteria, Evaluation Methods, Evaluation Needs, Evaluation Techniques, Program Evaluation, Teacher Evaluation, Teacher Proficiency, Teacher Rating, Effective Teaching, Teacher Education Curriculum, Teacher Behavior, Teacher Certification, Task-Performance, Observation, Behavior Change, Professional Education, Professional Training, Objectives, Measurement Techniques, Preservice Education.

The rubrics used within Psychological Abstracts were Job Evaluation, Evaluation, Teacher Training, Training.

In all I followed up some 200 references, which from their titles seemed appropriate. That the elephant labored and gave forth a mouse will quickly become apparent as I read on.

Studies Evaluating Graduates of Teacher Preparation Programs Against Internal Criteria

Evaluation studies of graduates of teacher preparation programs which use specified objectives of the program as criteria require two general components, viz., first, a set of specified objectives describing the abilities, the characteristics and dispositions which graduates of the program are expected to exhibit; second, a set of instruments and techniques for measuring the extent to which graduates of the program exhibit these abilities, characteristics and dispositions. To the extent that we may also wish to say that the abilities, characteristics and dispositions exhibited by the graduates are due to the effects of the program we will also have to have a set of instruments and techniques to obtain pre-measures of these same graduates when they entered the program. But that is a slightly different question which need not concern us directly here. However, we should keep in mind that evaluations of program effectiveness as contrasted with evaluation of graduates of this program may have to use this pre-test, programming, post-test model.

Large scale studies which actually have attempted to determine the extent to which graduates of a teacher preparation program have acquired the behaviors and characteristics described in the program objectives are rare, and in any pure form, seem to be non-existent. Thier frequency may increase, however, for the recent USOE-sponsored Elementary Teacher Education Programs have all been formulated around the central idea of specified teacher competencies (Fattu, 1968). For example, Dickson et.al. (1966) have listed 618 program objectives, formulated in terms

of specific teacher behaviors and, in what is frequently a very general manner, have also described how participants in the program will be evaluated to determine if they have met the criteria. The description of the evaluation techniques is general in the sense that frequently there is no mention of the specific instruments and techniques by which the evaluation will be carried out. As the design and validation of such instruments is normally a demanding, lengthy and expensive task we should recognize the significance of this lack of specificity. Nevertheless, the availability of teacher preparation programs built around specified objectives presumably means that the attempt will now be made to evaluate the extent to which these objectives have been attained.

One study, though by no means a model, may suggest something of the state of the art and of the problems still to be solved. The Final Progress Report of a Ford Foundation-sponsored teacher education project carried out at University of Missouri at Kansas City and published in 1967 is devoted largely to evaluation. While evaluation of several sorts was attempted, only those parts of the evaluation study which concerne themselves with certain pre-specified verbal behaviors of the graduates approximate the type of evaluation study here under review.

Graduates of the program were evaluated during their first year of teaching, to see if their teaching behaviors reflected the specific objectives of the part of their program which had dealt with the teaching of cognitive behaviors. This program component had attempted to train them to teach so as to give particular emphasis to higher level behaviors. Specifically, audio-tapes were made of two lessons for each of a group

of Experimental teachers and each of a group of Controls (total N = 40). These tapes were then analyzed to determine 1) the percent of teacher verbal behavior which fell into each of Bloom's categories for the Cognitive Domain, and 2) the number of pupil responses induced by teacher questions. No significant differences were found. However, when the Experimental group was divided in two,to form a group with high academic achievement and a group with low academic achievement, significant differences between certain of the sub-groups of these High and Low groups emerged, favoring the High group. With the exception then of these small sub-groups, there was no evidence that graduates of the program were teaching in a manner to reflect the objective criteria of the program. Whether the n.s.d. results are due to lack of treatment difference or to reliability and sampling problems is not apparent.

While I was able to locate no other large scale studies which attempted to evaluate their graduates against internal criteria, there are two studies which I would like to mention in this section. In both cases the behavior of the graduates of the program was measured, but in neither case were there explicit pre-specified program objectives against which the behavior measured could be evaluated. Sandefur et.al. (1967) devised an experimental program which attempted 1) to identify and to organize knowledge related to teaching and learning; 2) to design and to implement a series of laboratory experiences; and 3) to evaluate the extent to which teacher behavior was affected. Essentially, they attempted to coordinate lab experiences allowing observation and participation with appropriate readings, and to conduct the whole program in a relatively informal, non-threatening, seminar context.

Sixty-two members of this experimental program were then compared with fifty-two members of a conventional program within the same institution. Data on classroom behavior were collected during student sessions using Ryan's Classroom Observation Record (Ryan, 1960, pp. 83-92) and Hough's Modification of Flanders' system of interaction analysis. Additional data were collected using student-teaching grades and the National Teachers Examination. Hypotheses looked for differences in teacher behavior, teaching patterns, pupil behavior, student-teaching grades and professional knowledge. In all categories except professional knowledge, as measured by the National Teachers Examination, student teachers from the Experimental Group and the pupils under their direction showed significant differences in the direction of behaviors generally held to be desirable. For example, Experimental teachers showed significantly more use of behavior which could be categorized as praise, acceptance and use of pupil ideas, student talk, demonstration, etc. Their pupils were judged more alert, responsible, initiating, fair, democratic, etc., etc. Thus, while no program objectives had been pre-specified, the program designers were prepared to say that the classroom behavior of participants was of the sort which they wished to produce by their program. In a sense, the "desirable" and the "undesirable" behaviors which the instruments were designed to record provided an implicit set of behaviors to serve as objectives of the program. Obviously, it would be a relatively simple matter to make these objectives explicit.

While there may be limitations to this approach, it seems not a bad idea for program designers to concern themselves with behaviors for which there already exist measuring instruments of some demonstrated reliability and validity. Approximately eighty such direct observation instruments and techniques are summarized in the Simon & Boyer anthologies (1968,1970)

A second and somewhat similar case is provided by Corle (1967) who compared sixteen intermediate mathematics teachers who received in-service training via a 15 week ETV program and 16 who did not view the program. Ss were visited seven times before the in-service training began and 23 times during the program. Behavior was recorded on a modification of Medley and Mitzel's OSCAR, designed for elementary mathematics classrooms. Only one behavior category of the six recorded showed a significant difference in favor of the Experimental Group. Lack of feedback, lack of shaping and short duration of the training program are given as possible reasons for the lack of behavioral change evident. However, the point I wish to make is that while the author had no pre-specified objectives for his program, he was prepared in his discussion section to judge certain of the behavior categories of the OScAR (EM) as more or less desirable and to imply that his course was successful to the extent that it moved teachers towards these desirable categories. Thus, he, like Sandefur, was using the behavior categories of his instruments as the implicit objectives of his program.

Studies Evaluating Graduates of Teacher Preparation Programs Against External Criteria

I was unable to locate any studies whatsoever which evaluated graduates of a teacher preparation program against the criterion of pupil achievement. Studies attempting to use this "ultimate" criterion of pupil achievement are still small scale and concerned with developing criterion instruments or concerned with mapping teacher behavior in order to identify significant teacher variables. The study which came closest to a headlong assault on the problems surrounding the use of pupil achievement as a criterion of teacher effectiveness was that reported by Popham and Baker, (1965), and Popham (1967). This study attempted to determine if teachers who differed greatly in terms of experience and training would be differentially effective in promoting pupil change. The underlying purpose of the study was to validate a test of teacher effectiveness using pupil achievement as a criterion. The study directors, building on the observation of Turner and Fattu (1960) that the relative effectiveness of teachers could be judged only when they were attempting to teach to the same objectives, provided teachers with a set of instructional objectives, à la Mager, suggested a variety of means to teach these objectives, spelled out the subject matter content and, finally, provided a pre and post-test which the participating teachers neither saw nor administered. In the hope of obtaining differences between teachers, two apparently very different groups were formed, one consisting of trained teachers who 1) had

received A in a curriculum and instruction course emphasizing the construction and use of behavorial objectives, 2) had social studies majors and, 3) had been judged superior by their supervisors. The other group was made up of housewives who 1) had had no formal teaching experience or teacher training, 2) had at least two years of college and, 3) had been enrolled as social studies majors. There were no significant differences whatsoever between the achievement scores of the pupils whether taught by the experienced teachers or the inexperienced teachers. Nor were there any differences in attitudes expressed by the pupils, nor did the teachers themselves differ in their reactions to the materials, the objectives etc. which were provided for them.

Popham suggests that the principal reason explaining why there were no differences in pupil achievement may be that "experienced" teachers are no more experienced than intelligent lay people in bringing about change in pupils. This is not to say that the trained teachers do not possess certain specialized skills and knowledge. It is just that this skill and knowledge does not seem to be particulary related to pupil change.

I have dwelt at some length with this study, even though it does not specifically set out to evaluate graduates of a program, for two reasons. First, I have been able to locate so little else to report, and, second, I have wished to emphasize for you the complexity of the problem of evaluation which we are considering. Popham is an extremely imaginative, intelligent researcher who spent a lot of time, and devoted

a lot of resources to design a test which would discriminate between teachers. To increase the likelihood of his obtaining differences he took two apparently very different groups of teachers. Despite these efforts he was able to detect no differences. If nothing else this suggests that there are no simple-minded easy solutions to the problem of evaluating graduates of programs using pupil achievement as the criterion.

Conclusions

I am afraid that this paper advances our understanding of the nature and problems of evaluating graduates of teacher preparation programs very little. Perhaps it will be of some use if it brings to our attention the fact that while many writers have advocated the approach to evaluation now suggested in the Recommended Standards, almost no one has attempted it. Some writers (eg. Woodruff, 1968) believe we are right on the edge of being able to evaluate our products satisfactorily. Woodruff writes, "It is doubtful that we could have taken this direction (i.e., the evaluation of program products) earlier with any realistic chance for success, but I am convinced we can do so now, and indeed that we must for the sake of professional responsibility". (p. 245).

Fattu (1968) however, raises the question of whether all components necessary for an invention (in our case the means and technology of product evaluation) are available to the people trying to do the inventing. For example, do we have any reasonably satisfactory set of criterion behaviors around which to design our programs and against which to evaluate our graduates? Dickson et.al. (1968) states "What a teacher does as he performs his tasks must be determined before the knowledge and experience needed in developing these teaching skills can be ascertained". (p. 90). We need to ask ourselves to what extent the significance of the various teacher behaviours which are offered as program objectives has been empirically determined and to what extent their significance is merely conjectured.

The Recommended Standards state that it is recognized that the means now available for making such evaluations (i.e. the evaluation of program products) are not fully adequate. This may turn out to be the understatement of the year. There is no doubt that much rigorous and imaginative basic research is being done in the area of program product evaluation. For example, McGuire (1968, b) writes, admittedly in the context of medical education, that

> products of medical education are being studied by systematic evaluation procedures which include: empirical determination of essential components of professional competence, employment of simulation techniques to supplement more conventional methods of assessment, application of pre-established standards, and utilization of numerous feedback mechanisms to assure fuller exploitation of evaluation data. Such evaluation studies are being employed not only to assess individual achievement of critical performance requirements, but also to identify differential rates and patterns of progress toward these goals, to determine the relation between these patterns and important independent variables in the learning situation, to guide curricular development, and to provide evidence of value in redefining the goals themselves. (p. 51)

Some of these same kinds of studies, only focussing on teacher education, are undoubtedly being attempted right now. All of them are being advocated. A balanced set of the kinds of studies listed by McGuire, above, actually would contain all the sufficient and necessary components for the evaluation of program graduates. But the very fact that research and developmental-type studies are being undertaken which focus on individual components of the evaluation process, serves to raise the question, "Have we as yet the means and techniques to conduct evaluation of teacher preparation programs of the sort advocated in the

Recommended Standards?" My feeling is that we do not, despite the fashionability of the product evaluation approach. Most of us have underestimated the difficulty of such an approach and have ignored the conceptual and measurement problems which remain to be solved. Two of the most sobering reminders of this are expressed by Travers in two papers (1966, 1968), one dealing with the nature of theory building, and the other with some problems of the product-oriented approach to instruction and evaluation.

In summary, it seems to me that examples of the problems which must be solved before we can begin to attempt, with any hope of success, to evaluate the graduates of programs of teacher educations are of the following classes.

1. Problems of criteria: e.g., Which behaviors and characteristics of teachers are going to be specified as the proposed outcomes of the program against which the graduates will be evaluated? Which characteristics and behaviors of pupils will be measured to determine teacher effectiveness?
2. Problems of criterion relevance: e.g., What is the evidence that the criterion behavior specified in the outcome is relevant to the teaching task, and has utility in facilitating learning, and is practical in the real world of teaching? With which situational and pupil characteristics does it interact?

3. Problems of measurement: e.g., For which classes of teacher and pupil behavior and teacher and pupil characteristics have we reliable and valid measurement instruments and for which have we not? If we attempt measurements of natural settings (ongoing teaching-learning) as opposed to measurements of constructed "artificial" settings, how can we decrease the liklihood of sampling error?

All of these and other similar problems actually are problems for research in teacher effectiveness. The evaluation studies which are attempted can only be as good as the research basis on which they rest. And what can we say of this research basis? Biddle (1964) states unequivocally (p. 3), "we do not know how to define, prepare for or measure teacher competence". Farther on in the same work he writes (p. 12), " ...a general classification of teacher behaviors appropriate to the study of effectiveness has not been advanced - nor does it seem likely that a satisfactory system will be produced in the next decade." Flanders (1969), in contrast, in a review based largely on his own and other related work concludes that empirical cause-effect relationships exist between certain characteristics of teachers and pupil change and that adequate instrumentation is available to permit measurement of these characteristics on a large scale. Travers (1968), however, in what is, unfortunately, merely a passing reference to studies using interaction analysis, questions the extent to which we can use their results as a basis for constructing training programs.

I do not wish to belittle the import and direction of the Recommended Standards. Nor do I wish to discourage others here to attempt to undertake product evaluation studies. But I hope that teacher educators who may have jumped on a bandwagon will recognize that at the moment the product evaluation movement is mostly just talk and that a tremendous amount of research and development awaits us before we will have licked this problem. If this is so, I believe our strategy should be to attempt many, many, reasonably small studies each of which attempts to increase the fund of knowledge and the supply of instruments and techniques. Only in this way will we secure a better foundation for the design and evaluation of teacher education programs than presently exists.

References

AERA, Report of the Committee on the Criteria of Teacher Effectiveness, Review of Educational Research, 1952, 22, 238-263.

AERA, Second Report of the Committee on the Criteria of Teacher Effectiveness, Journal of Educational Research, 1953, 46, 441-458.

American Association of Colleges of Teacher Education, Recommended Standards for Teacher Education, Washington, D.C., 1969.

Barr, A.S., The Measurement and Prediction of Teacher Effectiveness, Madison, Wisconsin, Dembar Publications, 1961.

Biddle, Bruce J., "The Integration of Teacher Effectiveness Research", Contemporary Research on Teacher Effectiveness. New York, Holt, Rinehart and Winston, 1964.

Corle, Clyde, Mathematics Teaching Behavior Changes made by Intermediate Grade Teachers During a 15 Week Period of Instruction by ETV. Final Report, Pennsylvania State University, Univ. Park, 1967.

Dickson, George E. et.al., Educational Specifications for a Comprehensive Elementary Teacher Education Program: Vol. II, The Specifications. Final Report. 1968. Consortium of the State Universities of Ohio; Toledo.

Fattu, Nicholas S., Nine Proposals for Elementary Teacher Education., a Description of Plans to Design Exemplary Training Programs. OE, DHEW, Washington, D.C., 1968.

Flanders, Ned, Teacher Effectiveness, Encyclopedia of Educational Research, 4th Edition, New York, Macmillan, 1969.

Hough, John B., An observational system for analysis of classroom instruction, AERA, 1965.

McGuire, Christine H., "Testing in Professional Education" Review of Educational Research., 1968, 38, 1, 49-60.

McGuire, Christine H., An evaluation model for professional education - medical education; Proceedings of the 1967 Invitational Conference on Testing Problems, ETS, Princeton, N.J., 1968,b.

Mitzel, H.E., Teacher Effectiveness, Encyclopedia of Educational Research. 3rd Edition, 1960, Macmillan, New York.

Morsh, J.E. and Wilder, E.W., Identifying the effective instructor: a review of the quantitative studies, 1900-1952. San Antonio, Texas. U.S. Air Force Personnel and Training Research Centre Research Bulletin, 1954, No AFTPRC-TR-54-44.

Popham, W.James, Development of a performance test of teaching proficiency, Final Report, UCLA, Rep. No. BR5-0566, August, 1967. 172 p.

Popham, James and Baker, Eva. A Performance test of teaching effectiveness, AERA, Chicago, 1966.

Rabinowitz,W. and Travers, R.M.W., Problems of defining and assessing teache effectiveness. Ed. Theory, 3, 212-219, 1953.

Ryans, David G., Criteria-problems in validating teacher selection policies and procedures. N.Y.C. Board of Education, Brooklyn, N.Y. June, 1967.

Ryans, D.G., Characteristics of Teachers. Washington, D.C. American Council on Education 1960.

Sandefur, J.T. et.al., An Experimental study of professional education for secondary teachers, Final Report, 1967, Kansas State Teachers College, Emporia.

Simon, A. and Boyer, E.G.(eds) Mirrors for Behavior: An anthology of classroom observation instruments. Vol. 1-6, Philadelphia, Penn., Research for Better Schools, Inc. 1968.

Simon, A. and Boyer, E.G. (eds) Mirrors for Behavior: An anthology of classroom observation instruments, continued. Vol.7-15, Philadelphia, Penn., Research for Better Schools, Inc., 1970.

Stiles,L.J. and Parker,R.P., "Teacher Education Programs" Encyclopedia of Educational Research, 4th Edition, New York:Macmillan, 1969.

Travers, R.M.W., Some deficiencies of the Industrial production model when applied to education. AERA, Chicago, 1968..

Travers, R.M.W., Review of Conceptual Models in Teacher Education, by John Verduin, Washington, D.C., AACTE, 1967.

Travers, R.M.W., Towards taking the fun out of building a theory of instruction, Teachers College Record, 1966, Vol. 68, 1, 49-60.

Turner, R.L. and Fattu, N.A., Skill in teaching, a reappraisal of the concep and strategies in teacher effectiveness research; Bulletin of the School of Education, Indiana University, Bloomington, 1960, 36, 3.

University of Missouri at Kansas City, Final Progress Report, The Teacher Education Project of the School of Education, UMKC, Kansas City, 1967.

Woodruff,A.D. Accreditation by evaluation of the product: What it means for teacher education programs. American Association of Colleges of Teacher Education, Yearbook (1968) 21; 237-245.