Following the acceptance and implementation of the new National Standards for Accreditation of Teacher Education, the American Association of Colleges for Teacher Education prepared this model for evaluating teacher education graduates as an aid to education institutions. The paper briefly reviews research on teachers and evaluation and places the research into three thematic clusters: teaching methodology, effective climate of the classroom, and characteristics of teachers. On the basis of generalizations drawn from this research, the model suggests that the proposed evaluative data can be derived from four sources: career line data (systematic collection of data on wastage from teaching, promotions, advanced degrees, writing, and research); direct classroom observations (two suggested systems: the Classroom Observation Record and a 14-category modification of the Flanders system and Hough variation of interaction analysis); student, peer, and supervisor ratings (for example, the Student Evaluation of Teaching developed by Heldman and Peck); and standardized measures. The document closes with recommendations for utilizing the model and a discussion of the cost. (JA)
An Illustrated Model for

THE EVALUATION OF TEACHER EDUCATION GRADUATES

by J. T. Sandefur

Dean, Graduate College, Western Kentucky University

for the AACTE

Commission on Standards

American Association of Colleges for Teacher Education
One Dupont Circle
Washington, D. C. 20036
AACTE COMMISSION ON STANDARDS MEMBERS

CHAIRMAN: Margaret Lindsey, Professor of Education, Teachers College, Columbia University, New York, New York 10027.

Frederick R. Cyphert, Dean, College of Education, University of Virginia, Charlottesville, Virginia 22903.


Herbert Hite, Chairman, Department of Education, Western Washington State College, Bellingham, Washington 98225.

J. T. Sandefur, Dean, Graduate College, Western Kentucky University, Bowling Green, Kentucky 42101.

STAFF:

Karl Massanari, Associate Director
Shirley Bonneville, Program Assistant
Brenda Greenhowe, Secretary

Published by the
AMERICAN ASSOCIATION OF COLLEGES FOR TEACHER EDUCATION
One Dupont Circle, Washington, D. C. 20036

September, 1970

Standard Book Number 910052-69-7
Library of Congress Catalog Number 72-91036
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preface</td>
<td>v</td>
</tr>
<tr>
<td>1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>II. The Research on Teaching and Evaluation</td>
<td>3</td>
</tr>
<tr>
<td>Research-supported Generalization on Teaching and Teachers</td>
<td>4</td>
</tr>
<tr>
<td>III. A Proposed Model for Evaluating Teacher Education Graduates</td>
<td>11</td>
</tr>
<tr>
<td>Career Line Data</td>
<td>12</td>
</tr>
<tr>
<td>Direct Classroom Observation</td>
<td>12</td>
</tr>
<tr>
<td>Pupil, Peer, and Administrator/Supervisor Evaluation</td>
<td>17</td>
</tr>
<tr>
<td>Standardized Measures</td>
<td>23</td>
</tr>
<tr>
<td>Summary</td>
<td>28</td>
</tr>
<tr>
<td>Recommendations</td>
<td>30</td>
</tr>
<tr>
<td>IV. Projected Costs of the Proposed Model</td>
<td>33</td>
</tr>
<tr>
<td>Notes</td>
<td>35</td>
</tr>
<tr>
<td>Bibliography</td>
<td>40</td>
</tr>
<tr>
<td>About AACTE</td>
<td>43</td>
</tr>
<tr>
<td>Order Form for Recent AACTE Publications</td>
<td>45</td>
</tr>
<tr>
<td>Publication Order Form for PBTE Papers</td>
<td>46</td>
</tr>
</tbody>
</table>
Preface

Following acceptance and implementation of the new national Standards for Accreditation of Teacher Education, many institutions requested assistance from AACTE and NCATE regarding appropriate application of the Standards. Most requests centered on the effective application of the Standards which concern institutional efforts with respect to evaluation, program review, and planning. The particular aspects of the Standards that seemed to offer difficulty dealt with the requirement for a well-defined institutional plan for evaluating the teachers being prepared and with explicit feedback and utilization of evaluation results in the improvement of programs.

In response to such requests, the Commission on Standards decided to sponsor the development of descriptive case studies or alternative strategies to focus upon the application of these Standards. This paper represents the first of such studies. It is anticipated that additional illustrative approaches to the problem of evaluating teachers and the use of feedback techniques will be forthcoming at a future date.

J. T. Sandefur, dean of the Graduate College at Western Kentucky University, generously undertook to develop this initial paper. The charge to Dean Sandefur was to survey relevant literature and to illustrate an operationally feasible approach to the evaluation of teacher education graduates. The plan described here is Dean Sandefur's response to the Commission's request. Currently the project he explains is being applied at Western Kentucky University.

Dean Sandefur would be the first to acknowledge that the Western Kentucky University plan does not represent a comprehensive model for the evaluation of teacher education graduates at all institutions. It is, as the title of the paper suggests, but one illustration. Unquestionably, there are many different ways in which one can approach the entire problem of evaluating graduates of teacher preparation programs. It will be apparent to many, for example, that the research and the plan summarized by Dean Sandefur emphasize affective dimensions of teaching. In other plans, conceivably, greater attention would be given to the assessment of cognitive aspects of teaching performance, or perhaps different evaluative strategies would be employed.

AACTE is indebted to Dean Sandefur for his response to the Commission's "assignment." The Association also wishes to express its gratitude to the several critical respondents who offered insights and comments on the first draft of this paper. Readers are invited to offer their reactions or to report alternative approaches to the critically significant problem of how to evaluate graduates of teacher education programs and how to use
evaluation feedback for the improvement of such programs. We hope this serves as an example of a systematic, practical approach that will, in turn, stimulate the development of many alternatives useful for the improvement of teacher education.

Edward C. Pomeroy  
Executive Director, AACTE

Margaret Lindsey  
Chairman, Commission on Standards  
Professor of Education,  
Teachers College, Columbia University
I. Introduction

A major and continuing problem of teacher education is the evaluation of its "product"—namely, the teacher. It is trite to point out that industry spends a significant portion of the overall cost of a product on its evaluation. No industry could long survive in a competitive marketplace without a systematic, valid procedure for testing how well its product performs the functions for which it was designed. Teacher education is analogous to industry in that it needs to evaluate the "product" and to feed that evaluative information back into the program of preparation in order to improve the quality of the teaching profession.

The new Standards for the Accreditation of Teacher Education, adopted in January 1970 and made mandatory in the Fall of 1971, have focused the attention of NCATE (National Council for Accreditation of Teacher Education) and of institutions of teacher education squarely upon the problem of evaluation.

Standard 5.1 is prefaced by this statement:

The ultimate criterion for judging a teacher education program is whether it produces competent graduates who enter the profession and perform effectively. An institution committed to the preparation of teachers engages in systematic efforts to evaluate the quality of its graduates ... when they complete their programs of study, and after they enter the teaching profession.1

Standard 5.1 reads: "The institution conducts a well-defined plan for evaluating the teachers it prepares."2

During the 1970-71 academic year the use of the new Standards was optional. More than thirty institutions, however, chose to use them in their bid for initial accreditation or reaccreditation from NCATE. In the process of evaluating these institutions by the new criteria, it became evident that none of them used a systematic approach to the evaluation of their graduates which complied totally with the spirit and intent of Standard 5.1.
The search for a systematic approach to the evaluation of graduates of teacher education programs is thus a major concern of institutions preparing for NCATE accreditation visits. How effectively these institutions are able to evaluate their graduates is a concern as well of NCATE visiting teams, and subsequently, of the evaluation boards which must ultimately either recommend or deny accreditation to NCATE.

Evidence that Standard 5.1 is of major importance to evaluation boards was found by the author, who observed an evaluation board in July 1971. Of twelve institutional cases reviewed by the evaluation board, and recording 288 questions asked by the board of the institutional representatives, the author found that 25 questions, representing 12.75 percent of the total, were directed at Standard 5.1, thus ranking first in the total number of questions asked.

It is evident that teacher education institutions have largely ignored the evaluation of their graduates. This failure has been due primarily to the profession's inability to determine what constitutes effective teaching, and partly to the lack of evaluative tools and techniques with which to measure effective teaching. Fortunately, both of the conditions which discouraged evaluation have been at least partially removed. In support of this position, namely, that the conditions which have prohibited evaluation have been removed and that teacher education institutions must now move ahead with systematic approaches to evaluating their products, two premises are offered:

- A sufficient body of research now exists from which inferences may be drawn, and substantiated, on the characteristics of good teaching and good teachers. The findings of research on teaching and learning form a configuration which is subject to order and can be incorporated into instructional schemata.

- Classroom observational systems and other evaluative tools have been developed which enable educators to assess the teaching behavior in a systematic fashion.
II. The Research on Teaching and Evaluation

Evaluating teacher effectiveness has been the most difficult of all problems faced by the education community. It is not surprising that little has been done in this relatively unexplored area. The diverse opinions of authorities as to what constitutes effective teaching has unquestionably retarded and restricted the development of tools designed for uniform assessment of teaching behavior.

Historically, many new theories of teaching and learning have been advanced and each, accompanied by supportive methodology, was added to those already in existence. As a result, practitioners have had an almost infinite number of unvalidated theories from which to choose models for their teaching behavior. The teaching profession, as a consequence, lacked uniform terminology to describe teaching. Its evaluation and study depended primarily upon the value judgments of the observer.

Prior to 1960 little experimental research had been done in teacher education either to determine the characteristics of good teaching or to measure them. Beginning in the early Sixties, however, the situation changed. Significant amounts of money and research expertise were made available to teacher education. According to Robert Peck, a quantum leap occurred somewhere between 1963 and 1965 in the quality of both the design and the reporting of research, probably as a result of the influx of substantial federal money for graduate training and research in education.4

Peck and Tucker surveyed the research in teacher education conducted during the period of 1955-1971. With appropriate caution they identified 6 themes which "seemed" to emerge from the recent body of research:

1) A "systems" approach to teacher education, often called "instructional design," substantially improves its effectiveness. A good deal of research clustered around three special cases of the general model: training teachers in interaction analysis, microteaching, and behavior modification.

2) Teacher educators should practice what they preach. That is, when teachers are treated in the same way they are supposed to treat their pupils, they are more likely to adopt the desired style of teaching behavior.

3) Direct involvement in the role to be learned, or such close approximations as sensitivity-training laboratories or classroom simulation laboratories, produce the desired teaching behavior more effectively than remote or abstract experiences such as lectures on instructional theory.

4) Using any or all of the techniques just mentioned, it is possible to induce a more self-initiated, self-directed,
effective pattern of learning, not only in teachers but, through them, in their pupils.

5) Traditional ways of educating teachers have some intended effects, but they also have some quite undesired effects.

6) One long-needed methodological advance is beginning to appear in the research: the use of pupil-gain measures as the ultimate criterion of the effectiveness of any given process of teacher education.

If we accept the Peck and Tucker trends or "themes" which have emerged from recent research on teacher education, there is reasonable evidence that the research on teaching and evaluation of teachers forms thematic clusters from which generalizations may be drawn and documented. In designing and offering a model to teacher education for the evaluation of its product, we must first identify those broad generalizations about the characteristics of good teaching and good teachers which may be drawn from the research and which may be documented with some degree of objectivity. Secondly, it is necessary to identify the evaluative tools which offer promise of valid assessment of those characteristics. Despite the problems of such an undertaking, the following pages represent the author's analysis of the research findings relevant to teaching and to the evaluative tools available to teacher education. It is recognized that not all of the research available has been included and, consequently, no claim has been made relative to the completeness of the review of the research.

Research-supported Generalization on Teaching and Teachers

1. Good teaching utilizes maximal involvement of the student in direct experiential situations.

Amidon and Hunter ask "Why do researchers engaged in classroom observation find that teachers are so controlling, restrictive, and inhibiting?" Regardless of the answer, the fact is that they have been found to be so. Flanders found that 70 percent of the verbalization in the average classroom comes from the teacher. Arno A. Bellack and others, in research involving fifteen secondary teachers, found that teachers spoke 72.6 percent of the total verbal classroom discourse. Numerous other research studies have been conducted which dealt with verbal interaction in the classroom. Overwhelmingly, they presented evidence that the typical classroom is dominated by teacher talk.

Contrary to what happens in the typical classroom, available research supports the active involvement of the learner in the instructional process and appears to discourage student passivity and vicarious experience.
insofar as possible. Specifically, the research points out the following sub-generalizations:

- Good teachers attempt to foster problem-oriented, self-directed, actively inquiring patterns of learning behavior in their students.
- Good teachers elicit pupil-initiated talk and allow more pupil-initiated exploration and trial solutions.
- When teachers try to elicit independent thinking from their students, they get it.
- Good teachers involve students in decision-making processes in active, self-directing ways.
- Teachers who are interested in student involvement are less prone to dominate the classroom with lecture and other teacher activities.

It appears evident that a thrust of Flanders' system of interaction analysis and its descendants, which include Hough's Sixteen Category System, 9 and Amidon's Verbal Interaction Category System,10 among others, is to provide a means of assessing the extent to which student involvement is obtained in the classroom as well as to measure teacher influence. Peck and Tucker state that the "teaching laboratories" at many places make this philosophy (student involvement) an explicit part of their practice. In summarizing the research on the subject of self-directed learning, Peck and Tucker quoted supportive data from studies conducted by Peck, Burrell, Fleming and Trione.11

Two studies by the author have provided supportive, although implicit, data that when teachers are prepared to minimize their own direct involvement and encourage student involvement they get it in the form of more responsible and initiating student behavior.12 Perhaps the most complete study to date is one conducted over the past five years dealing with disadvantaged children in Durham, N. C.13 One of the goals of the study was the use of the discovery method in several subject fields utilizing individualized, ungraded, non-competitive instruction. It was found that children so taught increased their independence, assertiveness, and productivity. In addition, they made significant gains in Stanford-Binet IQ scores over a matched control group which declined slightly during the study.

Research by Johns supplied impressive evidence that when teachers solicited student involvement and then used the students' ideas in the instruction, they elicited more thought-provoking student questions.14 Finske, in 1967, found that teachers trained in interaction analysis elicited more pupil-initiated talk.15
Throughout the research dealing with involvement of students, interaction analysis has played a prominent role as a research tool to determine the extent of "indirect" teacher influence. Most systems of interaction analysis, particularly those developed by Flanders, Amidon and Hunter, and Hough, have categories designed to measure "indirect" teacher influence. These categories are, as in the Flanders system, the acceptance of student feeling, the use of praise and encouragement, the acceptance and use of student ideas, and the asking of questions. The "direct" influence categories are lecture, giving directions, and criticizing or justifying authority. A number of experimental studies have used systems of interaction to determine the extent to which teachers used "indirect" influence (expanding the freedom of the student) as opposed to "direct" influence (restricting the freedom of the student). With a surprising degree of agreement, the studies have pointed out that effective teachers use significantly more "direct" influences than do poorer teachers. These findings have led to the second major generalization on what research says about good teaching.

2. Good teaching encourages maximal "freedom" for the student.

In describing the characteristics of teachers who use indirect teaching to promote student freedom, several sub-generalizations can be made:

- Good teachers use significantly more praise and encouragement for the student.
- They accept, use, and clarify students' ideas more often.
- They give fewer directions, less criticism, less justification of the teacher's authority, and less negative feedback.
- They use a relaxed, conversational teaching style.
- They use more divergent questions, do more probing, and are less procedural.
- They are more inclined to recognize the "affective climate" of the classroom and are responsible to student feelings.
- Teachers with low dogmatism scores are more likely to use indirect methods than those with more closed-minded attitudes.

So many studies have been done utilizing some aspect of "direct - indirect" teacher influence that only a representative sampling has been presented here. Volumes could be written on the research and implications of the research in this area.
The author and others conducted a study using 115 preservice secondary teachers to determine whether a specific experimental program of professional education would produce more favorable teaching behaviors than would a more conventional program. Upon finding that the experimental program did indeed provide significantly more desirable teaching behaviors, it was concluded that the desired changes occurred because the instructional process used was characterized by 1) constant efforts to reduce tensions and threats in the classroom, 2) persistent effort to recognize and use principles of good human relations based on a feeling for individual worth and dignity, 3) efforts to assure internal motivation, and 4) constant use of student involvement in the teaching-learning process through problem-solving, free discussions in seminars, and laboratory experiences.16

Ned A. Flanders and Anita Simon, after research on teacher effectiveness for the Encyclopedia of Education, wrote:

It can now be stated with fairly high confidence that the percentage of teacher statements that make use of ideas and opinions previously expressed by pupils is directly related to average class scores on attitude scales of teacher attractiveness, liking the class, etc., as well as to average achievement scores adjusted for initial ability.17

Flanders and Simon reported several studies which supported the position that the acceptance and use of a student's ideas lead to improved student attitude and achievement. Morrison, studying 30 sixth-grade teachers' use of student ideas and comparing pupil-gain in language usage, social study skills, and arithmetic, found a significant relationship.18 LaShier, studying 10 teachers and 239 students in eighth-grade science classes also found a significant relationship.19

Pankratz studied in depth 10 high school physics teachers, 5 of whom were rated as "good" teachers and 5 of whom were rated as "poor" by previously determined criteria. He found that the good teachers used significantly more indirect teacher influence as measured by a 16-category system of interaction analysis.20

There is impressive evidence that by teaching interaction analysis to student teachers and inservice teachers, these individuals became more indirect in their teaching behaviors. Studies conducted by Amidon,21 Sandefur,22 Bondi,23 Finske,24 Hough, Lohhman, and Ober,25 Kirk,26 Parrish,27 and Simon,28 lend supportive evidence to this fact. Peck and Tucker, viewing the problem empirically as well as in the light of research, state that since most classrooms are overwhelmingly dominated by teacher talk, "To propose that teachers be trained to allow somewhat more scope for pupil-initiated exploration and trial solutions of problems seems no more than a modest redressing of the balance."29
Hough and Amidon did significant research which indicated that teachers with low dogmatism scores who were taught interaction analysis became significantly more indirect in their teaching behavior than did open-minded teachers who had not been taught interaction analysis or closed-minded teachers who had been taught interaction analysis.30 Barr reported correlations between teaching success and objectivity which include open-mindedness in 26 out of 27 studies dealing with this variable.51

Open-mindedness and other personal traits lead into the research which has been conducted on teacher characteristics. One of the best-known studies is, of course, David G. Ryans' Characteristics of Teachers.32 This study, along with many which preceded and followed it, has led to a third major generalization about good teachers.

3. Good teachers tend to exhibit identifiable personal traits broadly characterized by warmth, a democratic attitude, affective awareness, and a personal concern for students.

Broad generalizations require more specificity. Therefore, at the risk of limiting the research findings, and in the full awareness that it would be impossible to do more than provide a limited sampling of the available research, the following sub-generalizations have been formulated:

- Good teachers exhibit characteristics of fairness and democratic behavior.
- They are responsive, understanding, and kindly.
- They are stimulating and original in their teaching.
- They are responsible and systematic.
- They are poised and confident, and emotionally self-controlled.
- They are adaptable and optimistic.
- They are well-versed in subject matter and give evidence of a broad cultural background.

Barr, in summarizing the research on teaching competencies, listed the following as factors which correlate positively with teaching success: resourcefulness, emotional stability, considerateness, buoyancy, objectivity, drive, self-reliance, attractiveness, refinement, cooperativeness, and reliability.33 Apparently relying heavily upon the categories of characteristics developed by Barr, the Teachers Characteristics Study, headed by Ryans, determined 18 dimensions of teaching behavior which seemed critical to the investigators. On a bi-polar scale they were: partial - fair, autocratic - democratic, aloof - responsive, restricted -
understanding, harsh - kindly, dull - stimulating, stereotyped - original, apathetic - alert, unimpressive - attractive, evading - responsible, erratic - steady, excitable - poised, uncertain - confident, disorganized - systematic, inflexible - adaptable, pessimistic - optimistic, immature - integrated, and narrow - broad.

The Teachers Characteristics Study identified three teaching patterns which emerged from the extensive data:

TCS Pattern X - Warm, kindly, understanding, friendly versus aloof, egocentric, restricted teacher behavior.

TCS Pattern Y - Responsible, businesslike, systematic versus evading, unplanned, slipshod teacher behavior.

TCS Pattern Z - Stimulating, imaginative, surgent versus dull, routine, unimaginative teacher behavior.

The relationship between certain personality characteristics and teaching effectiveness has been the subject of numerous research studies. Getzels and Jackson report a number of studies dealing with authoritarianism. One of the most relevant was conducted by McGee who correlated Classroom Observation Records from the Teachers Characteristics Study with the California F Scale for measuring authoritarianism. He found (at the .005 level) a significant relationship between a measure of antidemocratic potential and a measure of teachers' overt authoritarian behavior in the classroom.

In a controlled experiment on responsive-directive dimensions of teacher behavior, Miller found that junior high school students under the direction of responsive teachers had significantly more positive attitudes and used significantly higher levels of thinking than did pupils in classes in which the opposite treatment was used.

In a study which used the classroom observation record to evaluate the teaching behavior of 115 secondary teachers, the author found that the teachers who were rated highest by independent observers in other areas of teaching effectiveness also rated higher on characteristics of fairness, democratic behavior, understanding, kindliness, stimulation, originality, alertness, attractiveness, responsibility, steadiness, poise, confidence, adaptability, and optimism.

After having examined the generalizations drawn from the research on teaching and learning, one may be struck with the realization that no generalization has been made about the relationship between the amount of teacher preparation and pupil gain. The relationship between the amount of academic training of the teacher and the gain in learning of the pupils still remains a moot question. Several researchers, however, have attempted to deal with this problem. Seymour Metzner reviewed 17 research projects which seemed to be based on some dimension of the premise that...
extended teacher preparation produces more knowledgeable teachers who are therefore better equipped to impart this knowledge to their students. He concluded, "There is not a single study that, after equating for pupil intelligence and socioeconomic status, has found the length of teacher preparation variable to be even peripherally related to pupil gain...."39

One may be equally impressed by the fact that there is almost complete unanimity among researchers that teacher behavior (as opposed to teacher knowledge) proves to be a significant research variable. This is particularly evident when the teacher behavior being investigated falls into some category of the affective realm.

In summary of the research on teaching and teachers, it should be pointed out again that the research reported herein has been little more than a representative sample. These studies do, however, in the opinion of the author, characterize the vast majority of recent studies and serve as valid bases for the derivation of the three major generalizations and the supportive sub-generalizations.

The research appeared to fall into three thematic clusters:

1. Research dealing with teaching methodology
2. Research dealing with the affective climate of the classroom
3. Research dealing with characteristics of teachers

The three generalizations and the supportive sub-generalizations were drawn from these thematic clusters.

The generalizations may be oversimplified and overgeneralized. However, even were they not documented with reputable research studies, many educators can accept them on the strength of empirical validity alone. In the absence of evidence which disproves the generalizations, they form the major bases for the recommended evaluation program presented on the following pages.
III. A Proposed Model for Evaluating Teacher Education Graduates

Any model for evaluating the product of teacher education will be inadequate and incomplete. The problems are too great and the knowledge about evaluation too limited to allow the presentation of a model which is not subject to criticism. The proposed model presented in this paper may be subject to an infinite number of valid criticisms. However, it should be explicitly understood that the plan has been offered only as a suggestion to those institutions facing the responsibility of evaluating the teachers they have prepared. Moreover, it should be recognized that the recommended model, despite its inadequacies, has been based on logically inferred generalizations drawn from relevant research.

Obviously, the evaluation of teachers must be based on many factors. The sum total of the program of preparation including both content and methodology, the personal characteristics of the teacher, the environmental determinants (including physical facilities, instructional resources, and administrative support) are illustrative of the scope of evaluative factors which merit consideration.

There is a growing conviction that evaluation of teacher effectiveness, in the final analysis, must rest upon the criterion of "pupil-gain," that is, evidence that the learner has achieved in some measure the intended objectives of the teacher. As desirable as pupil-gain measures may be, they are extremely elusive. In discussing pupil-gain measures, John Herbert has written:

While we should do more and better research on which teacher behaviors result in changes in pupil behavior, it is not expedient to evaluate teacher preparation programs by such changes in the schools where the teachers find employment....Combinations of variables--the school and home environment of the pupils and the decisions of the teacher's peers and administrators--may result in placing him in a position where, regardless of training received or the criteria used, he either cannot fail or cannot succeed. It would thus be no more reasonable to evaluate a teacher preparation program by the way pupils learn in the classroom of graduates than to evaluate a program of medical training by the health of the population its graduates serve. Therefore, though it is theoretically attractive to relate pupil behavior to accreditation, this seems unlikely to be feasible in the foreseeable future. As Ryans found: "With all the attractiveness of judgment of teacher behavior from its products (e.g., pupil changes)... the disadvantages of such approaches seem to outweigh their advantages."40

The proposed plan, recognizing the extreme complexity of using pupil-gain as an evaluative criterion, limited the model to measures of
teacher behavior based, insofar as possible, on research inferences. The proposed evaluative data can be derived from four categories:

A. Career line data
B. Direct classroom observation
C. Pupil, peer, and supervisory evaluations
D. Standardized measures

Career Line Data

It is readily apparent that teacher education institutions should collect information on the career lines of their graduates. According to Herbert, career line data would include information on such matters as wastage from teaching, types of teaching and administrative positions held, participation in research and program development, further training and education undertaken, and teacher mobility.41

The logic behind the collection of career line data is difficult to refute. Were an institution to find that the teachers it prepared averaged no more than 3 to 5 years in the classroom, that institution would obviously be faced with a problem of excessive wastage and should examine its preparation program in the light of that finding. Equally disturbing to an institution would be the discovery that, collectively, its product did not receive professional advancement and promotions, nor did they actively seek advanced degrees. Career line information must be collected on a longitudinal basis continuing throughout the professional career of the individual or as long as contact can be maintained.

Direct Classroom Observation

One of the newest and most effective techniques for determining teacher behaviors can be found in classroom observation systems. Classroom observation systems can be defined as "An organized and systematic attempt to assess and quantify through observation the behaviors of teachers and students engaged in the teaching-learning process."42 Among the best-known and most-used observational techniques are the interaction analysis systems which experienced ascendancy within the past decade follow the development of the Flander's System of Interaction Analysis the pertinent research conducted and reported by Flanders and others.43

Medley and Mitzel have stated that the true role of direct observation in research on teacher effectiveness must be one in which there is some attempt made to comprehend the nature of effective teaching.44 While it is true that most systems for direct classroom observation were developed primarily for research purposes, many are suited for aiding in the training
of classroom teachers and for the evaluation of inservice teachers. Brief descriptions have been presented in the following paragraphs of four classroom observation systems which appear to be effective in the evaluation of inservice teachers and would effectively measure verbal classroom interaction.

The Flanders System of Interaction Analysis. Flanders' intent in developing his system of interaction analysis was to record a series of acts of predetermined concepts with reference to the teacher's control of the student's freedom of action. Specifically, he attempted to distinguish those acts of the teacher which increased the freedom of the student from those which decreased the student's freedom of action, and to keep a record of both. This system, by far the best known, is also the simplest. It can be learned in 12 to 20 hours and has been shown by researchers to have both validity and reliability.

The Flanders system has only 10 categories: 7 for teacher talk, 2 for student talk, and 1 for silence or confusion. The categories are:

**Indirect Influence Categories** (expand student freedom)
1. Accepts pupil's feelings
2. Praises or encourages pupil
3. Accepts or uses pupil's ideas
4. Asks questions

**Direct Influence Categories** (restrict student freedom)
5. Lectures
6. Gives directions
7. Criticizes or justifies authority

**Student Talk**
8. Student talk-Response
9. Student talk-Initiation
10. Silence or Confusion

Use of the indirect influence categories encourages the student to participate in classroom discussion and gives him more opportunity to commit himself. Direct influence tends to inhibit student initiative and promote compliance.
Since the Flanders system is coded by the numbers of the 10 categories, it can be used effectively by training observers to record the number of the category which is occurring in the classroom at that given moment. To assure that all categories of verbal interaction are recorded, the observer writes the number of the category occurring every 3 seconds or every time the category changes. By writing the category numbers vertically in a column, the observer records 20 to 24 observations per minute and thereby acquires an objective record of the verbal interaction occurring in a classroom. This information can be recorded in a 10 x 10 matrix for statistical treatment.

Verbal Interaction Category System (VICS). The Verbal Interaction Category System developed by Edmund Amidon and Elizabeth Hunter is closely related to the Flanders system. In fact it simply expands the Flanders system to provide more detailed information. VICS contains 5 major categories for analyzing classroom verbal behavior: teacher-initiated talk, teacher response, pupil response, pupil-initiated talk, and other. Like the Flanders system, all of the categories must be memorized and are recorded in 3-second intervals by category number. The categories are:

Teacher-Initiated Talk
1. Gives information or opinion
2. Gives direction
3. Asks narrow question
4. Asks broad question

Teacher’s Response
5. Accepts
6. Rejects

Pupil Response
7. Responds to teacher
8. Responds to another pupil

Pupil-Initiated Talk
9. Initiates talk to teacher
10. Initiates talk to another pupil
Other

11. Silence

12. Confusion

As in the Flanders system, a matrix is used to plot the amount, sequence, and pattern of verbal behavior in the classroom. It can be determined from the matrix what kinds of behavior followed or preceded specific behaviors. Recurring patterns of behavior can also be seen.

The Hough System. A 16-category observational system was developed by John B. Hough. The system, another expansion of the Flanders system, is somewhat different in emphasis from the VICS. The 16 categories have been listed as follows:

Teacher Talk

1. Accepts feeling
2. Praises or encourages
3. Accepts or uses ideas of student
4. Asks questions
5. Answers student questions
6. Lectures
7. Corrective feedback
8. Gives directions
9. Criticizes or justifies authority

Student Talk

10. Student talk-response
11. Student talk-emitted
12. Student questions

Silence

13. Directed practice or activity
14. Silence and contemplation
15. Demonstration

Non-Functional

16. Confusion and irrelevant behavior

Three examples of systems of interaction have been presented and each has advantages and disadvantages. The Flanders system, for example, is simple and easy to code but does not provide the degree of specificity that does the VICS or the Hough system. Hough’s system provides specificity but is more difficult to code by virtue of the fact that the observer is working with 16 rather than 10 categories. For purpose of recommending a model for use in evaluation of teacher education graduates it seems appropriate to suggest a modification that combines some of the features of both Flander's 10-category system and Hough's 16-category system. Accordingly, a 14-category system is proposed:

Teacher Talk

1. Accepts feeling
2. Praises or encourages
3. Accepts or uses ideas of student
4. Asks questions
5. Answers student questions
6. Lectures
7. Corrective feedback
8. Gives directions
9. Criticizes or justifies authority

Student Talk

10. Student talk
11. Student questions

Silence or Non-Functional

12. Directed practice or activity
13. Demonstration
14. Silence or confusion
The 14-category system is essentially the same as Hough's system except that Hough's categories 10 and 11, student talk-response, and student talk-emitted, have been combined into one category of student talk. This recommendation has been made for two reasons: valid coding is considerably easier to obtain with fewer categories, and observers have difficulty distinguishing between student talk in response and student talk emitted. The other combination of categories has occurred with Hough's categories 14 and 16. Since both are non-functional behaviors, it seems unnecessary to retain both categories; therefore, they have been replaced with Flanders' category 10.

The second type of direct classroom observation which has had sufficient use to justify its recommendation in the proposed evaluation model is the Classroom Observation Record developed by David Ryans in the Teachers Characteristics Study sponsored by the American Council on Education. The Classroom Observation Record has attempted to assess 4 dimensions of pupil behavior and 18 dimensions of teacher behavior on a 7-point bi-polar scale. Each dimension of pupil and teacher behavior is carefully described and defined in a glossary which accompanies the Classroom Observation Record. A copy of the Classroom Observation Record (see page 18) and a sample page of the Glossary (see page 19) are included.

In preparing for the observation and assessment of teacher behavior through the use of the Classroom Observation Record, particular attention should be given to the selection and training of the observers. Only experienced teachers should be selected on the basis of 1) their ability to attend and perceive, 2) their familiarity with teacher behavior and its analysis and assessment, 3) their ability to set aside personal biases and employ an objective approach to the dimensions of teacher behavior, 4) their possession of social skill, 5) their general ability, and 6) their emotional adjustment.

The training procedure should include lengthy study of the Classroom Record and Glossary (COR), extensive practice with the COR and comparison of the results with trained observers, and finally, observation by both trained and untrained observers so that assessments can be compared and discussed.

Pupil, Peer, and Administrator/Supervisor Evaluation

It is evident that one means of gathering data on the behavior of teachers is to ask those who are in a position to know. Consequently, rating scales have been widely used in the research on teaching. Rating scales have the advantage of allowing the researcher to use a human observer to describe characteristics of another person. H. H. Remmers has written concerning rating scales:
Classroom Observation Record

Teacher Characteristics Study

<table>
<thead>
<tr>
<th>Teacher</th>
<th>No.</th>
<th>Sex</th>
<th>Subject</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>City</td>
<td></td>
<td></td>
<td>School</td>
<td>Time</td>
</tr>
</tbody>
</table>

PUPIL BEHAVIOR

1. Apathetic 1 2 3 4 5 6 7 N Alert
2. Obstructive 1 2 3 4 5 6 7 N Responsible
3. Uncertain 1 2 3 4 5 6 7 N Confident
4. Dependent 1 2 3 4 5 6 7 N Initiating

TEACHER BEHAVIOR

5. Partial 1 2 3 4 5 6 7 N Fair
6. Autocratic 1 2 3 4 5 6 7 N Democratic
7. Aloof 1 2 3 4 5 6 7 N Responsive
8. Restricted 1 2 3 4 5 6 7 N Understanding
9. Harsh 1 2 3 4 5 6 7 N Kindly
10. Dull 1 2 3 4 5 6 7 N Stimulating
11. Stereotyped 1 2 3 4 5 6 7 N Original
12. Apathetic 1 2 3 4 5 6 7 N Alert
13. Unimpressive 1 2 3 4 5 6 7 N Attractive
14. Evading 1 2 3 4 5 6 7 N Responsible
15. Erratic 1 2 3 4 5 6 7 N Steady
16. Excitable 1 2 3 4 5 6 7 N Poised
17. Uncertain 1 2 3 4 5 6 7 N Confident
18. Disorganized 1 2 3 4 5 6 7 N Systematic
19. Inflexible 1 2 3 4 5 6 7 N Adaptable
20. Pessimistic 1 2 3 4 5 6 7 N Optimistic
21. Immati 1 2 3 4 5 6 7 N Integrated
22. Narrow 4 2 3 4 5 6 7 N Broad
Teacher Behaviors

Partial-Fair Teacher Behavior

**Partial**
1. Repeatedly slighted a pupil.
2. Corrected or criticized certain pupils repeatedly.
3. Repeatedly gave a pupil special advantages.
4. Gave most attention to one or a few pupils.
5. Showed prejudice (favorable or unfavorable) towards some social, racial, or religious groups.

**Fair**
1. Treated all pupils approximately equally.
2. In case of controversy pupil allowed to explain his side.
3. Distributed attention to many pupils.
4. Rotated leadership impartially.
5. Based criticism or praise on factual evidence, not hearsay.

Autocratic-Democratic Teacher Behavior

**Autocratic**
1. Tells pupils each step to take.
2. Intolerant of pupils' ideas.
3. Mandatory in giving directions; orders to be obeyed at once.
4. Interrupted pupils although their discussion was relevant.
5. Always directed rather than participated.

**Democratic**
1. Guided pupils without being mandatory.
2. Exchanged ideas with pupils.
3. Encouraged (asked for) pupil opinion.
4. Encouraged pupils to make own decisions.
5. Entered into activities without domination.

Aloof-Responsive Teacher Behavior

**Aloof**
1. Stiff and formal in relations with pupils.
2. Apart; removed from class activity.
3. Condescending to pupils.
4. Routine and subject matter only concern; pupils as persons ignored.
5. Referred to pupil as "this child" or "that child."

**Responsive**
1. Approachable to all pupils.
2. Participates in class activity.
3. Responded to reasonable requests and/or questions.
4. Speaks to pupils as equals.
5. Commends effort.
7. Recognized individual differences.

Restricted-Understanding Teacher Behavior

**Restricted**
1. Recognized only academic accomplishments of pupils; no concern for personal problems.
2. Completely unsympathetic with a pupil's failure at a task.
3. Called attention only to very good or very poor work.
4. Was impatient with a pupil.

**Understanding**
1. Showed awareness of a pupil's personal emotional problems and needs.
2. Was tolerant of error on part of pupil.
3. Patient with a pupil beyond ordinary limits of patience.
4. Showed what appeared to be sincere sympathy with a pupils' viewpoint.
Note that the measuring device is not the paper form but rather the individual rater. Hence a rating scale differs in important respects from other paper-and-pencil devices. In addition to any limitations imposed by the form itself, ratings are limited by the characteristics of the human rater—his inevitably selective perception, memory, and forgetting, his lack of sensitivity to what may be psychologically and socially important, his inaccuracies of observation and, in the case of self-ratings, the well-established tendency to put his best foot forward, to perceive himself in a more favorable perspective than others do.51

There is considerable evidence in the literature that pupil evaluation of the teacher is a reasonably valid source of information. One study reports that if 25 or more student ratings of teachers are averaged, they are as reliable as the better educational and mental tests available as a predictor of teaching effectiveness.52 Remmers reported a large body of research on pupil ratings of teachers and drew the following generalizations from research conducted since 1927:53

1. Grades of students have little relationship to their ratings of the instructors who assigned the grades.

2. After 10 years, alumni ratings correlate highly (.92) with on-campus students.

3. Evidence indicates that students discriminate reliably among different aspects of the teacher's personality and the course.

4. Little if any relationship exists between the student's ratings of the teacher and the difficulty of the course.

5. The sex of the student bears little relationship to the rating.

6. Popularity of the teacher in extra-class activities is not appreciably related to student ratings of that teacher.

7. Teachers with less than 5 years' experience tend to be rated lower than teachers with more than 8 years' experience.

8. The sex of the teacher is generally unrelated to the pupil ratings.

The validity of using pupil ratings to assess teacher behaviors is well established in the research literature. Veldman and Peck conducted research with a rating system called the Pupil Observation Survey Report (POSR). On the 38-item POSR they found through a factor analysis 5 major dimensions of "space" within which pupils implicitly located their teachers. The 5 factors were: 1) friendly and cheerful, 2) knowledgeable and poised,
3) lively and interesting, 4) firm control (discipline), and 5) non-directive (democratic procedure). 54

There is a remarkable similarity between the first 3 of these factors and Ryan's 3 patterns which emerged from the Teacher Characteristic Study:

Pattern X - warm, kindly, understanding, friendly versus aloof, egocentric, restricted teacher behavior.

Pattern Y - responsible, businesslike, systematic versus evading, unplanned, slipshod teacher behavior.

Pattern Z - stimulating, imaginative, surging versus dull, routine, unimaginative teacher behavior. 55

In later research reported by Veldman and Peck, an effort was made to determine the extent to which pupil evaluations were related to supervisor evaluations of 609 student teachers. They found a definite relationship between factors 1, 2, and 3 which lends credence not only to Ryan's 3 patterns but also to the thesis that supervisors as well as pupils can consistently identify these important teaching behaviors. 56

As a result of their work with the POSR and the identification of the 5 factors within which pupils located their teachers, Veldman and Peck developed the Student Evaluation of Teaching (SET). The SET has several advantages to recommend its use as an instrument for the collection of data on student opinions of teaching behavior: the instrument is short, only 10 items; the SET has undergone extensive statistical analysis and shows correlation with other data-gathering devices; and the SET is easy to score--a FORTRAN program to score and print summary reports is available from the authors. The SET is copyrighted and permission to use or reproduce must be obtained from the authors. A copy of the SET follows on page 22.

While ratings from students seem a promising source of data on teaching behavior, there is evidence that agreement between supervisors and students, and even intra-supervisory ratings, has been difficult to obtain. Herbert, reviewing the research on supervisor/administrative ratings of teachers, concluded that the technique was subject to a number of limitations: 1) procedures and criteria for evaluating teachers vary considerably, 2) evidence on which ratings are based is very meager, 3) the personality of the principal seems to have a substantial effect on ratings of a teacher's ability and social competence, and 4) school district and college supervisors do not seem to agree on their ratings of teachers. 57

Despite the lack of research evidence that peer and supervisory evaluations have validity in assessing teaching behaviors, empirically a case can be presented for systematizing a peer-supervisor rating scale which can be quantified. The fact that little evidence is presently available which supports such ratings is no doubt largely due to the lack of a commonly acceptable scale which consistently seeks data on
STUDENT EVALUATION OF TEACHING

D. J. VELOMAN and R. F. PECK

TEACHER’S LAST NAME: ____________________________

SUBJECT: _______________________________________

SCHOOL: _______________________________________

CIRCLE THE RIGHT CHOICES BELOW

Teacher’s Sex: M F
My Sex: M F
My Grade Level: 3 4 5 6 7 8 9 10 11 12

CIRCLE ONE OF THE FOUR CHOICES IN FRONT OF EACH STATEMENT.
THE FOUR CHOICES MEAN:

F = Very Much False
f = More False Than True
t = More True Than False
T = Very Much True

This Teacher:

F f t T is always friendly toward students.
F f t T knows a lot about the subject.
F f t T is never dull or boring.
F f t T expects a lot from students.
F f t T asks for students’ opinions before making decisions.
F f t T is usually cheerful and optimistic.
F f t T is not confused by unexpected questions.
F f t T makes learning more like fun than work.
F f t T doesn’t let students get away with anything.
F f t T often gives students a choice in assignments.
teaching effectiveness. The peer/supervisor evaluation scale on pages 24-27 has been derived from faculty evaluation forms designed at Kansas State Teachers College. The original forms have been used at Kansas State Teachers College for 2 years and have been found to provide a range of discrimination. Therefore, in the absence of a validated form, the peer/supervisor form is offered as a means of collecting rating data on teaching behaviors.

Standardized Measures

As has already been pointed out, the use of standardized measures to determine student-gain is difficult in that there are multiple variables which influence pupil-gain in addition to effective teaching behaviors. The institution that chooses to use standardized measures to attempt to assess pupil-gain has an infinite number of instruments from which to choose. Intelligence tests (both short and long forms), achievement tests in subject areas, personality and attitude tests, all exist in abundance. They should be used with the understanding and awareness that there are few, if any, successful models on which to base their efforts. Basic questions, such as the effect of time as a variable on learning, the influence of extraneous variables such as home environment, previously-learned skills, personality, and countless other unknown factors make research in this area extremely difficult. The difficulty, however, should not prohibit or discourage further research in this area.

The use of standardized measures to assess certain personality characteristics which seem desirable in teachers has some precedents. Hundreds of research studies have dealt with personality and teaching effectiveness. Perhaps the greatest number of these have used 3 personality measures, the Minnesota Teacher Attitude Inventory (MTAI), the California F Scale, and the Minnesota Multiphasic Personality Inventory (MMPI). The failure of all this research to produce definitive results led Getzels and Jackson to write:

Despite the critical importance of the problem and a half-century of prodigious research effort, very little is known for certain about the nature and measurement of teacher personality, or about the relation between teacher personality and teaching effectiveness. The regrettable fact is that many of the studies so far have not produced significant results.

More research support exists for the use of the California F Scale possibly than for any other measure. Sheldon, Coale, and Copple, McGee, Hough and Amidon, and others have reported significant relationships between the degree of authoritarianism exhibited and certain teaching behaviors. Further research should be conducted in this area and the California F Scale seems to be a most promising instrument to use.
Teacher Evaluation
by
Peer/Supervisor

Name of Teacher Evaluated ____________________________

School ____________________________

Grade or Subject Taught ____________________________

Please check the appropriate items about yourself

☐ Female ☐ Male

☐ Peer (Colleague) ☐ Administrator or Supervisor

As a part of the continuing evaluation of selected faculty
of this School you are being asked to evaluate one of your fellow
faculty members. Please answer the following items as candidly
and consisely as possible. You may use the back of this page if
additional space is needed.

1. What are your particular qualifications for evaluating this
person?

2. Assuming this person is eligible, would you recommend promotion?
   Yes ☐ No ☐ Comment:

3. Assuming this person is eligible for tenure, would you recommend
tenure?
   Yes ☐ No ☐ Comment:

4. Assuming this person is eligible for reappointment as a pro-
bationary faculty member, would you recommend reappointment?
   Yes ☐ No ☐ Comment:
Teaching is the most important task of the school. In order to help the school to be informed regarding the quality of its teaching, you are requested to indicate your opinion of this instructor's performance in the four important dimensions of teaching described on the following pages. The highest rating is number 5; the lowest is number 1. Please encircle the number that represents your opinion of the instructor. Three of the five ratings for each dimension are described by words and phrases printed to the left of the numbers. The intermediate numbers may also be used for the expression of your opinions.

<table>
<thead>
<tr>
<th>DIMENSIONS OF TEACHING</th>
<th>DESCRIPTIVE WORDS AND PHRASES</th>
<th>RATING</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject Matter</td>
<td>Thorough, broad, and accurate knowledge of theory and practice; very able to organize, interpret, explain and illustrate concepts and relationships.</td>
<td>5</td>
</tr>
<tr>
<td>Competence</td>
<td>Adequate understanding; most interpretations and explanations are clear.</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Knowledge of subject is limited; does not give clear explanations and illustrations.</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Relations with Students</td>
<td>Excellent rapport; feeling of good-will prevails; very interested in students; easily approached; students are challenged yet individuality is respected.</td>
<td>5</td>
</tr>
<tr>
<td>Students</td>
<td>Adequate rapport; shows some interest in students; usually approachable; students are encouraged to participate; shows some sense of humor.</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Seems unfriendly and unresponsive; impatient; sometimes antagonizes students; too busy to be helpful.</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>DIMENSIONS OF TEACHING</td>
<td>DESCRIPTIVE WORDS AND PHRASES</td>
<td>RATING</td>
</tr>
<tr>
<td>----------------------------------------</td>
<td>---------------------------------------------------------------------------------------------</td>
<td>--------</td>
</tr>
<tr>
<td>Appropriateness of Assignments and</td>
<td>Assignments are challenging; he allows for differences of ability but expects superior</td>
<td>5</td>
</tr>
<tr>
<td>Academic Expectations</td>
<td>achievement; stresses important topics and concepts and avoids giving time to trivial</td>
<td></td>
</tr>
<tr>
<td></td>
<td>details; demands critical and analytical thought; tests seem valid.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Most assignments are clear, reasonable and related to class work; expects understanding not</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>memorization; recognizes individual differences among students but generally seems to ignore</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>them; tests are usually related to assignments and class work.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Assignments are unrealistic, often not clear, not related to class work; students do not</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>know what the teacher expects; tests seem unrelated to assignments and class work.</td>
<td>1</td>
</tr>
<tr>
<td>Overall Classroom Effectiveness</td>
<td>Lessons are carefully planned and show definite purpose; words come easily; well-organized</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>ideas and concepts are clearly related; enthusiastic and stimulating; raises thought</td>
<td></td>
</tr>
<tr>
<td></td>
<td>provoking questions; discussions are lively; pleasing manner, free from annoying mannerisms.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Usually well prepared, purposes are usually clear; presentations are fairly well-organized;</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>encourages student participation; objectionable mannerisms are not serious or numerous; asks</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>some good questions.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>DIMENSIONS OF TEACHING</td>
<td>DESCRIPTIVE WORDS AND PHRASES</td>
<td>RATING</td>
</tr>
<tr>
<td>------------------------</td>
<td>--------------------------------</td>
<td>--------</td>
</tr>
<tr>
<td>Lessons not planned, purposes are lacking or vague; relationships of concepts are not explained; asks few questions; subject seems uninteresting to him; repeatedly exhibits annoying mannerisms.</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

You may wish to comment further on this instructor's teaching performance. If so, you may use the space below and the back of this page.
The effort to design a model for the evaluation of teacher education graduates was based on 2 major premises: that a sufficient body of research was now in existence from which generalizations on good teaching and good teachers could be drawn, and that classroom observational systems and other evaluative tools had been developed which enabled educators to evaluate systematically the product of teacher education programs in the light of the research findings. The overriding premise was, of course, the position that institutions of teacher education had historically ignored the whole area of evaluation but were now required to face the issue because of the new Standards implemented by the National Council for Accreditation of Teacher Education.

Although no claims were made as to the completeness of the review of research on teaching and on teachers, an extensive review enabled the author to identify 3 thematic clusters into which the great majority of research projects could be classified. The 3 clusters were: 1) research dealing with teaching methodologies, 2) research dealing with the affective climate of the classroom, and 3) research dealing with characteristics of teachers. From these research clusters, 3 major generalizations, each with a series of sub-generalizations were drawn and presented as a synthesis of research findings relating to teaching and characteristics of good teachers. The generalizations were:

1. Good teaching utilizes maximal involvement of the student in direct experiential situations.
   - Good teachers attempt to foster problem-oriented, self-directed, actively inquiring patterns of learning behavior in their students.
   - Good teachers elicit pupil-initiated talk and allow more pupil-initiated exploration and trial solutions.
   - When teachers try to elicit independent thinking from their students, they get it.
   - Good teachers involve students in decision-making processes in active, self-directing ways.
   - Teachers who are interested in student involvement are less prone to dominate the classroom through lecture and other teacher activities.

2. Good teaching encourages maximal "freedom" for the student.
   - Good teachers use significantly more praise and encouragement for the student.
They accept, use, and clarify student ideas more often.

They give fewer directions, less criticism, less justification of the teacher's authority, and less negative feedback.

They use a relaxed conversational teaching style.

They use more divergent questions, do more probing, and are less procedural.

They are more inclined to recognize the "affective climate" of the classroom and are responsive to student feelings.

Teachers with low dogmatism scores are more likely to use indirect methods than those with more closed-minded attitudes.

3. Good teachers tend to exhibit identifiable personal traits broadly characterized by warmth, a democratic attitude, affective awareness, and a personal concern for students.

- Good teachers exhibit characteristics of fairness and democratic behavior.
- They are responsive, understanding and kindly.
- They are stimulating and original in their teaching.
- They are responsible and systematic.
- They are poised and confident, and emotionally self-controlled.
- They are adaptable and optimistic.
- They are well-versed in subject matter and give evidence of a broad cultural background.

Although another researcher reviewing the same research undoubtedly would have worded the generalizations differently and probably would have added some and deleted others, these generalizations represent the author's synthesis of the research findings and have served as the basis for the evaluation model.

The model recommended for the evaluation of teacher education graduates was based on data to be obtained from four sources: career line data; direct classroom observations; student, peer, and supervisor ratings; and standardized measures. From career line data it was suggested that institutions evaluating their graduates should systematically collect data on wastage from teaching, promotions, advanced degrees earned, writing, research, project activity, and teacher mobility.
From direct classroom observation systems, it was recommended that 2 systems be used: a 14-category modification of the Flanders system and the Hough variation of interaction analysis, and the Classroom Observation Record which was a product of the Teachers Characteristics Study. Both instruments produce valid quantitative data on the teaching behaviors used in the classroom and provide a high degree of objectivity to the assessment.

From the third source—student, peer, and supervisor ratings—the Student Evaluation of Teaching (SET) developed by Veldaan and Peck was suggested. The SET is a short rating device with only 10 items based on 5 factors determined by extensive research to be space within which students implicitly located their teachers. In addition, a teacher evaluation form to be used by peers and supervisors was developed by combining 2 rating forms developed and used at Kansas State Teachers College. The original forms were shown to provide a wide range of discrimination although no effort was made to validate them.

The fourth area, that of standardized measures, was treated vaguely. Since pupil-gain measures are difficult but desirable, it was suggested that institutions wishing to use some form of pupil-gain measure had an infinite variety of intelligence, achievement, personality, and attitude tests which could be employed. It was suggested, however, that the variable of effective teaching as a criterion of pupil-gain had proven elusive to even the most sophisticated researchers. In the area of teacher personality, however, it was pointed out that the California F Scale had correlated significantly with several dimensions of teaching behavior in several apparently valid research efforts and that further research seemed merited. In addition to the California F Scale, the Minnesota Teacher Attitude Inventory has been used extensively and has provided variable results.

Recommendations

An institution attempting to implement a comprehensive evaluation system for its teacher education graduates based upon this model would obviously have a number of decisions to make. Because the expense of a comprehensive evaluation program is considerable, the number of graduates who will be selected for evaluation becomes an important question. How often evaluation procedures will be replicated is an equally important question in terms of the commitment of financial and personnel resources.

In view of the many decisions that must be made and with a full awareness that individual institutions must adapt their evaluation program to their specific needs and conduct the program in keeping with their resources, the following recommendations have been made for a theoretical model.
1. The evaluation should be longitudinal and continue at least through a 5-year cycle.

2. The initial data should be collected during the preservice student teaching phase and should include:
   a. Personal data: name, age, sex, permanent address, major, minor, cumulative grade point average, and standardized test scores.
   b. Direct classroom observations:
      (1) Three 20-minute administrations of the interactional analysis system, each conducted on a different day. This will provide a 60-minute base for analysis of verbal classroom interaction.
      (2) Three different administrations of the Classroom Observation Record. (To be administered by the same observer-recorder who conducts the interaction analysis).
   c. Ratings
      (1) Toward the end of the student teacher period, every pupil of the student teacher should complete an unsigned SET to be administered by the observer-recorder.
      (2) The Peer/Supervisor Teacher Evaluation form should be completed by the observer-recorder, the college supervising teacher, and the public school supervising teacher.
   d. The California F Scale and other selected instruments should be administered during the latter part of student teaching.
   e. The evaluation program should be replicated at the end of the first, third, and fifth years.
   f. A minimum of 40 students should be randomly selected from a stratified sample annually. Consequently, 40 students would be evaluated initially, 80 students the first and second years, 120 the third and fourth, and the maximum number of 160 would be reached on the fifth year and continue thereafter. Attrition would, of course, significantly reduce all numbers after the initial year. Institutions replicating data at the tenth year would assess the remaining teachers and never exceed maximum 200 teachers per year.
   g. All data should be recorded for computerized statistical treatment and storage. A composite teaching profile should be
completed and institutional norms established.

h. Observers administering the system of interaction analysis and the Classroom Observation Record must be carefully trained. An inter-observer correlation in excess of .75 is generally acceptable.

i. The evaluation program will be significantly improved if the generalizations on good teaching and good teachers are restated as behavioral objectives of the teacher education program against which all data will be assessed.

j. There is considerable latitude which can be applied to the statistical treatment. The establishment of a bank of evaluative data is significant. It would appear that correlations should be computed between discrete variables. Changes in behavior indicated by testing intervals should be tested for significance and analyzed for causal factors. It is implicit that all data be examined in terms of the stated behavioral objectives and the total teacher education program revised in the light of findings.
IV. Projected Costs of the Proposed Model

Because of the large number of options available to institutions which would implement all or portions of the proposed model, cost estimates have been difficult to assess. There can be no doubt, however, that the recommended program of evaluation will require considerable added resources in terms of faculty time as well as minimal costs for the instruments used. The following estimates are undoubtedly inexact and have been offered only as a guide to interested institutions.

The authors of the Student Evaluation of Teaching and the Classroom Observation Record have not distributed their instruments commercially but with prior permission have permitted reproduction of the instruments without charge. The Peer/Supervision Rating Scale and the 14-Category System of Interaction Analysis have not been copyrighted and apparently may be reproduced for institutional use. Therefore, the 4 major evaluative instruments may be had for the cost of reproduction only, a negligible sum.

The standardized measures, intelligence tests, achievement tests, personality rating scales, and etc., may add significant expense if the institution opts to attempt measures of pupil gain and administers commercially-prepared examinations to large number of public school students. If standardized instruments are limited to preservice and inservice teachers, the maximum number will never exceed 200 and is likely to be 20-30 percent less. Administrations of most standardized measures will require inconsequential cost.

The major expense of the proposed model will be incurred in faculty and administrative time. The best calculation that may be made initially has been based on the experience of several researchers who have administered similar evaluation programs on an experimental basis. It seems reasonable to assume that the planning and administration of an evaluation program of the complexity and magnitude of the recommended model will require a minimum of one-half of a professional position. Each classroom visit must be calculated at three hours minimum time including travel, conference, and administration of the evaluative devices. Thus, 40 students x 3 visits x 3 hours equals 360 hours required for the initial administration. This requirement would increase to 720 hours for the second and third year, and to 1090 hours by the fourth year. The hours required should stabilize at somewhere between 1300-1500 hours of professional time. In short, it appears that one full-time equivalent position would be minimally required.

It is difficult to estimate the cost of computer time and the cost of data storage. Probably no more than one hour of computer time would be required for a typical analysis of evaluation data.

A broad estimate of minimal costs to an institution may appear as follows:
Travel, estimated at 50 miles per visit at 9¢ per mile - 50x.09x360 $1,620.00

Cost of reproductions of evaluation materials 200.00

Cost of one-to-two hours of computer time 375.00

One-half administrative position @$12,000.00 annual salary 6,000.00

One full-time equivalent position 12,000.00

Miscellaneous costs 500.00

$20,695.00

Research has shown that graduate students who are experienced teachers make excellent observers. The use of graduate students as trained observers could significantly reduce the personnel costs of the proposed model.
Notes


2Ibid.

3A report to the American Association of Colleges for Teacher Education's Committee on Standards prepared by J. T. Sandefur, October, 1971, p.16.


5Ibid. pp. 7-8.


11Peck and Tucker, "Research on Teacher Education" pp. 45-47.


22 Sandefur, "Professional Education for Secondary Teachers."


24 Finske, "The Effect of Feedback."


33 Barr, "Teaching Competencies" pp. 1449-51.

34 Ryans, Characteristics of Teachers, pp. 86-92.

35 Ibid. p. 382


41Ibid. p. 6.


43Flanders, Ned A., "Teacher Influence, Pupil Attitude..."


46Simon and Boyer, Mirrors for Behavior, Section Two, Amidon-Hunter p 3.


48Ryans, David G., Characteristics of Teachers.

49The Sample Classroom Observation Record and the page from the Glossary were taken from the "NEA Research Memo" published by the National Education Association, 1201 Sixteenth Street, N. W., Washington, D. C., September, 1962.

50Ibid. p. 3.


52Ibid. p. 367.

53Ibid. pp. 367-68.

54Veldman, Donald J. "Student Evaluation of Teaching," Research Methodology Monograph No. 10, R & D Center of Teacher Education, University of Texas at Austin, 1970.


56Veldman, "Student Evaluation of Teaching" p. 2.

57Herbert, "A Research Base...." pp. 5-6.
The Peer/Supervisor Teacher Evaluation Form was constructed from a Student Evaluation Form and a Colleague Evaluation Form designed by a Committee of the School of Education and Psychology at Kansas State Teachers College. Information on the total evaluation system can be obtained by writing Dr. Fred A. Markowitz, Associate Dean, School of Education and Psychology, Kansas State Teachers College, Emporia, Kansas.


Hough and Amidon, "The Relationship of Personality Structure."
Bibliography


Veldman, D. J., "Student Evaluation of Teaching," Research Methodology Monograph No. 10, Research and Development Center, University of Texas at Austin, 1970.
ABOUT AACTE

The American Association of Colleges for Teacher Education is an organization of more than 860 colleges and universities joined together in a common interest: more effective ways of preparing educational personnel for our changing society. It is national in scope, institutional in structure, and voluntary. It has served teacher education for 55 years in professional tasks which no single institution, agency, organization, or enterprise can accomplish alone.

AACTE's members are located in every state of the nation and in Puerto Rico, Guam, and the Virgin Islands. Collectively, they prepare more than 90 percent of the teaching force that enters American schools each year.

The Association maintains its headquarters in the National Center for Higher Education, in Washington, D. C.—the nation's capital, which also in recent years has become an educational capital. This location enables AACTE to work closely with many professional organizations and government agencies concerned with teachers and their preparation.

In AACTE headquarters, a stable professional staff is in continuous interaction with other educators and with officials who influence education, both in immediate actions and future thrusts. Educators have come to rely upon the AACTE headquarters office for information, ideas, and other assistance and, in turn, to share their aspirations and needs. Such interaction alerts the staff and officers to current and emerging needs of society and of education and makes AACTE the center for teacher education. The professional staff is regularly out in the field—nationally and internationally—serving educators and keeping abreast of the "real world." The headquarters office staff implements the Association's objectives and programs, keeping them vital and valid.

Through conferences, study committees, commissions, task forces, publications, and projects, AACTE conducts a program relevant to the current needs of those concerned with better preparation programs for educational personnel. Major programmatic thrusts are carried out by commissions on international education, multicultural education, and accreditation standards. Other activities include government relations and a consultative service in teacher education.

A number of activities are carried on collaboratively. These include major fiscal support for and selected of higher education representatives on the National Council for Accreditation of Teacher Education—an activity sanctioned by the National Commission on Accrediting and a joint enterprise of higher education institutions represented by AACTE, organizations of school board members, classroom teachers, state certification officers, and others.
The Association headquarters provides several secretariat services which help make teacher education more interdisciplinary and comprehensive; the Associated Organizations of Teacher Education, International Council on Education for Teaching, Society of Professors of Education, and The John Dewey Society. A major interest in teacher education provides a common bond between AACTE and fraternal organizations.

AACTE is deeply concerned with and involved in the major education issues of the day. Combining the considerable resources inherent in the consortium--constituted through a national voluntary association--with strengths of others creates a synergism of exceptional productivity and potentiality. Serving as the nerve center and spokesman for major efforts to improve education personnel, the Association brings to its task credibility, built-in cooperation and communications, contributions in cash and kind, and diverse staff and membership capabilities.

AACTE provides a capability for energetically, imaginatively, and effectively moving the nation forward through better prepared educational personnel. From its administration of the pioneering educational television program, "Continental Classroom," to its involvement of 20,000 practitioners, researchers, and decision makers in developing the current Recommended Standards for Teacher Education, to many other activities, AACTE has demonstrated its organizational and consortium qualification and experiences in conceptualizing, studying and experimenting, communicating, and implementing diverse thrusts for carrying out socially and educationally significant activities. With the past as prologue, AACTE is proud of its history and confident of its future among the "movers and doers" seeking continuous renewal of national aspirations and accomplishments through education.
ORDER FORM FOR RECENT AACTE PUBLICATIONS

Number of Copies

"The Profession, Politics, and Society" (1972 Yearbook)
- Volume I and Volume II @ $6.00
- Volume I (Proceedings) Only @ $4.00
- Volume II (Directory) Only @ $3.00

"Power and Decision Making in Teacher Education" (1971 Yearbook) @ $6.00

"What Kind of Environment Will Our Children Have?" @ $2.50

"Social Change and Teacher Education" @ $2.50

"Systems and Modeling: Self-Renewal Approaches to Teacher Education" @ $3.25

"Excellence in Teacher Education" (Limited Supply) @ $1.00

"Beyond the Upheaval" @ $1.00

"In West Virginia, It is Working" @ $2.00

"Educational Personnel for the Urban Schools: What Differentiated Staffing Can Do" @ $2.00

"An Illustrated Model for The Evaluation of Teacher Education Graduates" @ $2.00

BILLED ORDERS: Billed orders will be accepted only when made on official purchase orders of institutions, agencies, or organizations. Shipping and handling charges will be added to billed orders. Payment must accompany all other orders. There are no minimum orders.

DISCOUNTS: A 10 percent discount is allowed on purchases of five or more publications of any one title. Also, a 10 percent discount is allowed on all orders by wholesale agencies.

Payment enclosed _______ Bill me _______ Amount _______
Purchase Order No. _______

NAME __________________________ (Please type or print)

ADDRESS __________________________
ZIP __________________________

Please address: Order Department, American Association of Colleges for Teacher Education, Suite 610, One Dupont Circle, Washington, D. C. 20036

45
<table>
<thead>
<tr>
<th>Number of Copies</th>
<th>PBTE Series</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>#1</td>
</tr>
<tr>
<td></td>
<td>&quot;Performance-Based Teacher Education: What Is the State of the Art?&quot; by Stan Elam @ $2.00 per copy</td>
</tr>
<tr>
<td></td>
<td>#2</td>
</tr>
<tr>
<td></td>
<td>&quot;The Individualized, Competency-Based System of Teacher Education at Weber State College&quot; by Caseel Burke @ $2.00 per copy</td>
</tr>
<tr>
<td></td>
<td>#3</td>
</tr>
<tr>
<td></td>
<td>&quot;Manchester Interview: Competency-Based Teacher Education/Certification&quot; by Theodore Andrews @ $2.00 per copy</td>
</tr>
<tr>
<td></td>
<td>#4</td>
</tr>
<tr>
<td></td>
<td>&quot;A Critique of PBTE&quot; by Harry S. Broudy @ $2.00 per copy</td>
</tr>
<tr>
<td></td>
<td>#5</td>
</tr>
<tr>
<td></td>
<td>&quot;Competency-Based Teacher Education: A Scenario&quot; by James Cooper and Wilford Weber @ $2.00 per copy</td>
</tr>
<tr>
<td></td>
<td>#6</td>
</tr>
<tr>
<td></td>
<td>&quot;Changing Teacher Education in a Large Urban University&quot; by Frederic T. Giles and Clifford Foster @ $3.00 per copy</td>
</tr>
<tr>
<td></td>
<td>#7</td>
</tr>
<tr>
<td></td>
<td>&quot;Performance-Based Teacher Education: An Annotated Bibliography&quot; by AACTE and ERIC Clearinghouse on Teacher Education @ $3.00 per copy</td>
</tr>
<tr>
<td></td>
<td>#8</td>
</tr>
<tr>
<td></td>
<td>&quot;Performance-Based Teacher Education Programs: A Comparative Description&quot; by Iris Elfenbein @ $3.00 per copy</td>
</tr>
</tbody>
</table>

BILLED ORDERS: Billed orders will be accepted only when made on official purchase orders of institutions, agencies or organizations. Shipping and handling charges will be added to billed orders. Payment must accompany all other orders. There are no minimum orders.

DISCOUNTS: A 10 percent discount is allowed on purchases of five or more publications of any one title. Also, a 10 percent discount is allowed on all orders by wholesale agencies.

Payment enclosed ____ Bill me ____ Amount ______
Purchase Order No. ______

NAME __________________________________________
(Please print or type)

ADDRESS ________________________________________

ZIP __________________________

Please address: Order Department, American Association of Colleges for Teacher Education, Suite #610, One Dupont Circle, N. W., Washington, D. C. 20036