DOCUMENT RESUME

ED 049 268 TE 500 624

AUTHOR Larson, Richard L.
TITLE The Evaluation of Teaching College English.
INSTITUTION Modern Language Association of America, New York, N.Y. ERIC Clearinghouse on the Teaching of English in Higher Education.
SPONS AGENCY Office of Education (DHEW), Washington, D.C.
PUB DATE [71]
NOTE 101p.

EDRS PRICE EDRS Price MF-$0.65 HC-$6.58
DESCRIPTORS *Colleges, Educational Responsibility, English Curriculum, *English Departments, *English Instruction, English Programs, *Faculty Evaluation, *Humanities, Instructional Materials, Lesson Observation Criteria, Program Effectiveness, Program Evaluation, Program Improvement, Quality Control, Student Attitudes, Teacher Attitudes, Teacher Evaluation, Teacher Rating, Teacher Role, Teaching Styles

ABSTRACT

This monograph reviews theory and procedures of evaluation as reflected in the professional literature and in correspondence from chairmen of departments of English and the Humanities. Introductory comments are followed by chapters on: the evaluation of teaching; teaching and values; student evaluation of faculty and courses; observation of classes; inspection of teaching materials and annotated student papers; assessing the results of instruction; circumstantial evidence; teacher self-evaluation; conclusions and questions; and some recommendations. Appendixes contain: (1) sample forms for student rating of faculty; (2) a form for peers to use in making an evaluation; (3) a form for faculty self-evaluation; (4) a procedure for handling classroom observation; and (5) part of a department's chairman's evaluation form, emphasizing results of instruction. A selected bibliography is included. (RL)

THE EVALUATION OF TEACHING COLLEGE ENGLISH

Richard L. Larson
Department of English
University of Hawaii

ERIC

Clearinghouse on the
Teaching of English
in Higher Education

Modern Language Association of America
62 Fifth Avenue
New York, New York 10011

Published by the MLA/ERIC Clearinghouse on the Teaching of English in Higher Education, 62 Fifth Ave., New York 10011.

Copies of this pamphlet may be purchased from the Materials Center, MLA/ACTFL, 62 Fifth Ave., New York 10011.

CONTENTS

ACKNOWLEDGEMENTS

In the preparation of this study, I have had the help of many people, and I want to recognize that help here. Dr. Michael Shugrue, Secretary for English of the Modern Language Association, first suggested this study to me and offered helpful comments on my plans for it as well as on the drafts of the various chapters. Professor George Anderson, then Chairman of English at the University of Hawaii, supported the study and facilitated my correspondence with chairmen, deans, presidents, and officers of educational organizations. Professor David Ryans, Director of the University's Educational Research and Development Center, provided information about his research on characteristics of teachers. Miss Lorraine Nagata, Secretary of the Department of English at the University of Hawaii, handled much of the correspondence with administrators, and Mrs. Rosemarie Zimmer typed with her usual expertise successive drafts of a difficult manuscript. I appreciate greatly the help of each of them.

I also appreciate, and want to acknowledge here, the help of the hundreds of people to whom I wrote inquiries as a part of the study. Questionnaires about policy and procedure are among the trials of an administrator's life, and I apologize again here for adding one more burdensome request to the pile of papers that confronted my correspondents. Some of my correspondents answered my queries briefly, while others responded at considerable length, taking a great deal of time to supply informative comments on my questions. Everyone's reply, short or long, helped me, by disclosing attitudes toward and assumptions about evaluation even if not by supplying specific information. These replies are the heart of my study, since they enable me to be confident that my report deals directly with the procedures followed by departments of English and Humanities. Without these replies, then, this study would not exist. And I appreciate, also, the permission given me by several schools to reproduce forms and documents they use in evaluation of teaching. Such illustrations say better than any descriptions of mine how evaluation is practiced at schools whose faculties have thought a good deal about it.

At the close of my letters of inquiry, I asked for permission to quote from letters if the comments they contained struck me as especially pertinent to the questions I am discussing. Most respondents indicated their willingness to be quoted, but the majority of those who indicated such willingness also asked to remain anonymous. Out of respect for them, and in order to maintain a consistent policy, I have therefore not ascribed to their authors any of the comments I quote, unless the author of the statement has made the same remarks elsewhere in print, in which case I quote his remarks as published rather than as they appeared in his letter to me. I trust that this procedure will seem equitable even to those who agreed to be quoted by name, and I again thank those who supplied me with answers I could quote.

Honolulu, Hawaii
31 October 1970

Richard L. Larson

INTRODUCTION

This monograph was commissioned by the Modern Language Association as an early number in a series of reports on the "state of the art" in several instructional activities currently under review within the profession. Reliable procedures for evaluation of teaching (in all subjects) have been the goal of scholars in education for half a century, and a considerable body of literature on the subject has been developing for the last four or five years. Today, confronting students' complaints on many campuses that, however well-informed some members of the faculty may be, their teaching abilities are meager and their techniques outmoded, the profession finds it increasingly necessary to know how it makes judgments about the teaching that students are experiencing. A comparably urgent reason for concern with the evaluation of teaching at this time is the national pressure applied by legislators and administrative agencies, including some in the federal government, to hold teachers "accountable" to their employers and to citizens at large for the "effectiveness" or "success" of their teaching. For if a teacher is to be held "accountable" for the success of his teaching, how is he to demonstrate his success to those who demand the accounting, and how might they in turn demonstrate—if they are so moved—that the teacher's accomplishments fall short of justifying the privileges, opportunities, facilities, and money that have been given to him?

Two articles by teachers specifically of English that appeared after the beginning of the study help to emphasize further the need for it. Writing in the May 1970 issue of the *ADE Bulletin*, Basil Busacca of Occidental College comments on the difficulty members of his department experienced in being sure of the rightness of their selection for the MLA Certificate of Excellence in Teaching:

> Do we know, or know how to discover, who is a good teacher? Are our beautiful arguments for judging colleagues in terms of publication and other forms of public professional recognition really more than a persuasive (and comfortable!) way to pass the buck? Do we perhaps know who is a good teacher without knowing how to defend our certainty against demands for "hard" evidence? Is there any "hard" evidence which is actually authentic? Are there any real alternatives to what we do?[1]

In the same publication, John Gerber, Chairman of English at the University of Iowa, suggests another reason for attending to the evaluation of teaching: "We have not gotten enough [in our colleges and universities] from our investment in teaching.... We can recognize good teaching far more than we have in awarding raises and promotions."[2] "Recognition" in the sense of "honoring" surely requires the ability to "recognize" in the sense of "identifying with some certainty." It is our *uncertainty* when asked who is a good teacher, and our *need* for at least enough certainty so that we *can* reward good teaching, that forces us to inquire how teaching is evaluated now—and how it can perhaps be evaluated more accurately and dependably in the future.

I have sought solutions to these problems principally in two directions. First, I have reviewed much of the accumulated literature on the evaluation of teaching, particularly the important books and articles that have appeared within the last few years. And, since most writing about the evaluation of teaching has not been based on the teaching of any one subject, I have relied even more heavily upon mail inquiries: I sent out the inevitable questionnaire—this one to the chairmen of departments of English or Humanities in every state university in the country and in many state colleges, as well as in many four-year private universities and liberal-arts colleges and in a selection of both public and private two-year colleges. In response to suggestions from some chairmen, I wrote to deans and presidents at some of these colleges, to students—particularly student leaders of organizations that compile and distribute evaluations of courses on campus—and to the executive secretaries of regional associations of schools and colleges. Wherever a chairman indicated that he or a member of his department regularly visited the classes of other members of the staff, I followed up the initial questionnaire with another asking for specific information about the focus of the observers' attention during their visits to classes. I received replies from administrators and students at a goodly number of schools in each group, and in all parts of the country, although the total number of responses to my questions (which called for discursive answers, not simply multiple-choice responses or short factual statements) was less than half the number of inquiries sent out.

Not being a professional researcher in education, I was not equipped to follow formal statistical procedures in drawing up the list of schools to which I would send letters. Nor did I classify the responses by source, in order to demonstrate that the number of replies from a particular

kind of school constituted a statistically fair sampling of such schools. The goal of this study is not to establish what percentage of universities with graduate departments offering the Ph.D., or of two-year urban colleges, follow a given evaluative procedure; in this report, therefore, I cite relatively few numerical values. Those I do cite are given in round figures (like one in ten, one out of four, one out of two, and so on) and are not advanced with any claim for their statistical significance. My goal is to report what appear from my correspondence to be typical and useful procedures (procedures that particularly please some users) for determining whether good teaching is taking place, who is doing it, and how. I can describe the "state of the art" of evaluating teaching, indeed, only to the extent that practitioners of that art have been willing to share with me specifically the secrets of their art.

Some restrictions adopted to expedite the study need mention here. I define a "teacher" as one who is guiding students in a program of academic study. My remarks do not deal with the teacher's miscellaneous contributions to his department, to the university or college as a whole (for example, by service on committees), to the community as a whole, or to the scholarly profession (even though publication is regarded by many administrators as a highly significant form of teaching, and these administrators resisted the differentiating of research and teaching in my letters of inquiry). Nor do I include any reference to the teacher's function as counselor to student, as a helper in students' selection of a major, of courses, or of a career. And I have not tried to discuss training programs for teachers except insofar as such programs include evaluation of his work in training. These are of course arbitrary exclusions, but they seemed necessary to me in order to keep the subject of discussion clearly in focus: the teaching of English (whether "English" is thought of as an academic discipline or a process of growing). Besides looking for ways of identifying successful teachers, I also looked for effective procedures for discovering effective teaching *techniques*, valuable *curricula*, and satisfying *courses*, though I found that most evaluative procedures focus principally on the teacher rather than on how he teaches or features of the course he is teaching. It is hard, I find, to identify promising teaching *techniques* and worthwhile *courses* by the instruments presently employed in departments of English and Humanities for evaluating instruction.

When I began work on the project, I had in mind, among other goals, two that I have been unable to meet; they deserve brief mention here. First of all, I had hoped to be able to focus quite sharply on the teaching of English (and the Humanities), and to differentiate the art of evaluating teaching in English from the art of evaluating instruction in other subjects. That focus proved much more difficult to retain than I had expected. Some procedures, to be sure, are much more applicable—perhaps solely applicable—to English and the Humanities: for example, inspection of writing assignments and supporting materials in courses in composition, as well as the inspection of annotated student papers in such courses. But other evaluation procedures unique to English are had to discover, and perhaps impossible to devise. Comments by chairmen of English, deans, and even students mentioned almost no bases or procedures for evaluation intended solely for courses in English. Most of my chapters, therefore, discuss the art of evaluation in ways that might apply equally to any nonscientific subject, and even my illustrations of the *application* of evaluative techniques are not often uniquely drawn from the teaching of English—because evaluators of teaching in English often do not illustrate their procedures by showing how they handle events in the teaching of *English*. Indeed one pressing need in the teaching of English is for discussion of exactly what we think of as "good teaching" in our subject.

Second, I had been asked, and I had hoped, to explore the differences between the way evaluation of junior teachers is carried out by their senior colleagues, the way evaluation of older teachers is carried out by their peers, and the way evaluation of faculty is carried out by students. Comments on students' efforts at evaluation of faculty members, of course, are essential, and they appear in the long chapter on student ratings. But further differentiations quickly came to appear impossible, for a relatively simple reason. Most of the evaluation procedures described in this monograph are carried on explicitly with teaching assistants, instructors, and assistant professors not on tenure, but these techniques, *except* the use of student ratings, are much less often used with tenured or senior staff members. The only institutions in which systematic evaluation of teaching by older as well as younger members of the department is carried on with some regularity, it would seem, are the two-year colleges, where it is often salary rather than rank or tenure that separates veterans and newcomers. Within some two-year colleges, evaluation is a fact of life for all members of the staff, and differences in treatment of "upper" and "lower" "ranks" simply do not exist. But in many, perhaps most, four-year colleges and universities, once a teacher has earned tenure and promotion to one of the two upper ranks, and especially when he is teaching graduate courses, evaluation of his teaching (except by student

ratings, which are often voluntary and private) is nonexistent or inconsequential. This inattention to the teaching done by senior staff is changing in a few schools, which are beginning to give more attention to student ratings of all faculty, but the change is slow. Thus my chapters describe mainly the procedures employed for apprentice teachers and probationary teachers.

My initial inquiry attempted to identify the qualities of teaching insisted upon by chairmen or other administrators, the ways of gathering and interpreting evidence on whether each instructor's teaching was exhibiting these qualities, the action taken in response to data gathered, and improvements desired by the administrator or student in the techniques of gathering data. My questions were kept as broad as possible to enable each respondent to comment freely, either answering the questions directly, or rephrasing them so that he would find them more congenial, or giving reasons why he though them inapplicable. Any reply, even an objection to the question itself, was significant in disclosing attitudes of the respondent toward the possibility or process or problems of evaluation.

I learned a good deal more than I had at first envisaged about teachers' and administrators' attitudes toward evaluation—and not all of what I learned was encouraging. My questions clearly touched on matters that had puzzled many administrators; some shared with me both their procedures and their puzzlements, and others, though admitting the urgency of the questions, replied quite seriously that they had not time to answer in the kind of detail that the questions required, or that their department or their university had embarked on an examination of procedures for evaluating teaching and that they were not ready to respond when I wrote. Others responded with helpful completeness to my queries, addressing each of them squarely and honestly. Some chairmen, however, responded by describing the machinery by which a man is considered for promotion, tenure, or a salary increase, but not the grounds on which or the methods by which a judgment is made concerning his fitness. Several administrators wrote of the process followed in moving papers and dossiers from the department upward, as if that process were the act of evaluation itself. Other chairmen named the officials in the college or university responsible for carrying out the evaluation or for making a final decision, without recognizing that the crucial question is not who makes the judgment but rather the bases on which he makes it and his methods for ring the data with which to make it. It seems he *person* making the judgment (for example, tment chairman or dean) is, in the minds of some administrators, the same as the *basis* for judgment, and the machinery for ratifying an administrator's recommendation is more important than the *grounds* on which the recommendation is based.[3]

Some chairmen evidently refuse to have anything whatever to do with evaluation; one chairman from a prestigious private university told me that although he would like to help me with my inquiries, he simply had no way of knowing how to answer my questions. Although some administrators advanced serious arguments that teaching is an act inherently not susceptible of evaluation, many chairmen and deans expressed, in their hesitation about replying to my questions, the recognition that "evaluation" is very nearly a tabu notion on their campuses—one that immediately inspires revulsion among their colleagues and, perhaps as a result, timidity or even resentment among chairmen. The reaction was puzzling, because chairmen inevitably and incessantly engage in evaluation; that is a major part of their job. What these chairmen resis , it would seem, is the request that they bring to consciousness, for discussion and examination, the intuitive judgments on which they now rely. What they may also resist is the suggestion that the gathering of specific, observable data for evaluation be made a process visible for all faculty to see—a process that may cause those who see it to feel threatened, and to become self-conscious (to the detriment of their teaching), at the thought of its operation. Some teachers—and, possibly because of their feelings, some chairmen—may prefer to assume that if evaluation isn't discussed and isn't seen, it isn't going on. That assumption, of course, is a delusion.

Several chairmen were put off by the question about the standards they expected members of their faculty to meet in their teaching, arguing that to speak of standards in the context of teaching implies a possibility of codifying and systematizing what "teaching" includes—a possibility that does not exist. These chairmen hesitated to consider that anyone who looks at an object or an act and makes judgments upon it is applying some notion of what the act ought to be or ought to do; that is what evaluation is. A question about standards upheld in the evaluation of teaching invites self-examination; it invites the respondent to ask himself what elements, characteristics, or results he values when he looks at teaching, and, by implication, what elements, characteristics, and results leave him less satisfied. Granted the absolute need for flexibility in the assessment of teaching, and granted that successful teaching may take an infinite variety of forms, there is still a point at which one must say that in one particular or another the work of

Teacher A is as good as, better than, or less effective than that of Teacher B, and must explain his judgment—unless one is going to say that all teaching is equally good, and no chairman who wrote me appeared ready to take that position.

A final paradox: some chairmen were suspicious of the request for comment on how they weight the various kinds of evidence they receive concerning a person's teaching. Yet clearly some decision-makers are more impressed with some kinds of evidence than with other kinds—an assertion supported by the diversity of answers from chairmen to that very question about how they weight different sorts of evidence. Again I had hoped for some self-examination; I had hoped that chairmen would ask themselves: to what kinds of evidence do I give more credence? To what kinds less credence? Granted that the answers to these questions may vary from circumstance to circumstance, the fact of variation is itself important, as is the fact—if such it turns out to be—that in many or most cases one kind of evidence is preferred by a particular administrator over another kind of evidence. Anyone who wishes to work toward a more fundamental concern with the evaluation of teaching in American universities will need, it appears, to encourage administrators to examine their assumptions about teaching more deeply than some administrators who wrote me were able to do or had the time to do.

Hopefully this monograph, detailed and extended though it be, may encourage a few chairmen to attempt that reexamination.

[1]"A Postscript to 'The Department Awards a Certificate of Excellence,'" *ADE Bulletin*, No. 25 (May 1970), p. 30.

[2]"Response to Austerity," *ADE Bulletin*, No. 25 (May 1970), pp. 51-52.

[3]John W. Gustad, *Policies and Practices in Faculty Evaluation* (Washington, D.C.: American Council on Education, 1961); the matter is discussed in Ch. i.

CHAPTER I
THE EVALUATION OF TEACHING: SOME ISSUES

Not many issues divide teachers—particularly teachers of English—as sharply as do those of whether their teaching should be evaluated according to some sytematic plan and, if so, how such evaluation should be carried out. The suggestion, whatever its source, that teaching in the Humanities be the subject of deliberate, formal evaluation is endorsed as long overdue by some teachers and administrators, but is rejected as intolerable, even abhorrent, by many. The arguments on each side run roughly as follows:

Teachers and administrators who support systematic evaluation argue that teaching is a profession whose practitioners serve a large and important group of clients—the students in their classrooms—and should be held quite as strictly to account for the diligence and effectiveness of their professional work as the members of any other profession. At a time when students are more and more demanding lively, informed teaching, and ever more forcefully insisting upon a voice in the improving of their education, it is all the more necessary to know what good teaching is so that it can be fostered, to know who the good teachers are so that they can be rewarded, and (by implication) to know who the less successful teachers are so that students can be helped to avoid their courses (if the teachers themselves can neither be helped nor eliminated from the profession). Quite apart from the need to meet, wherever reasonably possible, the increasingly militant demands from students, argue the advocates of evaluation, it is no more than simple justice to assure the student a fair return for the tuition he pays to a private college or university, and to assure the citizen full value for the taxes he pays to support a public university.

Opponents of formal evaluation of teachers argue that teaching, quite simply, is an art—and as such not susceptible of evaluation by any systematic, mechanical procedure. They bring to bear on the issue testimony from well-known writers on teaching, such as Gilbert Highet, whose Preface to *The Art of Teaching* makes essentially that point. Also, teaching—opponents of evaluation contend—is a private act, a human transaction between a teacher and a learner; transactions of this sort are too delicate, too intangible, to be judged arbitrarily by any of the crude techniques currently available for the making of such judgments. The prospect of evaluation, moreover, is a threat that will lead teachers to compromise their best judgment concerning what they should be doing for their students, in the interests of reaping the rewards of a favorable evaluation, or, at least, in the hope of avoiding any penalties that might attach to an unfavorable judgment. To intimidate teachers by threatening them with formal evaluation is to discourage experimentation in teaching, and thus to deprive both students and citizens of that very excellence in professional service which a system of evaluation is meant to secure. Evaluation procedures, say teachers and administrators who resist them, create tension among teachers and damage morale, thereby weakening the ability of colleges and departments to promote effective teaching.

Assuming that this is a fair, even if by no means exhaustive, summary of the arguments on both sides of the controversy over evaluation, it is hard to deny that the arguments on both sides have some merit. To take up first the arguments in favor of systematic evaluation, one notes at once that very nearly all institutions of higher learning (according to their administrators) consider effective teaching the most important or one of the two most important criteria for awarding tenure and promotion to faculty members.[1] If this insistence on effective teaching is not to be just an empty piety, these administrators presumably need some way of determining who is an effective teacher worthy of retention and promotion, and who is a mediocre teacher not worthy of being retained. That is, in order to make the necessary decisions about faculty on the criteria they say they employ, these administrators must take for granted some effective system of evaluation. But when asked to identify the sources of data used in evaluation, many administrators refer vaguely to judgments by chairmen and deans and to comments by students.[2] Unless students' comments are carefully and systematically gathered, the remarks that carry weight are likely to be the occasional observations students make to advisers, to chairmen, and to other faculty members—quite possibly a haphazard group of comments hardly substantial enough to support any decision about a faculty member's professional future. To assert that evaluations of teaching are based on the reports of chairmen and deans is to beg the question: where do the chairman and the dean get their information? Unless they gather comparable information systematically for all faculty

members—for example, by observing faculty members as they teach in their classrooms—chairmen and deans are likely to have nothing more than occasional bits of hearsay on which to base their judgments.

For all their protestations about the importance of effective teaching in personnel decisions, then, administrators who capitulate before faculty resistance to the gathering of specific data about teaching must either base their judgments on sketchy and potentially unreliable information, or admit that they cannot honestly consider a faculty member's teaching when making decisions about his future. Indeed, it appears that many administrators today prefer (or are virtually required) quietly to assume the competence of every faculty member in teaching, unless there is massive evidence to the contrary, and to base decisions about promotion and tenure on other data. It may be the inaccessibility of data about teaching that leads to what many students regard as disproportionate emphasis on a faculty member's publications—which, after all, are tangible and durable and which for many administrators are an important form of teaching—as a basis for determining his academic advancement. Those who argue that teaching should be reinstated as the major criterion for academic advancement can properly insist on some reliable procedure for evaluation of teaching as a step toward a sound judgment of every faculty member's professional abilities.

Some observers of teaching, however, regard the search for evaluative techniques simply to assist in decisions on tenure, promotion, and salary as a wearisome occupation.[3] Although administrative judgments on individual faculty members are inevitable and ought to be made as justly as possible—the decision to retain a teacher despite want of evidence concerning his teaching skills is just as firm and important a commitment as the decision to fire a man on evidence of his incompetence—these observers have an important point. It is not personnel decisions alone that establish the quality of teaching on a university campus; the goal of any evaluation procedure ought not to be primarily to threaten teachers, but to improve the overall quality of teaching and learning on the campus. Wise administrators are properly interested in discovering what constitutes successful and effective teaching under various conditions and in publicizing their findings so that any teacher can, if necessary, alter his teaching practices in response to those findings. Wise faculty members, too, if they want to improve their own teaching (it is reasonable to expect a professional person to want to improve *his own* performance, even if he resists evaluation by others) may also be interested in discovering when and in what ways their work is most effective.[4] Indeed, according to Professor Joseph Katz, a faculty member may become intellectually and professionally stagnant unless he has the kind of feedback about his teaching that can be gained only through evaluation:

> In the absence of sufficient concepts of outcome, almost anything seems to be acceptable and in the end nothing seems to make much difference, so evanescent a thing do college courses seem to be. The absence of supervision or assessment of teaching is presented, in the prevailing ideology, as a condition of freedom and independence in teaching. But it has an unrecognized side effect in that it deprives most if not all teachers of a firm sense of accomplishment, or a firm sense of mistakes to be learned from. Teaching, without assessment, leads almost necessarily to a cycle of repetition. Only detailed, sophisticated, continuing assessment can make teaching experience cumulative, instead of repetitive, and thus instruct the instructor.[5]

To get information for this kind of assessment, which can suggest where the teaching of one man or of a whole faculty needs improvement and can suggest ways of making that improvement, is a worthy goal. It justifies efforts by administrators and faculty alike to find ways for determining when effective teaching is taking place and for understanding what makes it effective. Professor N. L. Gage, a ranking authority on research into teaching, emphasizes the usefulness of evaluation in giving insights that will help in "understanding, predicting, and controlling the teaching process."[6]

But even though the case for systematic evaluation may seem to be unshakable, teachers and administrators who resist the adoption of formal evaluation procedures offer many strong arguments for their position. To attempt development of a formal, routine system for judging a complex and humane "art," they argue, is fundamentally self-contradictory. (Who can systematize the evaluation of a poem, a painting, or a concerto?) Moreover, they argue, teaching is infinitely varied. What one does as a teacher depends upon one's goals, the kind of students one faces, the background of these students, the physical conditions in which one is working, one's own particular strengths as a teacher, and so on. Different kinds of courses require different kinds of teaching; a teacher must be subtle, sensitive, and flexible in perceiving what is needed in each course he teaches. Granted that different teaching styles can be identified and described in broad terms, as Joseph Axelrod does when he names five kinds of teachers—The Drillmaster, The Content-Centered Faculty Member, The Instructor-Centered Faculty Member, The Intel-

lect-Centered Faculty Member, and The Person-Centered Faculty Member[7]—one cannot say that any style is absolutely good or bad in itself. And each style has many variations. Each style, says Axelrod, "has its own excellence," and the young teacher needs to choose and perfect his own style.[8]

On this line of reasoning, it it outrageously presumptuous to say what constitutes "good teaching," and, by implication, what does not. Furthermore, the work of any one teacher will vary in effectiveness from course to course, from section to section, from class meeting to class meeting, perhaps even from minute to minute within the same class meeting. For instance, a teacher may be outstandingly effective in talking about the poetry of Yeats, but boring and irrelevant in talking about Dryden. Even teachers who are regarded by students and colleagues as outstanding admit that they enjoy good days and bad days. And teaching is a many-sided act, in which some parts (e.g., organizing a course, drafting assignments for essays) can be performed well at the very same time that other parts (e.g., classroom instruction) can be performed poorly. It is thus preposterous to try to label a teacher as "good," "mediocre," or "weak," and it is equally unsupportable to assert that one kind of teaching or one method of organizing a class period is absolutely superior to others and worthy of emulation. Absolute standards for "effective" or "ineffective" teaching, then, are impossible to establish; yet without such standards any system of evaluation lacks a scientific or philosophic foundation—or so the opponents of systematic evaluation argue.

Even allowing for the possibility that some precise bases for evaluating a particular job of teaching in a particular course on a particular day might be established, critics of evaluation plans point out that "reliable" evaluation of teaching—in which different judges would arrive at similar assessments of similar teaching carried on under comparable circumstances—is exceedingly difficult or impossible to obtain. Every judgment is a human act (even if a written rating form is used, the real rater is the human being who completes the form), and human beings differ in their judgments of the events they see. Persons observing the same job of teaching evaluate it differently because of different expectations, different preferences in teaching methods, different values about education, different assumptions about what should have taken place, different views about what the observers might have done in the same situation, and so on. Even if an attempt could be made to train observers to respond in the same ways to the same job of teaching, some observers might still perceive different events more sharply than would other observers, and might give different weights to various parts of what they observe, with the result, again, that the final assessments by different observers of the same act of teaching might diverge noticeably.[9] Rather than accept unc'ependable judgments about the professional success of a teacher or about the utility of various teaching procedures, critics of systematic evaluation prefer to see judgments not attempted. For these critics, fairness to teachers and wise caution in the endorsement of teaching techniques are principles to be nourished. Pragmatic gains, such as the possible elimination of persons unsuited for teaching and possible slight changes in the teaching styles of a few people, are not compelling reasons to sacrifice fairness and caution.

These arguments may convince us that absolute judgments about teachers and teaching are ill-advised. It is probably unwise to say categorically, for example, that Professor X is "no good" as a teacher or that lecturing to large classes is an ineffective method of teaching. But once determined to guard against sweeping judgments, one's next steps are not so clear. One probably has to admit, first, that teaching, though all of us regard it as a profession (and rightly so, if a profession can be defined as an activity in which practitioners, after suitable training, make complex human decisions about ways of achieving desirable human goals), is unlike most other professions in important ways. Some other professions, such as medicine or dentistry, or piloting an aircraft, are practiced publicly under strict regulation by employers and by the government, on the assumption that the physical safety and well-being of a large clientele depend in measurable ways on competent professional service. The client's physical health or personal comfort, easily observable by an outsider, is an adequate indication of how well the doctor has carried out his responsibilities. The lawyer, too, practices his profession publicly, adhering to standards of conduct established by a professional association, and his skill in carrying out his duties is recorded in the responses that his arguments evoke from judges and jurors, in the form of verdicts returned in the cases he has argued. The verdicts, one may argue, are sometimes adverse, but his "batting average" is a visible record of his success, and so is the esteem in which he is held by professional colleagues who have watched him work. Teaching, however, involves the production of subtle, often unmeasurable, effects in human beings; the evidence of professional success is not visible to the naked eye nor recorded unambiguously in official documents. For such a profession, if there is to be any effort at all to distinguish kinds

and degrees of success, distinctive and varied methods for determining success and failure may be needed. And they may need to be more systematic, more varied, more thorough to cope with the invisible, silent workings of the teacher's art.

Indeed, some argue that it is fairer to view teachers as artists (in language, in rhetoric) than to treat them as practitioners of a profession; that perspective may permit better recognition of the distinctive goals sought by teachers and the distinctive methods they employ. But, if teaching is an art, it may still be open to evaluation. Other forms of art, after all, are open to judgment by observers; it is possible for a trained observer to distinguish between a masterpiece and an amateur's painting, and as teachers of English we claim to be able to help students discriminate between, say, the poems of Donne and those of Abraham Cowley. If teaching is an art, says Wilbert McKeachie, it is an art "that builds upon knowledge and skill." And in support of systematic study of how teaching works, he adds that "research can help to lay bare the deepest properties of our teaching while revealing to us more wonderful intricacies."[10]

In the interests of giving their employers reasonable assurance of the quality of their services, then, teachers of English may want to look for ways of evaluating their art, rather than, by denying the possibility of evaluation, leaving assessment of that art to the whims of administrators and investigators who may be anxious to ride prejudices or promote biased attitudes. Before concluding that the evaluating of instruction is a lost art—if evaluation is an art at all—or one that can never be discovered, we should examine methods of evaluation to see whether they offer us any means of attaining the purposes sought by administrators and citizens—purposes that will be served, wisely or unwisely, by decisions of administrators and votes of legislatures—without substantiating the fears of those who resist deliberate evaluation. In this examination, we have a good deal of help. There is no dearth of effort among students of education to guide research into what is and what is not effective teaching. Professor Gage's *Handbook of Research on Teaching*, the *Encyclopedia of Educational Research*, and periodicals such as *Improving College and University Teaching* are only a few of the professional publications that support studies of the art of teaching.

Three distinctions should be borne in mind as we proceed. First, though the act of teaching is performed by diverse human beings in almost equally diverse ways, the evaluation of teaching and the evaluation of teachers are not necessarily the same. Besides guarding against absolute judgments on any method of teaching, we will find it worthwhile, as the Introduction indicated, to distinguish comments on techniques of teaching from comments on the overall work of any one teacher. Research on teaching procedures, as they are employed in different circumstances, may permit some general statements about these procedures; it is much more difficult to gather enough evidence about the possible strengths of a teacher to permit useful generalizations about him.

A second important distinction to keep before us is that between the performance of a teacher and the results of his teaching. When we talk of a teacher's performance, we focus upon what he does, on the observable characteristics of his art; when we talk of a teacher's results or accomplishments, we look at what happens to his students. Exactly what a teacher does before he enters the classroom, while he is in it, and after he has left it is amenable to some kinds of description; what his students do as a result of, or just plain after, his performance, requires quite another kind of observation and description. The comments we make on a teacher's performance often assume that we have the power to predict how different performances will affect students. Our ability to make these predictions wisely is much more doubtful than we sometimes think; wise doubts about it might reduce the confidence of some administrators in their methods of evaluation.

Finally, we need to distinguish the purposes for which evaluation may be carried out, or, to put the matter another way, we need to differentiate the points of view from which evaluation can be carried on. An administrator may want to evaluate teaching in order to get evidence on which to base personnel decisions, or to see what guidance his faculty needs toward the improvement of their teaching. A teacher may evaluate his work in order to improve that work. The researcher may carry on evaluation in the hope of publishing findings, stimulating experimentation in teaching methods, or even encouraging the use of specific teaching techniques for particular purposes. Students may carry on evaluation in order to advise their fellows about courses or major fields. An individual student may evaluate the teaching he has experienced in order to help himself understand the purposes of his education, the benefits gained from his study to date, and perhaps the kind of education he wants in the future. The citizen paying taxes to support a public university or the alumnus asked to contribute to his alma mater may want the results of evaluation in order to determine whether the institution is serving its community or its student body well. (Citizens, though unable to escape paying taxes, can write to

legislators; alumni can withhold checks if their schools do not live up to expectations.) No doubt other points of view and other purposes for evaluation are equally possible. Indeed the kinds of data sought, the criteria on which these data are judged, and the kinds of inferences extracted from the data will differ a good deal depending on who is carrying out the evaluation and why. Some techniques for evaluation will serve the needs of some groups of evaluators better than other techniques, and it will be necessary to incorporate such discriminations into the analysis of different evaluative techniques. The importance of scrupulous fairness in the making of personnel decisions, for example, may force administrators to reject data and data-gathering techniques that might be acceptable for students whose principal purpose is to advise fellow students concerning the attractiveness of particular courses.

With these distinctions in mind, along with some notions of why evaluation procedures are needed to begin with, we can proceed with a review of current practices in the evaluation of teaching. But first it will be useful to look at statements by various administrators and teachers concerning the acts or events that they understand to be included under the amorphous and capacious phrase, "good teaching."

[1]For data showing how widely this assertion is made, see John W. Gustad, *Policies and Practices in Faculty Evaluation* (Washington, D.C.: American Council on Education. 1961), pp. 6-8.

[2]Gustad, pp. 10-11.

[3]See, for instance, one recent monograph, Arthur Cohen and Florence Brawer, *Measuring Faculty Performance* (Washington, D.C.: American Association of Junior Colleges, 1969), pp. 52-56.

[4]See Cohen and Brawer, pp. 58-59, for a few suggested questions that look toward improved understanding and performance of teaching.

[5]"Interpersonal Relations in the Classroom," in *The American College*, ed. Nevitt Sanford (New York: Wiley, 1962), p. 376.

[6]"The Appraisal of College Teaching: An Analysis of Ends and Means," *Journal of Higher Education*, 32 (1961), 20.

[7]"Teaching Styles in the Humanities," in *Effective College Teaching*, ed. William Morris (Washington, D.C.: American Council on Education, 1970), p. 43.

[8]Ibid., p. 53.

[9]For some comments on problems in attaining "reliability" in the systematic observation of teaching, see Donald Medley and Harold Mitzel, "Measuring Classroom Behavior by Systematic Observation," in *Handbook of Research on Teaching*, ed. N. L. Gage (Chicago: Rand McNally, 1963), pp. 250 ff.

[10]"Research in Teaching," in *Improving College Teaching*, ed. Calvin B. T. Lee (Washington, D.C.: American Council on Education, 1967), p. 231.

CHAPTER II
WHAT VALUES DO WE PROFESS ABOUT TEACHING?

Evaluation, to define it very broadly, is mainly the gathering of data to determine whether or not particular events that the evaluator wishes to see take place are occurring. That is, the evaluator explicitly or implicitly adopts a set of values that he wants upheld in the activities he is looking at, and he tries to determine whether those values are being sustained. In the evaluation of teaching, of course, it is not only "events" (discrete bits of experience) at which the observer is looking, but recurrent characteristics of the teacher's behavior, prevailing conditions observable in the classroom, and recurrent observable behavior of students. But whatever the observer is look at, he is gathering data and determining whether they disclose the achievement of goals he values or the presence of characteristics he likes.

When one talks of the evaluation of teaching, this rudimentary concept of evaluation identifies the problem that we cannot avoid: is it possible to develop values that one can insist be upheld by teachers in their teaching? If teaching, as everyone concedes, is infinitely varied, there is extreme danger that by adopting a set of values concerning teaching an observer may unwittingly exclude from the category of "good teaching" actions that another observer might praise but that are not accommodated within the first observer's definition. The question is insistent: can one talk in general terms about good teaching without subverting one's efforts to be flexible in recognizing the many varieties of good teaching and the many subdivisions of the act of teaching? Certainly no one can enumerate the full range of possible manifestations of good teaching any more than one can enumerate the full range of possible manifestations of beauty. Chairmen across the country regularly express suspicion of any effort to define or circumscribe the notion of what good teaching might be or to specify characteristics that might regularly be associated with "good teaching."

In reacting so, chairmen share the feelings of authorities on education like H. S. Broudy, editor of *Educational Forum*, who writes in *Teachers College Record* that an important misconception about teaching "is the notion that one can set down in verbal form a definition or description of good teaching, such that a layman could use to identify and judge teaching performance."[1] Some educational researchers concur. David Ryans writes at the beginning of a long research study that we have made "embarrassingly little progress" in finding out "the details that are necessary for describing competent teaching."[2] Bruce Biddle writes flatly that "We do not know how to define . . . or measure teacher competence," adding that we will not be able to define teacher competence until the effects desired of the teacher are decided upon, and enumerating the fearsome array of problems that beset anyone who seeks to describe teacher competence.[3]

How then to reconcile the impossibility of establishing firm standards on which to conduct evaluation with the need for making judgments about teachers and teaching? One is tempted to try to rephrase the question "what is good teaching?" into "what do good teachers typically do?" but the rephrasing, even more than the original question, invites the futile effort to compile an exhaustive list of activities engaged in by good teachers. Similarly, to rephrase the question as "what do good teachers achieve?" implies that in order to be a good teacher, one has to achieve something, and that signs of student achievement are exclusively the result of the teacher's effort, regardless of how bright or intractable the students with whom a teacher had to work.

Despite these difficulties, evaluation is carried on. It is carried on because observers assume, despite their uneasiness with definitions and despite the recognition that no one can list the full range of "good" teaching activities, that some elements associated with good teaching can be named and will be agreed to by the majority of persons who are concerned with the improvement of teaching. This belief has acquired strong support recently from thoughtful investigators of teaching across the country. In particular, Dr. Kenneth Eble, Director of the AAUP-sponsored Project to Improve College Teaching, is convinced that one can talk usefully about the characteristics of good teaching and good teachers. "For all that the diversity of students and teachers draws attention to the variables that make up good teaching," he writes, "such variables are best perceived as diverse personal manifestations of commonly agreed upon qualities." Talking about teachers' work in class, we find, says Dr. Eble, "a relatively small number of characteristics which relate to a teacher's effectiveness."[4] Since the methods of collecting data for use in the evaluation

of teaching must be judged on how well they help an observer to see whether his values in teaching are satisfied by a particular course or teacher, reviewing the conceptions of effective teaching currently held by department chairmen (and deans) is a useful preliminary to any survey of methods for gathering data. Some chairmen spoke out willingly on their values in teaching; others, while voicing reluctance to formalize conceptions of good teaching, revealed indirectly their perspectives and emphases in their assessment of teaching in their departments or schools.

In what follows, then, I record from the general statements of administrators (mostly chairmen, a few deans), or infer from comments on problems of evaluation written by those who did not themselves generalize about teaching, qualities that my correspondents evidently associate with what they consider "good teaching." To make sense out of the diverse conceptions of good teaching, I have tried to group into reasonably discrete classes the qualities that chairmen suggested they look for. For each group of qualities, I add a comment on its usefulness as a basis for evaluation of teaching.

(1) *Administrative Dependability.* Quite a few chairmen, whether they were trying to evade judging what takes place in the classroom and in the student or were simply unconcerned with substantive matters, identified conscientious execution of specific procedures as their principal criterion of professional competence. The standards they used were built around such routine accomplishments as: meeting classes regularly and on time, keeping a minimum number of office hours and being available to students, returning annotated written work to students promptly, giving examinations on time according to prescribed departmental procedures, getting grades to the registrar on time, and so on. Although few chairmen listed procedural conformities as their only criterion in evaluation, some listed these kinds of behavior as their principal concern, and several enumerated these items of behavior alongside more substantive acts as if the two sets of activities were, for practical purposes, of equal importance. It is hard to believe anyone can assume that good teaching is equatable with meeting routine administrative expectations, but if an administrator distrusts procedures for gathering information on less tangible, more subtle qualities of teaching, a teacher's conformity to administrative procedures may be all he can judge. Administrative conformity, at least, is physically and empirically observable, like a technician's following a manual in operating a machine.

(2) *Diversity of Activities.* A few chairmen expect that a good faculty member will teach "variously." They expect him, that is, to be willing to handle lecture courses, seminars, courses in the literary history of a period, courses in techniques for literary analysis, and (presumably) individual instruction and guidance of students, with equal vigor. Evidently they also expect him to develop new courses and new instructional techniques—courses that reflect new ways of looking at a field of study and new procedures for getting students interested in it. The willingness to try new things, the refusal to persist in a single pattern of teaching or to teach only one or two kinds of courses, is for these chairmen evidence of intellectual energy and up-to-dateness. The emphasis in these comments fell upon teachers' doing the things mentioned—at least passably well, one must presume—rather than upon the way they are done or upon the results, good or bad, from the doing of them. Of course, the diversity of a teacher's activities is fairly easy for an administrator to check on. The courses a man is teaching and the use he is making of films, tapes, records, and other modern techniques are fairly easy to find out. The information may, indeed, give a clue to the teacher's energy and originality.

(3) *Personal Qualities.* Several chairmen and some students who answered my inquiries listed as characteristic of a good teacher some qualities observable as well outside the classroom as in it—indeed observable in ordinary conversations about subjects other than teaching. (My correspondents assumed that the good teacher would reveal these qualities in the classroom; I list them separately from classroom acts because they can be observed outside of class.) Among these qualities are a clear and pleasing voice, an ability to communicate complex and abstract ideas clearly, and a broad awareness of developments in contemporary society and culture. Many chairmen said that they valued objectivity of outlook and openness of mind, a willingness to consider the value of new approaches to subjects under discussion, flexibility in argument, and fairness in the appraisal of evidence. One chairman, speaking of the teacher's work in his classroom as well as outside it, insisted upon the importance of personal and intellectual humility. Also to be counted, perhaps, among "personal" characteristics (because they do not connect only with a man's activities as a teacher) are his knowledge of his subject—the subjects of the courses he teaches—and the up-to-dateness of his scholarship on these subjects.[5] A desire constantly to improve his teaching is another "personal" quality valued by many of my correspondents. Since these qualities are inferrable from social

interaction with a teacher, as well as from professional interchange, they can be measured by what we will later call "circumstantial evidence."[6] The trouble with emphasizing personal qualities thus observed, of course, is that a man's behavior in the classroom may be quite different from his behavior outside it.

(4) *The Planning of Instruction.* For some chairmen, it is clearly one of the teacher's responsibilities to organize his courses according to some plan, so that they will proceed at a reasonable and suitable pace; it may also be part of his responsibility to prepare clear and informative syllabi for his courses. Obviously, not every teacher is going to prepare syllabi and other written teaching materials; many faculty members teach largely without printed handouts, and assessment of the organization of their courses without continuous attendance at class meetings is difficult. Still, where syllabi are prepared or where documents that reveal organizational plans for a course are distributed, most chairmen hope that these plans will be orderly, comprehensible, and useful to students. The value of these documents, where available, can also be assessed without direct observation of classroom activites and possibly without polling students or gathering supplemental data about how classes are managed. (See Chapter v for further discussion of the use of teaching materials in evaluation of instruction.) The number of chairmen who said they valued orderly syllabi and handouts was not large, but several were emphatic in saying that they put great value on orderly, specific instructional plans and materials.

(5) *Classroom Procedures.* The performance of a teacher in his classroom—to be distinguished from his effect on students, which will be considered later—was naturally an element in the teacher's work to which chairmen gave much attention (though it is only one part of the complex art of teaching). One important characteristic of a good teacher, for many chairmen, is that he prepares conscientiously for his classes. While careful preparation for class is probably desirable in most teachers (though some teachers can teach "well" with little or no preparation), assessing a teacher's level of preparation is often difficult—a fact not always observed by those who cited this feature as a characteristic of a good teacher. Second, chairmen value the clear, accurate communication of ideas to students—in such a way that the ideas can be retained. To the extent that a class period is devoted to communication by an instructor to students (some would deny that such ought to be the principal emphasis in any class), it may be reasonable to expect clear and precise communication, though to assess that clarity and precision, again, is not easy. Also expected by many chairmen, and also hard to assess, is clear organization of materials presented in the classroom, with due emphasis on the important concepts and suitably less attention to the unimportant. A preference for order in the presentation of data and in class discussion is understandable, though a lively discussion or an exemplification of a mind working energetically and illuminatingly on its subject will not always be an "organized performance." Also expected by some chairmen is the use of illustrative materials to illuminate large subjects and the application of general ideas to specific problems (e.g., specific literary works)—the bringing down to earth of theoretical concepts to show the student how these concepts apply to today's world, and how they relate to other fields of study. This preference for illuminating presentation is related to the desire for clear transmission of data; it seems a reasonable wish, although some teachers are able to move their students even if what happens in their classes does not exhibit the pattern of generalization plus illustration or application. However the teacher organizes his material, he is expected by chairmen and students alike to display "enthusiasm" for his subject, though chairmen are notably reticent in describing the signs of this enthusiasm. Chairmen evidently believe that every good teacher should try to communicate to students the order, beauty, and value of his subject—even, it seems, where the "subject" is material whose direct value to the student appears negligible.

Finally, many chairmen expect teachers to be quite flexible in their classroom techniques, and to adjust their teaching carefully to the needs and interests of their students. Many chairmen also stress the importance of having a fair, well-planned program for evaluating students' accomplishments, though this element in a teacher's work is less prominent in chairmen's comments than the teacher's thoroughness of preparation, clarity of communication, and enthusiasm for his subject.

Some chairmen are quite specific, too, about desirable teaching methods, preferring discussion and sharing of ideas among teacher and students, rather than lecturing. Although lecturing is recognized as a necessary teaching technique in large classes,[7] the insistence on student participation in discussions is marked in most responses to my inquiries. These respondents also prefer teaching in which the faculty member is responsive to and considerate of comments and ideas from students. The model of a community discussion, a fellowship of equals, is for many chairmen the preferred way of envisaging classroom activities. One chairman summed up what he valued in classroom activity by

saying that in a good lesson students' observations and questions should be tightly incorporated into the flow of ideas; another said, similarly, that he thinks the good teacher draws students into the process of teaching—by which he meant, I take it, that the good teacher helps students contribute to a learning process in which teacher and students are sharing alike. The class that proceeds in such a manner may ultimately be more "effective" than one that consists simply of a lecture or a neat sequence of questions and answers.[8]

Conspicuous by its absence from the comments of most chairmen about good classroom techniques is praise for the use of audiovisual media and other technological aids to instruction. Few chairmen in English and Humanities, evidently, are committed to the view that use of media is essential in "modern" teaching or that it necessarily constitutes a point in the teacher's favor.

(6) *Eliciting Student Work.* Far less prominent than the emphases just discussed, but still noticeable among the preferences of chairmen and the compilers of student-rating instruments, is skill in setting assignments for writing that will encourage students to engage in fresh, creative thought. The ability of the instructor to devise and administer assignments and respond to student writing (and speaking), whether on examinations or essays, is for some chairmen an important attribute of a good teacher. Related to skill in setting assignments for writing is the ability to evaluate student work fairly and to criticize it constructively. The writing of perceptive, constructive comments on students' work is a central part of effective teaching—as important in some courses as classroom activities.

Assignments for papers, and examinations, especially if they have been duplicated, are of course available for examination. So are student papers, if faculty members or students allow others to read them. They remain tangible indications of what took place in a course (and what has happened in students' thinking), and thus are among the more durable and perhaps more trustworthy kinds of data that can be gathered. In Chapter vi, I discuss further the use of teaching materials in the evaluation of instruction.

(7) *Relationships with Students.* Chairmen and the compilers of student questionnaires place considerable value on productive relationships between instructor and students; inside and outside class, the good teacher is expected to achieve that intangible, indefinable condition we call "rapport" with students. I sense that, by "rapport," chairmen mean a feeling among students that they have a common purpose, a common goal with their teacher, that they are willing to collaborate with the instructor eagerly in the pursuit of knowledge. More deeply, perhaps, "rapport" implies that students are willing to trust an instructor, to believe that he has their best interests at heart, to carry on the inquiries he encourages and to accept the judgments he makes of them—all in the belief that the instructor is acting in the best interests of his students.

But this sense of common purpose is not all that a teacher is expected to achieve with students. Being accessible to students is, for some chairmen, essential in a good teacher; patience and a genuine friendliness of manner in dealing with students are likewise important. Willingness to help individuals in their courses and in their careers, respect for students as people, concern for students' total well-being—all these qualities, too, are expected of a good teacher by various administrators. Such sympathy and affection are, of course, impossible to measure objectively. They are *felt* by students, perhaps by a faculty member's colleagues. But the availability for conferences, the designing of special projects for particular students, the posing of questions that draw out the interests of individual students, may be signs of such feelings. So may efforts to identify particularly promising students and to encourage them to follow congenial academic careers. So, too, may the teacher's ability to describe specifically the strengths, weaknesses, and problems faced by individual students in his class. That ability, indeed, is taken by some administrators as a crucial index of the teacher's perceptivity and skill.

(8) *Individualizing Instruction.* The good teacher's task is not only to develop confidence and trust in students. He must, in the judgment of many chairmen, make special provision for the needs of individual students. For many chairmen, the adaptation of teaching procedures to the needs of individual students is essential, as are the inventiveness and resourcefulness needed to meet the needs of students with special problems. Individualizing instruction, of course, is a part of a teacher's technique; possibly some chairmen and instructors might not agree on the value of such a technique for much college teaching.

(9) *Impacts on Students.* In their recent book on the assessment of faculty performance in junior colleges, Arthur Cohen and Florence Brawer say that the ultimate criterion for the evaluation of teaching and teachers is whether learning results.[9] For them (as for many writers on teaching), what counts in the assessment of a teacher's work is his effectiveness in causing learning, not the characteristics either of his performance or of his personality (which are viewed only as factors *contributing* to his effectiveness). Cohen and Brawer urge us to ask

whether a teacher has had an impact upon students, and what kind of impact, before saying whether or not he is successful.

Many chairmen and most compilers of student questionnaires give some attention—some a good deal of attention—to a teacher's impact on his students, though they give less attention to teachers' impact than Cohen and Brawer would presumably like to see. Such attention, of course, brings at least some risk of unfairness to the instructor, because although he may be in control of his performance, of the plan for his course, of his assignments, and of the attitudes he exhibits to students, he is not necessarily in control of the responses of his students, who have after all a lifetime of experience before they enter any teacher's classroom and whose experience may so condition their attitudes that no teacher, however skillful, can have much of an impact. On the other hand, for Peter Elbow, a teacher of English, students' unpromising backgrounds do not exonerate the instructor who does not "produce understanding." Regardless of students' abilities it isn't enough simply to "show up" in the classroom "with the goods," i.e. with well-ordered knowledge and sophisticated teaching techniques, if the student doesn't learn.[10]

Quite apart from possible unfairness to the teacher in making the judgment of his effectiveness turn on what happens to his students, the criterion of whether or not a teacher "causes learning" brings up immediately the question of what constitutes "learning" and what kinds of learning a teacher may be expected to cause. A few chairmen and a few compilers of questionnaires think it sufficient if a teacher causes his students to learn facts from memory or to learn the teacher's judgment on, say, a work of literature. But the "learning" desired by most chairmen is the spirit of critical inquiry and independence of thought—curiosity, willingness to explore, determination to reject pat answers in favor of more accurately complex perceptions, willingness to see familiar materials in fresh ways. The teacher who is thought by most chairmen to be making a desirable impact is the one who sets before his students a challenge to their interpretive powers and who guides them in meeting that challenge to the student's satisfaction. To put the point another way, a successful teacher is the one who causes his students to want to become solvers of problems that face professionals in the discipline and who helps them learn ways of solving those problems. Such is clearly the view of Peter Elbow, for whom teaching is helping the student discover "how to open out new questions, new problems, unnoticed premises," and "how to deploy and assess various and contrasting sorts of thinking." The good teacher produces "in the student the ability to make problems soluble by being able to see and feel new questions."[11]

Obviously the student will want to become a problem-solver in a discipline only if he develops interest in that discipline; a good teacher, therefore, encourages students to develop that interest, to inquire further into the subject, to desire mastery of it. The desire to exhibit excellent workmanship, and joy in the pursuit of learning, are other attitudes students will develop if a teacher is good, according to some chairmen. A few chairmen expect teachers to lead majors and nonmajors alike toward a desire for active participation in learning.[12]

Although it places a teacher partly at the mercy of his students in ways that he cannot possibly foresee, or, having foreseen, may be unable to control, ability to "cause learning" or "produce understanding" seems far more important in identifying a "good" teacher than some chairmen and deans appear to allow. A teacher, after all, is a professional person, and professional persons can be judged in part on the results they obtain in practicing their professions. Engineers, for example, are judged in part—maybe more than in part—on whether the structures they design stand up. But pushing this analogy very far discloses at least one serious difficulty in the proposition that "causing learning" is the ultimate criterion for evaluating teachers. If the engineer's bridge stands, it is there for all to see and traverse. But, particularly in English and the Humanities, if the teacher's student has learned, how do we know it? We cannot measure it by the student's ability to name authors, dates, and works; we cannot measure it by the student's ability to tell us the various interpretations of an ode by Keats and the originators of each interpretation; we cannot measure it by the publication of a student's thesis on a familiar or obscure twentieth-century poet. To be sure, it may be possible to discern at the end of a course whether a student can talk more "perceptively" about the art of a poem than he could at the beginning, and it may be possible through well-chosen examination questions to discover whether a student has a better understanding of the development of the novel in America since 1900 than he had before he began the course. But the gradations of improvement in these skills or kinds of knowledge are at best subjectively perceived, and it is hard indeed to compare the effectiveness of two teachers—each of whom may be competing for a single vacancy on a permanent faculty—by, say, asking them to teach sections of the same course and then comparing the results of examinations written by students who may be from altogether different backgrounds. Thus, as Chapter

vii will establish further, difficulty in defining the kinds of learning desired in students and in saying exactly how these learnings are manifested in what students say or write is, in English and the Humanities, a major obstacle to use of the learning he "causes" as a major criterion in determining a teacher's effectiveness.

Although not every chairman or compiler of a student questionnaire would use the same categories I use in grouping statements about what administrators value in teaching, these nine categories sum up at least approximately the items mentioned today when chairmen talk of what they look for in teachers (and when one looks at the qualities of teaching that seem to get the most favorable loadings in student questionnaires).[13] To exhibit most of these qualities requires no particular teaching techniques or procedures, although some (such as attentiveness to students' comments in class and the establishment of a classroom community of equals) come closer than others to mandating a particular teaching procedure; a teacher could achieve most of the desired qualities while using any of a number of teaching techniques. Our main task, to which we now proceed, is to determine what procedures are used for gathering data that will reveal how well a teacher achieves what is expected of him, and also to determine the value of each procedure and the intelligence with which it seems currently to be applied.

[1]"Can We Define Good Teaching?" *Teachers College Record*, 70 (1969), 585.

[2]*Characteristics of Teachers* (Washington, D.C.: American Council on Education, 1960), p. 2.

[3]"The Integration of Teacher Effectiveness Research," in *Contemporary Research on Teacher Effectiveness*, ed. Bruce Biddle and William Ellena (New York: Holt, Rinehart, and Winston, 1964), pp. 3-4, 18-19.

[4]*The Recognition and Evaluation of Teaching* (Salt Lake City, Utah: Project to Improve College Teaching, 1970) p. 9.

[5]Professor Wilbert McKeachie voices some doubt about whether extensive knowledge of his subject is truly an important element in a good teacher's success. The assumption that it is, Dr. McKeachie says, "has never been checked, and, conceivably, students might become better educated by a confused or ill-informed instructor who motivated his students to clear up the confusions than by a professor with great depth of knowledge. . . . The whole area of content has been neglected in research on teaching." ("Research in Teaching," in *Improving College Teaching*, ed. Calvin B. T. Lee (Washington, D.C.: American Council on Education, 1967), p. 212.)

[6]David Ryans' *Characteristics of Teachers* is an important work for those interested in what is known about the personal and temperamental qualities of teachers and the connection of those qualities to the performance of teachers. Arthur Cohen and Florence Brawer also discuss personal qualities of teachers in their monograph, *Measuring Faculty Performance* (Washington, D.C.: American Association of Junior Colleges, 1969), pp. 17-20. Their conclusion is that correlations between a teacher's personality and his performance are very difficult to establish.

[7]On the values of lecturing, see W. J. McKeachie, "Research on Teaching at the College and University Level," in *Handbook of Research on Teaching*, ed. N. L. Gage (Chicago: Rand McNally, 1963), pp. 1126-27.

[8]For comments on the value of orderly discussion, see Joseph Katz, "Interpersonal Relations in the Classroom," in *The American College*, ed. Nevitt Sanford (New York: Wiley, 1962), p. 385. See also comments in Chapter vii, below, that tend to support the preferences of the administrators cited here about teaching methods.

[9]*Measuring Faculty Performance* (Washington, D.C.: American Association of Junior Colleges, 1969), pp. 52-54.

[10]See his "The Definition of Teaching," *College English*, 30 (1968), 187.

[11]"The Definition of Teaching," pp. 187-88.

[12]Further comments on problems of evaluating teaching by reference to learning caused are found in Cohen and Brawer, *Measuring Faculty Performance*, pp. 61-66.

[13]That the list is reasonably comprehensive can be confirmed by referring to a recent article by a student of the teaching of history: Charles E. Gray, "The Teaching Model and Evaluation of Teaching Performance," *Journal of Higher Education*, 11 (1969), 638-41. The preceding paragraphs incorporate almost all of the positive values associated with teaching, as listed by Professor Gray.

CHAPTER III
STUDENT EVALUATIONS OF FACULTY AND COURSES

By Rating Instruments

Student rating instruments have been in use on some campuses around the country for over forty years, but the recently intensified demands by students and members of the community that teachers be held accountable for the quality of their professional work have greatly increased interest in the use of such instruments as a means of evaluating teachers' performance. These pressures have combined with students' newly awakened desire for some control over the quality of their education to force many institutions that had not previously employed student ratings to search for formal, even elaborate, procedures for gathering and interpreting such ratings. Nearly half of the chairmen who responded to my inquiries about evaluation of teaching said that some use was made of student rating instruments at their schools, and several other chairmen said that faculty and/or student groups were actively at work on plans to use such instruments. In some institutions, these instruments are optional, informal means for helping faculty members to get some feedback on their teaching; in other institutions, however, they are an important means by which faculty and administration together monitor the quality of teaching throughout the school. In this discussion, I will first consider rating instruments used by faculty and administrators for their own purposes (some of these comments will apply to student-administered rating forms, as well); afterwards I will turn to special problems evidently associated with student-managed rating programs.

It is worth noting at the outset that in schools where students rate teachers, ratings are usually gathered for courses taught by faculty in all ranks; in this respect student ratings differ from other techniques for evaluation, discussed later, which often apply only to nontenured teachers. Most schools using rating instruments developed by the administration or with administration support make the instruments available to all courses in all departments, or require their use in all courses. On the other hand, participation in the rating programs is as often voluntary as it is compulsory; in many schools, faculty are invited or urged, but not required, to secure evaluations from students. sometimes departments vote as a unit to participate or not to participate in the programs. In a few schools, special rating forms are regularly used to monitor the work of teaching assistants and new instructors, while they are optional for other teachers. The number of schools using special forms for younger teachers, according to my information, is not now large.

The central argument in favor of student ratings is that students are the only persons who see teachers regularly while they are in the act of practicing their profession, and that consequently students are much better situated to judge the performance of a teacher and even the effects of his teaching than any other persons. Administrators seldom see a teacher at work in his classroom; colleagues in some schools see each other in the classroom little more frequently than do administrators; persons outside the university almost never see the teacher at work. Even where observation of classes is a regular practice, observers may see at most three or four of any one teacher's classes in a given semester. A teacher's students, however, see him forty to forty-five times a semester in class and may see him several times in private consultation. Even if the student lacks skill as an observer, he has uniquely frequent opportunities to get an impression of whether or not the teacher is effective.

And, say the advocates of student ratings, the student—as "consumer" of the teacher's "services"—is morally entitled to make known his judgment of the quality of those services and to demand their improvement if they are less than satisfactory. If those who are taught are satisfied, pleased, or excited by the teaching, their responses constitute an entirely adequate judgment of the quality of the service rendered. Indeed, says the coauthor of one study of methods for gathering student ratings, "good teaching" almost has to be defined operationally as teaching that the students being taught think is good. "Good teaching . . . produce[s] students who believe that they have been well taught and have learned."[1] Students' conviction that they are experiencing inferior teaching, advocates of evaluation argue, ought to impel teachers and administrators alike toward the improvement of that teaching or the elimination of poor teachers. On the other hand, student ratings can protect good teachers against "evaluation by gossip or hearsay," and from complaints of a dissident few. And they can help to disclose strong and weak points in the teaching throughout a school.[2]

Further, student ratings can offer teachers feedback that will help them improve their teaching.

That possibility is of primary importance.[3] And in the process of completing the rating forms, Wilbert J. McKeachie suggests, students may be led to consider the purposes of higher education and the value of their own education.[4] Students may be encouraged, also, to consider thoughtfully what they consider to be good teaching, and why. And if the results of the ratings are made public, they can generate discussion of what is to be valued in teaching—discussion that may stimulate interest in improved teaching on the campus. Results of students' ratings may also suggest, as participants in the recent Conference on Evaluation of the Project to Improve College Teaching agreed, some of the students' important attitudes toward education and some of the emphases they want to establish in their education. Last, but by no means least, student ratings of teachers, when combined with other techniques for analyzing and describing instruction, can lead to useful information about the process of teaching and learning, about what actually brings about learning, and, perhaps, about whether and how a teacher truly affects his students' learning. These last three gains, I think, are yet to be secured, since student ratings are nowadays employed mainly for improving instruction and (where students manage the rating program) to furnish students with advice about courses. But they are gains worth seeking and securing, if available, from any process of evaluation.

Most of the a priori objections to the rating of teachers by students—principally the arguments that students are incapable of making reasonable judgments on their teachers, and that their ratings will be capricious and biased—have been answered to the satisfaction of all but a few authorities on evaluation. Irvin Lehmann wrote in 1961 that "Students are perceptive, and they become more so when they realize that their opinions are seriously regarded."[5] Wilbert McKeachie cites research indicating that "students do seem to know when they are learning," and observes that "teachers rated as effective by students tend to be those teachers whose students learn the most."[6] The committee at Princeton that investigated the desirability of employing student ratings found that students "made clear discriminations between the parts of a course such as lectures, readings, and precepts."[7] On good questionnaires, say the Hazen Foundation's compilers of its booklet, *The Importance of Teaching*, students turn out to be capable of fine discriminations between such elements as a teacher's personality and the organization of his lectures.[8]

Professionally conducted research on student ratings indicates that they are on the whole reliable (different observers with similar training and observational ability looking at the same teaching act will record similar judgments), stable (raters will give the same judgments on the same teaching months and even years apart), and valid (corroborative evidence suggests that they measure what they claim to measure).[9] Fears that students' ratings will be biased by students' grades, sex, age, and field of concentration, and by the difficulty of the course or by the amount of work required appear to be unfounded,[10] although in one study students who expected to receive a high grade rated instructors and their grading policies more favorably than students who did not expect a high grade.[11] There is some evidence that upperclassmen and graduate students give higher ratings than do freshmen and sophomores, and that teachers in the lower academic ranks may be rated lower than teachers in upper ranks.[12] In some studies, instruction in large classes has been rated lower than in smaller classes,[13] and instruction in required courses lower than that in elective courses.[14] The University of Washington did find a tendency for students to rate instructors rather high on the whole, especially since the university administration was to see the results (a tendency corroborated by a shortlived experiment in student-managed evaluation of instructors at the University of Hawaii).[15] On the other hand, one committee investigating the possible use of student ratings noted with satisfaction a study that indicated no significant differences between student evaluations and supervisory evaluations of a group of 286 teachers[16]—a finding that corroborated earlier studies showing substantial agreement between student ratings and peer ratings of the same faculty members.[17] The fundamental objectivity and validity of student ratings is consistently established by investigators operating independently of each other, and most writers on the subject accept the proposition that student ratings have enough value in principle so that, at least, they cannot be overlooked.[18]

But if the principle that student ratings have some value in the assessment of teaching is rather widely accepted, the rating instruments used to gather student opinions admit few generalizations. Some instruments are one-page long and comprise ten or twelve questions; others run to six and eight pages, comprising seventy or eighty questions. Most of the instruments are designed for machine scoring and interpretation; this means that they invite the student to choose one (occasionally more) from a number of possible responses to questions or statements. Sometimes the responses are read by an optical scanner and sometimes they must be tabulated by hand before being analyzed on a computer.

In writing items for inclusion in a rating instrument, the compilers usually attempt to identify elements of a teacher's performance by which one can discriminate between the successful and the unsuccessful teacher. That is, they try to identify elements of behavior that most students would agree are characteristic of a good teacher and absent from the work of an ineffective teacher, and elements that most students would agree are characteristic of an unsuccessful teacher and absent from the work of an effective teacher. They then try, in short phrases, statements, or questions, to name these elements of behavior and invite students to signify whether the teacher whose course is being evaluated exhibited these kinds of behavior. By their responses, students both describe and evaluate the instruction they have received (some rating instruments invite more *description* of the teacher's actions than judgment on them). Elements of a teacher's performance to which students are often invited to respond include:

- the teacher's evident mastery of his material
- his ability to inspire confidence in his knowledge of his subject
- his organization of his course
- his ability to present his ideas clearly
- his ability to clarify abstract and difficult concepts or principles
- his self-confidence
- the degree to which he emphasizes the important points and minimizes unimportant details or tangential matters
- his evident preparedness for class
- his use of examples to illustrate general points
- his sense of humor
- his enthusiasm for his subject
- the clarity and reasonableness of his written assignments
- the appropriateness of his examinations
- the value of his comments on students' written work
- the fairness of his grading
- the extent to which he adjusts his emphases and techniques to the ability and experience of his students
- the "intellectual level" at which the course was taught
- the speed with which he lectured
- his personal mannerisms
- his use of classroom techniques that are apt for helping students achieve the goals of his course
- his application to current affairs and everyday life of the ideas he presented
- his use of new instructional materials

Most rating forms also ask about the instructor's interaction with his students, by inviting responses on such items as:

- his interest in whether his students are understanding and following his ideas
- his evident respect for his students and interest in them as people
- his ability and willingness to encourage discussion among members of the class
- his responsiveness to student questions and comments
- his tactfulness in dealing with students
- his tolerance for opinions that differ from his own
- his accessibility to students
- his willingness to give students credit when they have done well

And some forms ask students to consider the effects of the teacher on their desire to learn and their habits of mind, as in the following items:

- whether he gave the student new viewpoints or fresh ideas on the material
- his success in developing in students habits of critical thinking and cautious evaluation of evidence and ideas
- his ability to stimulate students to think for themselves
- whether the course, for the students, attained its stated objectives (students may be asked to say what the apparent objectives were)
- his ability to generate interest in his subject
- his ability to move students to produce original work
- his ability to move students to do their best work
- his ability to inspire students to take responsibility for learning about the subject on their own
- the extent to which he stimulates students to prepare for class carefully
- how well the course brought about for the student each of several specified outcomes
- his concern for the long-range as well as the short-range needs of his students
- whether or not the course increases, has no effect on, or diminishes the student's respect for and concern with spiritual values (this item is found only on rating instruments used at church-connected institutions)

For many writers on evaluation, questions about the effects that students perceive the course to have had upon them are among the most important and useful items, since they get at the success of the course in causing learning—which after all is what the course is for.

Sometimes the questions touch small details, referring to an instructor's making clear the objec-

tives of the course at the beginning, his use of audiovisual aids, his personal appearance, the tone of voice in which he lectured, his use of leading questions, the clarity with which he established deadlines for submission of written work, his willingness to extend office hours so as to see more students, the care he took to prevent cheating, whether he wrote difficult words on the blackboard and explained them, and so on. Most rating instruments are restricted to questions that might apply almost equally well to any course in any subject, regardless of how taught. Few instruments include questions specific to a particular course, although some instruments are designed for use by students taking a single course, and these instruments sometimes include questions that would not apply outside the one course.[19] Some rating instruments make provision for the instructor using them to supply additional items of his own devising, but this practice of permitting an instructor to adapt the instrument to his own purposes is not common.[20]

The ways in which the students can record their opinions on rating instruments also vary. Some of the typical kinds of responses invited by rating instruments include:

- an indication of whether a statement is true or false
- an indication of whether a statement is always, usually, often, sometimes, or never true
- an indication of whether the student strongly agrees, agrees, is uncertain, disagrees, or strongly disagrees with a particular proposition
- an indication of where on a scale of one through five or one through ten (or A through E) the student would rank the instructor on that particular item
- an indication of whether a student's judgment of the instructor in reference to a particular item is very favorable, favorable, uncertain, unfavorable, strongly unfavorable
- an indication of which of several (up to seven or eight) assertions about a given subject (e.g., the degree to which an instructor motivates his students to do their best work) applies most accurately to the instructor
- an indication of whether an instructor was judged, in reference to a particular item, to be outstanding, superior, competent, only fair, of less value
- an indication of which of four or five (or more) answers to a question (e.g., how well was the material of the course organized?) best describes the instructor's performance
- the selection of a word or phrase to complete a statement of which the first part is given
- the circling or checking of whichever propositions (from a given list) apply to the course
- the indication of whether the student would be sure to make, might occasionally make, would seldom make, or would never make a particular statement about the course (statements offered are sometimes statements of fact, sometimes statements expressing various favorable or unfavorable judgments about the course)

Of these different kinds of responses, the rating of an instructor on a numerical scale and the selection of one out of four or five possible statements about a topic (or the selection of one of several possible answers to a question) appear to be more common than the others. The goal of those who compile the forms evidently is to make possible swift and accurate tabulation of responses, so that results can be interpreted quickly by a computer.

Some rating instruments invite different kinds of responses to different groups of questions. One, for example, invites true-false responses to a series of items about the teacher, then asks for ratings on a scale of 1-5 on items asking how much the student has learned from the course on a given subject, and concludes by asking the student to rate on a scale of 1-5 the value of the course for a student to whose professional interest the course is related, and for a student seeking just a general knowledge of the subject.

Occasionally a "semantic differential" test accompanies other items. The student is asked to say where he would locate the course on a line representing a continuum running from one adjective to its opposite. For example, the student might be asked to say where he places the course on a continuum from "exacting" to "easy" or from "valuable" to "useless." Instruments that invite students to register a variety of kinds of responses are evidently designed on the assumption that a more sensitive monitoring of the instructor's performance can be obtained by asking the student to take a variety of perspectives on the instructor's work, as well as on the recognition that different elements of an instructor's behavior and of students' feelings are best identified by different kinds of statements and questions.

Rating instruments often end by asking the students summary questions such as these: How valuable is this course in comparison with all of the other courses [or sometimes only in comparison with the other courses in this field] that you have

taken at this school? How effective is the instructor in comparison with all of the other instructors whose courses you have had at this school? What is your overall rating of the effectiveness of this instructor? Would you like to take another course from this same instructor? Would you recommend this course or this instructor to a friend who asked your advice about courses to take? Some rating instruments also provide space for students to give their academic major, their expected grade in the course, and their grade-point average. Some correlations have been found, as mentioned above, between the grade the student *expects* in the course and the ratings he gives the instructor.[21]

One form of rating instrument that has generated some interest among specialists in evaluation but is evidently used very little in colleges and universities today is the so-called "forced choice" instrument. In such an instrument, the rater is presented with a series of statements that have been found experimentally to be about equal in "attractiveness" or "social acceptability" to raters—[illegible] one presumes, to raters who are commenting favorably on their instructors—but differ appreciably, according to experimental evidence, in the extent to which they discriminate effective from ineffective teaching. The rater is required to choose which one or two of the several statements within each group (there are usually four or five statements in a group) most closely or least closely describe the teacher. Since the statements look equally attractive to the rater, he presumably chooses those that really describe the teacher as the rater sees him, rather than as he might like the teacher to appear to those who are interpreting his responses. Instruments of this sort, research studies indicate, overcome student raters' tendency to leniency, and may increase the raters' objectivity, but the relative newness of the technique and the time-consuming professional effort (by men skilled in evaluation and testing) required to prepare the instrument probably account for the lack of frequency with which those who develop instruments use the technique today.[22]

Appendix A in this report contains a sampling of rating forms used for student evaluation of faculty. The first, recently introduced at the University of Washington, replaces there the instrument discussed in Dr. Langen's article (note 19, below). (The earlier instrument, discussed in the article, has been adopted as is or with slight modifications by several other schools.) Taken together, these eight forms (four of them from the same campus) illustrate the [illegible] of items included, the mix of items, and the [illegible] in which judgments are asked for on many [illegible] forms.

Relatively few rating instruments make provision for free written responses by students to general questions, or to questions that would require an opinion in the rater's own words. Some questionnaires do permit students to comment on their responses to multiple-choice questions. A few others invite the student to say what he liked most about the course and what he liked least. Still others give space for the student to make recommendations for the improvement of the course. Some ask the student to summarize what the course has contributed to his education, as compared with other courses he has taken. A very small number of questionnaires ask exclusively for free comments on various elements in the teacher's performance and ways in which he has affected students. And few questionnaires ask students to give their reactions to the subject they have studied, although, as a recent commentator observes, indirectly stressing the importance of a teacher's "affective" impact: "Criteria [for evaluation] . . . must be established . . . [which] take into account . . . student satisfaction. Most people would agree that we will have accomplished very little if as a result [of our teaching] the student [can] demonstrate his mastery of the concept taught while . . . his attitude toward that concept and his willingness to apply it or engage in further study of it have been destroyed . . ."[23]

Reports of the results obtained from using these rating instruments also vary widely in form, the form depending in part on the purposes for which the report is prepared (feedback to the instructor concerning students' reaction to his course, comparison—for the benefit of students—of the instructor with other instructors in the same subject, summary information for administrators, etc.). Some reports simply show the total number of students answering a particular item and the number of students who gave each possible answer (e.g., the number of "true" reponses on a true-false item). Many reports assign numerical values to the answers given by students, and report an arithmetic mean of the numerical values of the answers on each item. Some not only give an arithmetic mean but indicate the quintile or decile (among ratings of all faculty) into which the instructor's score on that item falls. Some reports show, in other ways, where the instructor stands on each item in reference to the average of all instructors rated in the same course, or in the same field (such as Humanities) or throughout the university. (Some researchers on evaluation suggest that the most useful comparisons are those among teachers of the same academic rank in the same kinds of courses—e.g., courses in the Humanities, courses that are electives for the students enrolled—

since they face the same problems and conditions in teaching, and may bring comparable experience to these problems.)[24] A few reports not only supply numbers of responses or arithmetic means but also provide correlations among different items so that an instructor can see at a glance a pattern of student reponses and the broad areas of teaching performance in which students think he does particularly well or particularly ill. One new technique for reporting results is the use of a graph or grid to display how an instructor is rated by student responses on a combination of items, such as the gains in knowledge together with the changes in attitude toward the subject that students feel as a result of the course. (An average of students' responses to items dealing with gains in knowledge may be plotted on one axis of a graph; an average of their responses to items dealing with changes in attitude might be located along the other axis.) If the student has been asked on the rating form to indicate what or how much he feels he has gained from the course, the report may correlate the intensity of students' conviction that they have gained much or little with average responses on individual elements of the teacher's work, so as to demonstrate to the teacher some possible reasons why the students thought they benefitted or did not benefit from the course. Indeed the statistical legerdemain performed with data from student rating instruments is often dazzling and sometimes bewildering; often a teacher needs some background in statistical analysis or educational research in order to understand fully the implications of the data presented to him in reports of student ratings. When free responses by raters are invited on the instrument in addition to multiple-choice answers on particular questions, the office compiling the responses may attempt a summary tabulation of the free responses, or may simply hand the completed questionnaires back to the instructor (after he has submitted his grades for the course to the registrar) so that the instructor can observe the free comments for himself.

As has been suggested, reports of student ratings can serve the needs of many kinds of readers. If reports of student ratings are given to the department chairman or the dean, they can be combined with reports of other courses taught by the same teacher and made to yield some generalizations about students' overall acceptance of that teacher, or, perhaps more usefully, can be made to indicate in what kinds of courses that teacher is more successful (in students' judgment) and in what kinds of courses less successful. The latter information, of course, can be of great value to administrators when they decide what kinds of courses a particular instructor will be assigned to teach. If reports concerning many or all of the teachers in a department are put together, the chairman can make some estimate of the subjects that his staff can present well to students, and he can infer what improvements are needed to make his department a more effective teaching unit as well as the areas of specialization in which more good teachers are desirable.[25] Displays of individuals' and departments' aggregate scores on student rating instruments can furnish evidence to alumni or legislators or citizens who may be in doubt about how well a particular teacher, department, or school is carrying out its responsibilities.

But evidently in almost half of the schools where students' ratings of faculty are systematically gathered or a form developed (or supported) by the administration, the ratings are for the private use of the teacher, unless he chooses to show them to his chairman. In these schools, students turn the rating forms back to the teacher directly, or they turn the forms over to a central agency (such as a university testing bureau) to be retained—and perhaps tabulated—and returned to the instructor after he has submitted his final grades. And some chairmen insist that student ratings, even when they come to the chairman, have negligible bearing on faculty salaries and no bearing whatever on decisions about promotion and tenure, although in some of these schools results of ratings in courses taught by graduate assistants and junior instructors are used by the course chairman or department chairman in advising the teacher and making decisions about his reappointment. In the other half (or more) of the schools using administration-sponsored rating plans, the ratings go to administrators (the department chairman, the dean, the personnel committee—sometimes two or more of these recipients) for use in activities ranging from counseling of the teacher to determinations about salary, rank, and tenure. But, as we shall stress later, rating forms developed and administered by student groups, even when they result in a published critique of courses (what one chairman called a "slam book"), have in the majority of schools no effect whatever on the faculty's appraisal of teachers or on administrative decisions about them. The converse also appears true: in relatively few schools are the results obtained from administration-sponsored rating programs made available to students to help them in the selection of courses and teachers.

Certainly as important as these informational—and occasionally rhetorical—uses of student ratings is the guidance teachers can draw from these ratings in changing their performance in their courses. Even a short rating form—one containing only ten or

twelve items—may offer a teacher (and maybe his chairman) indications of broad areas in which the teacher's work might be improved, such as motivating students to do their best work, or clarifying difficult concepts. But such forms usually leave to the imagination and initiative of the teacher the task of discovering exactly what changes he needs to make in his classroom techniques, attitudes toward students, manner of speaking, substantive emphases, and other elements of his performance, so as to alter the impact of his work on his students. On the other hand, some longer rating instruments, containing questions about details of an instructor's performance, may give immediate clues about steps he might take to improve students' response. For example, the instructor whose students report that he rarely provides specific examples to illustrate general propositions, that he often does not answer students' questions directly, and that he repeats the same point unnecessarily may plausibly guess that these features of his work, if students have observed them accurately, may be responsible for his being rated low on ability to make his subject seem clear and interesting to his students. Sometimes a perceptive interpretation of students' responses to a man's teaching can be proposed by the director of the university testing office, as has been suggested, through the application of sophisticated techniques for statistical analysis; this interpretation may help the instructor to see connections among ratings that are not obvious to an untrained observer. If an instructor is willing to take seriously the collected responses of his students, he may be able to discover broad areas or even specific bits of behavior on which he needs to work. Thus rating forms, as their proponents have argued, can indeed be both an incentive to the improvement of teaching and a source of guidance toward desirable improvements. From the rating forms an instructor can also gather, if he will study the ratings, potentially valuable insights into students' preferences and values that may help him when he teaches the course another time.

Despite the evident usefulness of rating instruments, and despite the growing number of schools that use such instruments, there is still among many faculty members (and some administrators) substantial objection to their use. Such resistance arises in part from disbelief in the findings of educational psychologists and researchers about the validity of student ratings and the relative freedom of the ratings from damaging bias. Indeed, anyone disposed to argue with the general conclusions of researchers that student ratings are on the whole fair can make a strong case for the view that biases do creep into such ratings, as N. L. Gage does when he brings together research findings that permit such a conclusion as this: "If a full professor is assigned a graduate, elective, off-campus course of intermediate size [30-39 students], . . . he is almost certain to get relatively high ratings from his students. If an instructor is assigned to teach an undergraduate, required, on-campus course of relatively large size, he will almost certainly get relatively low ratings from his students."[26] Even if one is rather positively disposed toward student ratings, it is one thing to quote a statistician's conclusion that student ratings of faculty are not affected by the difficulty of examinations, the amount of homework, or the grading curve in a course, and it is quite something else to convince Professor X that those twenty students who have been sitting in front of him for forty class periods will mark a rating instrument without expressing their private hangups about the university, the subject matter of his course, and features of his personality that he thinks he cannot alter or does not want to alter. Laura Kent observes that rating forms "depend upon the judgments of human beings who are necessarily subjective in their judgments,"[27] and quotes H. H. Remmers' comment about all rating instruments. After pointing out that the evaluator is not the instrument itself but the human being who completes it, Remmers observes that "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 . . ."[28]

Still other objections voiced by faculty members to student-rating instruments are to the generality (which some teachers translate as "ambiguity") of the language used in many items on some rating instruments, to perceived hostility of tone in the phrasing of some items on some instruments, and to genuine obscurities or excessively sophisticated nuances of meaning in the wording of some items. A teacher may be pardoned, perhaps, for hesitating to encourage students to express their judgments of him by responding to items that are, in his opinion, misleadingly phrased—especially if a summary of their responses is going to be handed to an administrator and used as a basis for a decision about the teacher's salary and tenure.

A much more significant source of resistance to student ratings may be the teacher's conviction that the individual items on a rating instrument, however clearly and crisply they may be worded, do not provide a pertinent comment on teachers' work and are thus inapplicable and misleading as sources of data with which to judge their teaching. These

objections have some support among commentators on evaluation. Hope Daugherty, for example, quotes an article written in 1952 by Max Marshall, arguing that "since it is to the student's future that the teacher is contributing, not his present, 'the only valuable rating of teacher by student would be after the student has been out of college for several years' "[29] Cohen and Brawer argue that since we do not yet know enough about "the relationship of teacher behavior to student learning," we cannot be sure that student ratings that focus on teachers' performance furnish truly useful data for measuring teachers' work by the ultimate criterion—how much students learn.[30] And the Muscatine report on education at Berkeley cites the views of a Special Committee on the Recognition of Distinction in Teaching that, while advocating the use of student ratings, recognizes that student appraisals "may reflect other aspects of teacher performance than those most central to the basic educational effectiveness of the teacher."[31] Even Dean Gustad has similar reservations about rating forms, as he has about many techniques for evaluation that are currently in use. "The evaluation of teaching," he says, "consists in finding out what contribution the teacher makes to what the student learns. . . . No one doubts that some instructors are more effective than others in assisting students to learn, but we have little more than a body of folklore to guide us in helping instructors do their jobs better and to help us decide how to allocate rewards equitably."[32]

Objections to the irrelevance of rating items are likely to become more frequent as the items themselves become more specific. An item referring to the teacher's use of audiovisual aids will seem thoroughly inappropriate on a rating instrument given to students whose instructor has considered and, with reason, dismissed the possibility of using audiovisual aids in his classes. A more subtle kind of misdirection might be alleged against an item about the constructiveness and specificity of the teacher's comments on students' written work, if the teacher has a policy of never writing on students' papers any comments other than those that will give encouragement or call attention to what was done well. Questions or items referring to the neatness of an instructor's dress, to the pace at which he speaks, to the punctuality with which he dismisses classes—to name only a few of the more vulnerable items found on some questionnaires—provoke spirited objections from teachers who believe that these elements of behavior have nothing whatever to do with their overall effectiveness as teachers. (Objections of this sort appear to be sustained by the results of at least one recent study of items included in a rating instrument.)[33]

Items referring to specific instructional procedures, therefore, are in the judgment of many faculty members "coercive," in that they virtually require the student to record a negative judgment regarding an element of behavior that the teacher may consider to be inapplicable in his teaching. Even worse, the presence of such items on a rating instrument may tend to suggest to students that the behavior named in the item is important—a part of good teaching—and may thus cause students to believe that they have somehow suffered because the instructor did not exhibit that kind of behavior—a belief that students would not have come to, or even thought about, on their own. These arguments, obviously, can carry a good deal of weight; clearly few items on any rating form will apply equally well to all kinds of courses and all kinds of teaching. These observations may argue against the use of detailed questionnaires as a means of giving teachers precise feedback on their work; they may argue instead for a shorter instrument containing only general questions likely to apply to any course or teacher.

But, since many rating instruments appear to describe, either through questions or through the affirmative and negative statements to which they invite response, a pattern of preferred teaching behavior, many faculty members resist the use of *any* rating instruments. No single pattern of teaching behavior, they argue, is demonstrably superior to other patterns, and virtually to impose the "preferred" pattern upon them by implying to students that it *is* preferable is a violation of academic freedom—a repudiation of the sacred axiom that it is the teacher's professional responsibility to teach in the style he finds most congenial and most effective.

Thus teachers and administrators who are anxious to gather data that will enable them to judge the quality of teaching in their schools confront the arguments of opponents that to describe an activity in words, so that it can be discussed, is to limit and perhaps to impoverish the activity, and that to devise verbal instruments for gathering data about teaching from the learners is to imply that teaching can be talked about precisely when in fact (so the argument goes) it cannot. Given these attitudes, it may be useful to offer a little advice here, based on responses to my inquiries, about the management of procedures for gathering student ratings. For if student ratings are going to be used as a source of information about teaching, it is almost mandatory that the hesitations of reticent faculty members be recognized and faced directly, *before* the rating instrument is completed and administered. The art of measuring the effectiveness of teachers by

student ratings may well be largely the art of getting a consensus about the form and use of the rating instrument.

If a college dean, a department chairman, a course chairman, or a group of concerned faculty members wish to establish student ratings as a method for gathering information about teaching, they will need to begin their planning well in advance of the time for the administration of the rating instruments. They will need to familiarize themselves with recent literature dealing with the use of such instruments, and they ought probably to gather a selection of the instruments in use at other schools to see how those schools have tackled the problem—even if they later reject as inapplicable in their school the rating instruments thus gathered. Since it is the students who will be making the ratings and since it is the students' education that is to be improved, students should be consulted early and regularly in the development of the instrument, and indeed ought perhaps to be on the committee that develops the instrument, especially if that instrument is intended for wide use and if its results are to be made available to help students in their choice of courses. A reasonable cross section of faculty members, particularly those known for successful teaching and recognized as committed to the improvement of teaching (or, if one prefers, to the *maintenance* of high quality teaching), should also be consulted, and some of the ablest, most respected faculty members should be drawn into the process of drafting the instrument. The form of the instrument and specific items should be carefully discussed and agreed upon among administrators, students, and faculty before a final decision is made to use the document with any large number of classes.

Indeed this process of developing the rating instrument, through wide discussion, can be one of the most useful steps in the whole process of securing student evaluations. To get useful responses, those who draft the rating form will have to ask useful questions; a useful question is one that directs attention to what is valuable and important in teaching. But values may be a matter of individual judgment; those who will use the ratings must know what they value, and what they take for granted, and why, when they talk about "good teaching." This self-scrutiny can be as beneficial as the ratings themselves in stimulating a serious inquiry among faculty, students, and administration into what constitutes effective teaching.

But whether the items to be included on the rating form are drawn from instruments in use at schools or from published reports of research ffective teaching, or are written by faculty members and students especially for the school's own form, the items proposed for inclusion in the form should be examined with care by administrators, faculty, and students alike—including many who were not on the original committee that drew up the form. Ambiguities should be eliminated wherever possible, vagueness reduced to a minimum, and a tone of neutrality and objectivity sought. Care should be taken to avoid giving the impressions that the document is endorsing any particular teaching techniques. Items asking the students to assess their own learning, their attitudes toward the subject, changes in their opinions and values should be included, in recognition of the widely accepted proposition that the effective teacher is one who causes learning—one who effects changes in students. If possible, the form should be pilot tested, to determine from the way it works in a few classes whether individual items are understood by the respondents and whether the responses to any items are distorted by feelings other than the students' honest convictions about the teaching they have experienced in the course. The pilot testing might well be in classes on quite diverse subjects, considering the number of kinds of instruction that students will evaluate by means of the instrument (if the instrument is for campus-wide use). After the testing, items can be checked to see whether they are yielding the information desired, and the language of each item can again be scrutinized to assure that it is as precise as possible.[34]

After the instrument has been drafted, discussed, tested, and refined, its use should be explained carefully to all faculty members who will be expected or encouraged to employ it. In this explanation, particular attention needs to be given to the disposition intended for summaries of the students' ratings. If the summaries are to go to any administrators, this fact should be clearly understood by everyone concerned before the instrument is used. If the summaries are simply to be returned to the teacher for whatever use he wishes to make of them, this fact should be made clear, and the implied promise that the ratings are confidential should be strictly kept.

Most rating instruments are administered during one of the last classes in a course, and the results are made known to the instructor only after he has turned in his grades. Some instruments, on the other hand, are administered at registration in the semester (or quarter) following the course, or even by mail after the students have completed the course. In either case, care is customarily taken to preserve the anonymity of the student completing each form. (A few instruments have been administered to seniors near graduation time and to alumni

several years out of college, to determine their recollections of courses they took as undergraduates.[35] One school polls students nearing the doctorate about courses they took during their doctoral program. Administering questionnaires to graduating seniors, alumni, and prospective recipients of graduate degrees, of course, recognizes that the real value of a teacher's work may not become clear until long after the student has left that teacher's classroom, and encourages respondents to put in useful perspective the various parts of their experience on campus; but the time lag between instruction and feedback reduces the possibility that the ratings can contribute directly and immediately to the improvement of teaching or even to the making of critical decisions about personnel.) Some care may need to be taken to assure that raters do not feel under pressure when completing the instrument; the instructor probably should not be in the room while the instrument is being administered (if it is administered at the end of a course), and a student volunteer can collect the completed instruments and take them to the office where they will be tabulated or retained. But the mechanics for administering instruments to students and even for tabulating and analyzing the responses (if computer tabulation is planned) are far less important to the program's success, my investigation suggests, than securing complete cooperation among administration, students, and faculty in the writing and refining of the instrument. Teachers will be much more likely to consider the ratings seriously if they are convinced that the instrument has undergone careful scrutiny and has the support of respected colleagues as well as of administrators and students. The developers of a rating instrument must take special care that the instrument is not regarded as a device invented by administrators to snoop without justification on the performance of professional colleagues, or as a technique for carrying on a popularity contest.

Whatever the merits of student ratings as a means of assessing the effectiveness of instruction, they represent a judgment from one perspective only (albeit an essential perspective), and, unless considerable free response is encouraged from the raters or unless the rating instrument is almost unworkably long, such instruments may not gather some significant data needed to help a teacher improve his effectiveness. Student rating instruments can bring home to an instructor when his course is poorly organized (in his students' judgment) or when he is insensitive to students' questions or when his comments on students' papers are thought inadequate, but if the instructor is to work on his weaknesses he may need other perspectives on why his course seems disorganized or how his responses to students' questions fall short or how his diagnoses of students' difficulties in writing are inadequate. Moreover, a course can be orderly and, indeed, have many other virtues, yet still be inadequate as a treatment of its subject; an instructor's responses to students' papers can be fair, friendly, and clear, yet still lacking in the kind of perceptivity that results in comments from which the students can learn what they need to know. These are only two examples of limitations that teachers and administrators may sense in student ratings, even if they find these ratings reasonable and valid as far as they go. These limitations suggest that student rating instruments might best be employed in combination with other techniques, in order to assure a comprehensive, adequately useful assessment of the instruction in a course or in a whole department. One hesitates to suggest that multiple techniques for evaluation be employed in any school; a concern for evaluation should not distract teachers' attention from efforts to advance knowledge and improve teaching. Still, the desirability of looking at a teacher's performance from several perspectives is not diminished even if we recognize the usefulness of rating instruments as a means of gathering data about a teacher's performance. Nor, of course, should the desirability of other perspectives obscure the unique values of student ratings, particularly as private feedback to the teacher. Indeed, used privately, while a course is going one, rating forms can aid in the immediate improvement of instruction—while it can still help students who are doing the ratings. There is in my correspondence, unfortunately, little evidence of interest in student ratings as means of "formative evaluation." Such use is one of the major unexplored applications of student ratings.

My discussion up to now has focused on rating instruments devised and administered primarily by faculty and administrators, sometimes (though not always) working in collaboration with students. On an increasing number of campuses, of course, students themselves (through the campus newspaper, political groups, or student government) initiate and carry forward programs for the evaluating of courses, partly just to tell other students more about the courses than the catalogue does. The desire to compile and publish ratings of courses and faculty is an understandable expression of students' concern for the quality of their education and their desire to help their colleagues make the best possible use of their college years. Student publications reveal these concerns clearly. Laura Kent finds that, in general, the focus in students' published ratings is on the instructor's ability to

communicate ideas clearly to students and inspire students' interest in learning about his subject.[36] Some publications, such as those issued by students at Harvard, Berkeley, and Michigan, are well known and have been appearing for several years. Many of the publications are read eagerly by students and act as a powerful sort of "informal advising."

Many, perhaps most, of these student-initiated rating programs are conducted with the best of intentions, are carefully planned, and are managed with every effort to assure fair and objective treatment for course and teacher. Instructors are often invited to supply descriptions of their courses and statements of course objectives for inclusion in the published booklet alongside reports of the ratings. Some editors offer instructors the opportunity to comment on summaries of the ratings before the summaries are published—the comment to appear with the summary of the ratings.[37] Some editors do not print evaluations of courses in which fewer than a given proportion of students (say, 50 or 60 percent) failed to complete the rating forms.[38] Miss Kent notes that many publications, too, "are careful to point out their own weaknesses and to disclaim infallibility."[39]

Despite these efforts, important difficulties frequently beset student-managed rating programs. In the first place, such rating programs often arouse suspicion and distrust among faculty members being rated. Rightly or wrongly, faculty members often view such rating programs as popularity contests, or as efforts to retaliate for unpleasant classroom experiences, to dictate teaching procedures, or to spread informal gossip about the idiosyncrasies of faculty members. Since faculty often do not participate in the development of student-administered rating forms, they sometimes feel that these ratings are an imposition. These feelings may be unjustified and even deplorable, but they sharply reduce the value of student-administered rating forms as instruments for the improvement of teaching. To be sure, on some campuses teachers pay considerable attention to student ratings, and administrators may even use them in making judgments about salary, tenure, and promotion. But faculty response to student-planned rating programs appears from my correspondence to be distinctly less favorable than to programs operated by administration and faculty.

In the second place, student-run rating programs are often conducted without much assistance from members of the faculty who could help students avoid common difficulties in the construction of rating instruments and in the phrasing of individual items. A project for the evaluation of teachers and courses can still be largely the work of students even if educational psychologists and experts in testing, not to mention teachers recognized on campus as effective, are consulted in the designing of the instruments. Students' desire to run their own programs sometimes results, according to my correspondence, in the administration of instruments that are demonstrably unwise or badly prepared; as a result, the very idea of student ratings becomes intolerable to many faculty and the possibility that such ratings might later be used to benefit both students and faculty is sharply reduced. Many chairmen deplored student-initiated rating programs; their letters to me expressed rather loudly the wish that the rating instruments would go away (if they were still in use) or be ignored.

Third, students frequently confront serious dilemmas when they try to report the results of the ratings. The more militant and aggressive members of student organizations frequently want to editorialize—to speak plainly and even caustically about some of the teaching their fellow students have experienced; these students insist on writing interpretive commentary that will guide other students to enroll in or stay well away from particular courses. More moderate members, often including the leaders, of the student organizations often hold out for objective, circumspect, noninflammatory comments about teachers and courses even if they do agree to publish interpretive commentaries on the ratings. Frequently the result is a publication that satisfies neither the more aggressive nor the more moderate students, and that may conceal or de-emphasize characteristics of a course or of an instructor that students might legitimately wish to know about. Sometimes the positions of the aggressive and moderate students are compromised by publication of a volume containing only the tabulations of ratings awarded instructors on individual items. Such a volume is likely to be large and confusing, and it may well discourage intended readers from using it. Furthermore, the publication of raw tabulations of ratings deprives alert interpreters of the opportunity to highlight what the ratings show about the strong points of a course or an instructor. It is, of course, possible to steer a middle course between flippant causticity and uninformative neutrality in the writing of comments on instructors and courses—and thus to avoid publishing nothing more than a forbidding computer printout—but the steering of such a course, indeed the mere writing of commentaries, takes time and requires great tact and diplomacy among the leaders of student organizations. Even careful, dispassionate editorial comment, which can hardly avoid containing some subjective assessments of the

data, may draw sharp charges of bias from teachers evaluated. In the face of these difficulties, student-managed rating programs are often short-lived, and all efforts to gather students' opinions on their courses and their teachers, for any purpose, may face lasting resistance as a result.

Fourth and last, the student leaders who manage rating programs often are with their programs only a year or two, thus leaving the program with minimal continuity from year to year and preventing the use in later years of experiences and information gathered in the difficult first years of the program. A student is likely to be elected student body president or executive officer of an academic council only when he is a junior or senior; by the time he has presided over one administration of a rating instrument, his academic career is nearly over, and the program that he has initiated or carried on from a predecessor may face an uncertain fate in the hands of his elected successor. (Such, at least, has been the experience on some campuses.) The student leader who initiated the program leaves the campus after graduation and is unavailable for consultation with those who may try to continue the program. As a result, mistakes are repeated, and enthusiasm for an important service to students may wane. Student programs in many schools, therefore, are irregular and undependable; they often do not exhibit the continuity and continuous improvement over time that one usually finds in programs run collaboratively by administration, faculty, and students.

It is fair to say, therefore, that the art of using student rating instruments as a technique for assessing the effectiveness of teaching is on the whole (and with notable exceptions) more highly advanced in schools where faculty members, particularly those with training in testing and statistical analysis, are collaborating on the development and administration of rating forms and on the interpretation of results. At the very least, faculty members appear more willing to pay attention to forms devised by their colleagues than, at least now, they pay to instruments devised and administered largely by student groups. If student ratings are to realize their potential for helping in the improvement of teaching, they might better be handled by groups representing the entire university community than by groups composed exclusively of students.

With all their limitations and despite all the difficulties and arguments that confront their supporters, student ratings of teachers are, then, a source of information about teachers' work that is nowhere else available and yet undeniably pertinent to the professional assessment of those who, after [illegible] must count these students as their major professional clients. The rating instrument is a data-gathering device; the data are students' feelings and reactions. Users of the instruments need to be sure that the data they gather—the feelings and observations they record—are those that will help toward useful generalizations about teachers and teaching. And users need to take care that their interpretations of their results are flexible, not mechanistic. For if we do not know, really, how students learn and which of their judgments about a teacher truly indicate how much they have learned, we cannot be sure we are asking students the "right" questions about their teachers, and we have to judge cautiously their answers to the diversely probing questions we do ask them. Still, by asking students questions about what they have observed, questions that can give us more or less "factual" knowledge about the way a teacher works, in combination with questions or items that invite judgments and require students to evaluate their own progress, we may be able to get reliable and valuable information about the art of teaching and about teachers from those who, after all, have observed both more closely than anyone else—except, perhaps, the teacher himself.

By Informal Student Comment

Even where student rating instruments are not in use, students' feelings about their teachers, informally recorded, are likely to carry weight with those who must judge the effectiveness of teachers—and in some schools those feelings carry a great deal of weight. Fully a third of my correspondents said explicitly that their assessment of faculty members' effectiveness depends at least to some extent on feedback received informally from students. (I suspect that the number of schools in which students' feelings about faculty are an important source of information for evaluating teachers is greater than suggested in my correspondence; if a department does not get tabulations from student rating instruments and does not observe classes, it is likely to have little to go on in judging the quality of its teachers except expressions of student feelings and teachers' impressions of their colleagues gathered in meetings and during informal conversation. The latter kinds of data are discussed in Chapter vii, below.) Some chairmen, in fact, prefer to rely on spontaneous, informal statements by students about their teachers, believing that these statements are more honest and genuine, hence more reliable, than computerized tabulations of check marks on rating forms filled out by students as a matter of routine, whether or not they have strong positive or negative feelings about the teacher. These chairmen assume that volunteered expressions of personal response,

unsystematically accumulated though they be, are preferable to data gathered deliberately by means of so-called "scientific" techniques for managing surveys. Among some chairmen, the preference for informally and unsystematically gathered data amounts to a triumphant conviction that in relying upon such data they are defeating pressures to recognize and make "scientific" that which formal procedures, as these chairmen see them, cannot possibly get at.

Procedures for gathering student opinion other than by rating instruments are, in most schools, hardly procedures at all. The chairman receives whatever comments come to him from students who "just drop around to talk." If no students come, the chairman is usually not troubled; indeed, he usually takes the absence of student callers as testimony that his teachers are all doing a good job. Some chairmen rely on comments by students who say they are moved to come in and give an opinion of a teacher; some rely on hints they pick up while signing students' petitions to drop or add courses. Some chairmen rely on information gleaned by advisors during consultations with students about their programs; indeed, gathering impressions about the quality of teaching by listening to advisors' reports of student reaction to courses appears to be common where systematic techniques for gathering data are not employed. Some chairmen, too, man registration tables, and pick up whatever they can hear as students register. Some chairmen admit that information thus gathered is hearsay or "grapevine" talk—casual remarks about teachers, sometimes at third or fourth hand.

A small number of schools collect "informal" opinion much more systematically than the schools just discussed. One school reported the existence of a committee of students that meets regularly with members of the department to give students' views on courses and teachers. A few schools elicit student opinion by commissioning student members of departmental committees to talk to their classmates and report back what the classmates say. In at least one school, members of the English department's personnel committee evidently conduct interviews with selected students to get impressions of teachers' work. Another school, when decisions about tenure and promotion are pending, establishes a committee of students specifically for the purpose of gathering information about the classroom teaching of each person under consideration. A small number of schools poll graduating seniors, asking them to appraise informally the quality of teaching at the school. And a few schools poll alumni—or alumni, students, and faculty together—get their views on who were the good and less good teachers on the campus. One school even reports that awards for outstanding teaching are conferred on the basis of polls of recent graduates.

Attitudes toward these informally gathered opinions from students (and alumni) are as varied as attitudes toward rating forms. Some chairmen recognize that complaints from students are more frequent than expressions of praise; they simply note the complaint and try to pacify the complainer, without paying undue attention to the complaint unless the same complaint is heard a number of times from different students. Other chairmen believe that they can distinguish between expressions of personal pique and legitimate objections to a teacher's procedure, though no one described exactly how he made the distinction; these chairmen invite teachers who have evoked student criticism to talk with them about the problems reported by students. Some chairmen seem to pay considerable attention to repeated conplaints; others appear to act only if action cannot be avoided. Some chairmen take seriously, and act on, complaints from students they regard as bright, but pay less attention to complaints from other students. Still other chairmen, on receiving *any* complaint, evidently carry on thorough investigations to determine its justness. In some institutions, particularly public two-year colleges, union contract provisions regulate the kinds of student complaints that can be considered in determining salaries, tenure, and promotion. Some chairmen use expressions of student sentiment (duly recorded as received, and filed for future reference) quite deliberately during discussions of promotion and tenure; some chairmen pay little or no attention to students' feelings when considering such issues. Quite a few believe that hearsay and grapevine gossip only corroborate judgments about teachers that a personnel committee would make from other evidence. But few considered whether the sort of personal popularity that leads to favorable comments about a teacher is a valid index of his effectiveness in teaching; few wondered whether a severe or aloof personality might cause an otherwise capable teacher to appear mediocre if, as a result of that personality, few students were moved to praise him spontaneously in the chairman's hearing.

These diverse estimates of the value of informally gathered student opinions, together with the recognition (shared by many chairmen) that administrators hear from only a small minority of the students who have worked with a given instructor, hardly encourage complacency among teachers and students in schools where students' opinions are gathered unsystematically, to be weighted arbitrarily in a fashion unknown to any of the persons

being evaluated. However free and plainspeaking a group of students may be, however small, intimate, and informal a department may be, however judiciously and circumspectly an administrator may elicit feedback from students, it still remains reasonable to doubt whether the real strengths and weaknesses of a teacher's performance come to light if informal expressions of students' feelings are a chairman's principal source of information about a teacher. Although chairmen who rely on informal expressions of student feeling assume, and some argue, that what they see and hear is the tip of the iceberg, there is not much to indicate that having glimpsed the tip, these chairmen take precautions to find out the shape of that portion of the iceberg they cannot see. Even though administrators listen constantly to students, as some administrators take pride in doing, the teachers in departments that rely on informal opinion may fairly believe that those who will decide their future have few reliable data about the quality of their work—fewer data than they would have if rating instruments or other data-gathering techniques were used. And what of those about whom no comment is heard: does the assumption that their work, though adequate, is mediocre do justice to them or to the students who are considering whether to take their courses? Does the absence of spontaneous feedback from students (which may be for reasons suggested above) justify the implication that there is little these teachers can do to improve their teaching, or that they are entirely adequate already? Without data, no one can say.

If it is still true, as Dean Gustad and others have shown, that deans and presidents evaluate faculty members' teaching largely on the recommendations of chairmen, it seems fair to conclude that major assumptions about the reliability of informal student comments on teachers are determining the professional futures of many members of college faculties, and thus affecting the quality of education offered to many students. Even if the assumptions are correct at a given institution, these informal procedures are not necessarily furnishing administrators with the information that will enable them to help teachers change their ways so that students' discomfort can be reduced or eliminated. Still, a number of chairmen take rather emphatically the position that the small size of their staffs and the informal relationships among faculty and students in their departments make anything more systematic than the informal gathering of opinions unnecessary, even potentially damaging to morale. How one reconciles reliance on these convictions—convictions no doubt earnestly and durably held as [illegible] of chairmen's experience and confidence in their perceptivity—with nagging reports of instances that cast doubt on these convictions, i.e., instances where other techniques for evaluation yielded findings opposite to those suggested by informal comments, is not easy to say. The question is one of reliability. In the assessment of teaching, is it worthwhile to use the advice of men whose profession it is to gather data that they are fairly sure is reliable, or is reliability of data irrelevant when one is talking about a multifaceted art such as teaching? That may not be the only question. Another might be: can we afford to ignore the reliability of our data when the professional futures of our colleagues are in question, not to mention the quality of the education to be offered our students?

By Indirect Evidence of Student Acceptance

Students have other ways, of course, than direct reports to advisors and administrators for indicating their response to the work of an instructor, and a few of the chairmen who corresponded with me indicated a lively awareness of how students can make their views known. Whether a teacher's courses fill up quickly or are largely avoided constitutes in the eyes of at least some chairmen an important comment on the teacher's effectiveness. (At least one chairman indicated that he recognized the need to look cautiously at this sort of data; many factors other than the teacher's effectiveness influence enrollment in courses.) For other chairmen, not merely the number but the kinds of students enrolling in a teacher's courses are revealing. In a multisection course, the instructor whose sections fill up quickly may be reliably assumed, according to some chairmen, to have demonstrated his effectiveness with students. One chairman spoke of paying attention to the number of students who continue working in a given teacher's field of interest, after completing a course with that teacher. Still another chairman notes the number and kinds of graduate students who make themselves protégés of a particular faculty member. The number of students who drop an instructor's courses can indicate the extent of student disappointment with what is going on in the classroom, though of course the amount of work assigned and the difficulty of the subject can cause students to drop courses quite as rapidly as ineffective instruction.

The dangers of overinterpreting statistics on enrollment and withdrawal from courses should not be lost on anyone. Instructors who demand a great deal of work or who are teaching material not inherently appealing to students or who are teaching at uncomfortable hours may draw and hold rela-

tively few students even though their teaching is satisfactory or better.

Summary

This long discussion of how students' opinions may help in the evaluation of teaching leads to a few conclusions, statable briefly. As the persons who gain little or much as a result of their teachers' work, students are entitled to share in the evaluation of those teachers; evidence indicates that, on the whole, they make fair and reliable judgments. If their opinions are worth having, they should be gathered carefully and systematically, not sporadically, haphazardly, or indirectly through hearsay or enrollment figures. The problem is to get from students accurate information about the features of a course and of teaching that truly affect learning, rather than on elements that do not matter, and to get their judgments on how well these important features of a course were managed. But, since there is little or no agreement on what elements of courses and teachers' performance affect learning, it is hard to be sure that we are asking for students' reactions on the right subjects. We sometimes end up asking students simply whether the teacher somehow (we do not ask how) made them want to work and made them feel they were learning. Furthermore, since we want to compare teachers and courses with each other (not that we *should* want to do so, but evidently we do), we try to ask for students' opinions on the same features in many different courses, ignoring the obvious point that in different subjects and with different teachers, quite different elements will affect learning in different proportions. And since we cannot talk with each student in order to be sure that we understand fully what his opinions are when he verbalizes them, we have to find words that will elicit from all kinds of students comparable opinions. Hence the variety of rating instruments, all of which try to solve the problem of standardizing our inquiries about an activity that, for each course and teacher, is in important ways unique. We come up against an ultimate problem: whether it is possible to standardize our questions, given the limited resources of our language, and still recognize the uniqueness of each teacher's work--the uniqueness that gives it distinctive value. We have not gone very far toward solving that ultimate problem.

But, as was suggested at the beginning, most schools that do gather students' opinions thoughtfully and systematically do so for teachers in all ranks, if not for all teachers. In this respect the gathering of students' opinions is an evaluation technique superior to many others, since these other techniques, being difficult to employ, are used only on selected groups of teachers, and not always with those groups where the greatest improvement in teaching is needed.

[1]George Benston, "The Value of Good Teaching to the Ambitious University," unpublished paper incorporated in a proposal to the Esso Foundation for support of a study of the evaluation of teaching.

[2]So argued Ruth Churchill, College Examiner of Antioch College, in 1966 before the American Association of Junior Colleges. Her talk, so far as I know, remains unpublished.

[3]But at the University of Washington, the *Ad Hoc* Committee on Student Evaluation of Teaching, in a report to the Faculty Senate dated 30 April 1969, said that it "found a surprising lack of objective analyses that demonstrate that student evaluation improves teaching" (Minutes of the Faculty Senate of the Univ. of Washington, 30 April 1969, p. 16).

[4]"Student Ratings of Faculty," *AAUP Bulletin*, 55 (1969), 441.

[5]"Evaluation of Instruction," in *Evaluation in Higher Education*, ed. Paul Dressel and Associates (Boston: Houghton Mifflin, 1961), p. 353.

[6]"Student Ratings of Faculty," p. 440.

[7]"Report of the President's Committee for a Pilot Study of Student Evaluation at Princeton" (Princeton: Princeton Univ., 1968), p. 5.

[8]The Hazen Foundation, *The Importance of Teaching* (New Haven, Conn., n.d.), p. 63.

[9]The main pieces of research are reviewed by H. H. Remmers in "Rating Methods in Research on Teaching," in *Handbook of Research on Teaching*, ed. N.L. Gage (Chicago: Rand McNally, 1963), pp. 367-69; more briefly, by Laura Kent in "Student Evaluation of Teaching," in *Improving College Teaching*, ed. Calvin B.T. Lee (Washington, D.C.: American Council on Education, 1967), pp. 339-40; by Wilbert McKeachie in the article already cited in the *AAUP Bulletin*; and by Cohen and Brawer in *Measuring Faculty Performance* (Washington, D.C.: American Association of Junior Colleges, 1969), pp. 12-14.

[10]Studies by Lehmann, Remmers, and McKeachie, in the articles cited, deal with these expected sources of bias.

[11]J. Stewart and K. Malpass, "Estimates of Achievement and Ratings of Instructors," *Journal of Educational Research*, 59 (1966), 347-50.

[12]Cited by Lehmann in "Evaluation of Instruction," p. 352.

[18]Cited in McKeachie, "Student Ratings of Faculty," p. 440, and in Cohen and Brawer, *Measuring Faculty Performance*, p. 14.

[14]Cited in Cohen and Brawer, *Measuring Faculty Performance*, p. 14.

[15]See Thomas D. F. Langen, "Student Assessment of Teaching Effectiveness," *Improving College and University Teaching*, 14 (1966), 25.

[16]B. W. Tuckman and W. F. Oliver, "Effectiveness of Feedback to Teachers as a Function of Source," *Journal of Educational Psychology*, 59 (1968), 297-301.

[17]Cited in Lehmann, "Evaluation of Instruction," pp. 352-53.

[18]Professors Cohen and Brawer imply a dissenting view in *Measuring Faculty Performance*, pp. 13-14. So does Duane Anderson in "A Report to the Committee on Instruction" (Washington, D.C.: American Association of Junior Colleges, 1970), when he says that "lack of competency, lack of maturity, and inability to see the overall function of the instructor pose real limitations on the use of this group of evaluators" (p. 6).

[19]Some rating forms are shorter than others, and no doubt many reflect a careful decision by their drafters concerning the elements in a teacher's behavior that will discriminate best between effective and ineffective teaching. Research findings have not yet helped us much, however, to decide which items are more useful and which less useful in the identification of good teaching. I have turned up no research at all that offers such assistance specifically for the evaluation of teaching in English. Dr. Langen found that the quality of lectures accounted for most of the differences perceived by students in the effectiveness of teaching, as measured by the rating forms at the University of Washington; reviews of class work (on papers and examinations), syllabi, and orientation talks at the start of the course made no significant contributions to the ratings of effectiveness ("Student Assessment of Teaching Effectiveness," p. 25).

Two recent studies of the value of different items in identifying those whom the raters really consider to be good teachers are Richard R. Perry, "Institutional Research," *College and University Business*, 47, No. 4 (Oct 1969), 18, 22; and Gerald R. Meredith, "Dimensions of Faculty-Course Evaluation," *Journal of Psychology*, 73 (1969), 27-32. Mr. Perry's top criteria, in order of their "value" in identifying teachers considered effective by raters, are: "Being prepared for class," "Establishing sincere interest in the subject being taught," "Demonstrating comprehensive knowledge of his subject," "Using teaching methods which enable students to achieve objectives of the course," and "Constructing tests which search for understanding on the part of students rather than rote memory ability." Dr. Meredith finds, unsurprisingly, that items relating to the value of the course and to the instructor's stimulating of thought and enthusiasm among students were prominent indicators of what students consider effective as distinct from ineffective instruction, while the difficulty of the material, the pace of the course, and the kinds of examination questions were less important in characterizing effective or ineffective instruction. But this is information at a high level of generality; it helps us little to identify elements of behavior that, in particular contexts, are associated with good teaching—if such elements can ever be found.

[20]This practice, however, is advocated by some recent writers on evaluation of teaching. See, for example, John O. Hunter, "Faculty Evaluation as a Liberal Persuasion," *Improving College and University Teaching*, 17 (1969), 90.

[21]See also McKeachie, "Student Ratings of Faculty," p. 440.

[22]For a full discussion of "forced choice" techniques, see Remmers, "Rating Methods in Research on Teaching," pp. 340-43. Also see George D. Lovell and Charles F. Haner, "Forced Choice Applied to College Faculty Rating," *Educational and Psychological Measurement*, 15 (1955), 302-03.

[23]Duane Anderson, "A Report to the Commission on Instruction," p. 3.

[24]Cohen and Brawer, *Measuring Faculty Performance*, p. 14; see also Gage, "The Appraisal of College Teaching: An Analysis of Ends and Means," *Journal of Higher Education*, 32 (1961), 18.

[25]Some administrators express doubt about the usefulness (for them) of reports about student ratings. One of my correspondents said that such reports help in identifying the very good teachers and the poor ones, but help little in differentiating among the large number of instructors "in the middle." One administrator said of statistical tabulations that they don't tell *how* the instructor is good, but only give relatively meaningless numerical rankings.

[26]"The Appraisal of College Teaching," p. 18. Gage also tells (p. 20) of one school where faculty have sought the benefits of student ratings without some of their disadvantages by asking colleagues to go into the class, interview the students, and then report fully and fairly to the teacher what the students (who remain anonymous) said of the class and the teacher.

[27]"Student Evaluation of Teaching," p. 339.

[28]"Rating Methods in Research on Teaching," p. 329.

[29]"Appraising the College Teacher," *Improving College and University Teaching*, 16 (1968), 205.

[30]Measuring Faculty Performance, p. 12.

[31]Select Committee on Education, *Education at Berkeley* (Berkeley: Univ. of California Press, 1968), p. 57.

[32]John W. Gustad, "Evaluation of Teaching Performance: Issues and Possibilities," in *Improving College Teaching*, ed. Calvin B.T. Lee (Washington, D.C.: American Council on Education, 1967), p. 280.

[33]Donald P. Hoyt, *Improving Instruction through Student Feedback* (Manhattan: Office of Educational Research, Kansas State Univ., 1969), pp. 14-15.

[34]The process of drafting a rating form is, obviously, expensive, and schools should probably hesitate to draft one unless they are prepared to spend the money to do the job well. At the spring 1970 conference on evaluation of teaching, sponsored by the Project to Improve College Teaching, first-year costs of $4,000 to develop a form for use at a school with 100 faculty and 2500 students were estimated. Princeton University is reported to have spent about $1.00 per student in developing its (rather elaborate) rating form.

[35]See Kent, "Student Evaluation of Teaching," p. 312.

[36]"Student Evaluation of Teaching," pp. 330-31.

[37]The Hazen Foundation, *The Importance of Teaching*, pp. 64-65.

[38]There is at least some evidence that even a relatively small number of returned rating forms will give a reliable assessment of a course's effectiveness. Donald Hoyt, at Kansas State University, reports that in his experiments with questions that asked students to assess their progress in a course, "Student progress ratings were made with acceptable reliability when there were 20-25 raters. Reliability of the overall progress measure was satisfactory when only 10 raters were used" (Donald Hoyt, *Instructional Effectiveness: Measurement of Effectiveness*, Kansas State Univ. Research Report No. 6, Manhattan: Kansas State Univ., 1969, p. 16). In recent faculty-run experiments with student evaluation of teaching at a private eastern university, a 20% response was considered the minimum needed before the results of student ratings would be published for use by students.

[39]"Student Evaluation of Teaching," p. 332.

CHAPTER IV
THE OBSERVATION OF CLASSES

Less popular as a method for evaluating teaching than the taking of student opinions, but still common—and sometimes offered to teachers as an alternative to the use of student ratings—is the observation of live or videotaped classes. In roughly one-third of the schools I heard from, the observation of classes is a major ingredient in evaluation procedures, and in some schools it is the technique relied on most heavily to furnish information about teachers' work. In almost all of these schools, classes are observed live; very few offer teachers the opportunity to have their classes videotaped for subsequent replay in front of the teacher and the observer, despite the popularity of this evaluation technique in some teacher-training programs.[1]

Some crucial differences between the use of student ratings and the observation of classes are immediately evident. First, where student ratings are employed, they are normally used with nearly all teachers in a school, and, if published, let the entire community know how each teacher's students are responding to him. With a few significant exceptions, however, only classes taught by graduate students, instructors, and—in a very few cases—nontenured assistant professors are observed; senior and tenured faculty members in four-year colleges are rarely evaluated by observation. Second, observation is mainly a step in a training process rather than a technique for evaluation. Few regular and tenured faculty are willing to withstand direct observation by their peers if the principal purpose is to determine how well they are performing their professional duties, and, of course, these people are scarcely prepared to admit that they need assistance in the improvement of their teaching or that colleagues might be able to offer such assistance. The principal exceptions appear to be service-connected schools, where observation of classes is a principal source of data for the annual efficiency reports that must be submitted on all officers who are classroom teachers, and some two-year colleges, where annual ratings by a chairman or superior are mandatory as support for recommendations about salary increases and "continuing contract" (i.e., tenure). Finally, the results of classroom observations never get to students or others in the university community; in some institutions absolutely no formal record is even made of observations, which in these schools are part of a strictly private relationship between colleagues.

The major premise underlying observation of classes is that looking at a sample or segment of the time during which a teacher is most directly and visibly engaged in the practice of his profession gives the information really essential for judging his success. If one assumes that the core of a teacher's work occurs in those moments when he is interacting with a class, evaluation—and training—may concentrate on what happens in these moments. From samples of these moments, so the argument runs, generalizations can be formed about the characteristics of a teacher's overall ability, and indeed concerning the effects he is having on his students; these effects are visible to the observer without his waiting for reports of judgments made by students on a rating form. A further and crucial assumption is that if a teacher conducts a successful class before an observer, he is at least capable of doing so and probably does so when the observer is not present; a corresponding assumption is that if the class observed is unsuccessful, other classes taught by the same instructor may exhibit problems as well—problems the observer and the teacher together must try to eliminate. A teacher's performance, in other words, is regularly an index of his competence—or so the supporters of evaluation by classroom observation often appear to assume.

The opportunity to see exactly what kind of performance the teacher gives and what effect he has upon his students is but one of the benefits cited by those who evaluate by observing classes. These administrators also contend that the judgments arrived at by peers are more dependable than those made by students. The peer being a teacher himself, and having presumably had considerable experience in both teaching and the observation of teaching, knows what to look at during a class session; he is a trained observer, while the student is often an untrained onlooker. Furthermore, the peer is likely to know well the subject being taught; he can judge the depth and breadth of the teacher's knowledge of his subject; can assess the clarity and accuracy with which the teacher is presenting data; can sense whether the teacher is setting forth opinions fairly, reasonably, and responsibly; and can appraise the justness of the opinions that students are encouraged to form. In short, the peer can determine how well the subject is being presented. Furthermore the peer, especially if he is a trained observer of people's behavior and reactions to

events, can see what is happening to students with a clarity not often achieved by the students themselves—or so the proponents of classroom observation argue—and by giving immediate feedback to the teacher, along with informed, humane advice about how he might try to proceed differently in order to get different results, can help him quickly to improve his work. A truly professional judgment, rendered immediately (not several weeks later, after student ratings have been tabulated), is obtainable in no other way than by direct observation of the teaching act.

The main arguments against the observation of classes may be better known.[2] What the observer sees, opponents of observation argue, is inherently an atypical class; the observer's very presence alters the relationship between instructor and students incalculably—and thus impairs students' ability to learn from the class, without giving any real benefits in return. In order to impress the observer, students may behave quite differently from normal; or they may be made nervous and for *that* reason behave atypically. The sample of teacher performance afforded by one or two visits—or even three or four visits spread over several weeks or several semesters—is unreliable; it presupposes a similarity in the instructor's performance from class to class that cannot be taken for granted, particularly since the classes observed are not likely to be representative of the teacher's daily performance. Whoever the observer may be, he looks at the class from a perspective different from that of students, and it is the students' perspective that counts since they are the ones who must learn from the class. Observation is not of much help to the novice or the young teacher, opponents argue, precisely because he is likely to be nervous and therefore to behave atypically before the observer. The comments offered by the observer may be irrelevant or downright misleading. Some educational researchers even argue that we do not know enough about the kinds of teaching that cause learning so that we can gather reliable and objective data from observation that will help in the evaluation of a teacher's work, though these writers admit that systematic observation may help researchers in studying patterns of teacher behavior to determine which patterns are more effective and which patterns less effective. Two recent commentators on the observation of classes put it thus:

> Teacher effectiveness must ultimately be defined in terms of effects on pupils, in terms, more specifically, of changes in pupil behavior . . . it is widely believed that a trained supervisor or expert of some kind can assess the effectiveness of a teacher by watching him teach . . . most of the many studies relating one variable or another to teacher effectiveness have used some such judgment as a criterion of teacher effectiveness. . . . In most cases, though, the effects of teaching on pupils cannot be observed directly in normal classroom behavior, but must be assessed by other means. . . . It may . . . become possible to measure teacher effectiveness in process by direct observation of the teacher. This is not yet possible. Attempts to validate process criteria by correlating them with measured pupil growth have been, on the whole, unsuccessful.[3]

But the climactic argument against observation of classes is that the classroom is the teacher's castle; for an outsider to invade this privileged sanctuary for the purpose of making judgments on the teacher's practice of his profession is an intolerable constraint on academic freedom and a visible expression of distrust in one's colleagues that is quite out of place in a humane profession such as teaching—a profession in which the setting of standards of performance that can be enforced by visual inspection is simply impossible. Each teacher, so the argument seems to go, should be assumed to be carrying out his duties competently unless eloquent evidence to the contrary is brought forth, presumably by aggrieved students, and the teacher should be encouraged to rely on his colleagues' making that assumption. An actor may perform before other actors, a physician may practice in the presence of his peers, a lawyer may plead cases before a judge and in the presence of a rival attorney; but teachers resist the mere suggestion that they practice their profession (or an important part of it) occasionally in the living presence of peers and colleagues. For an associate or full professor, the thought of being judged by peers as he performs his duties is simply unacceptable, and for a teaching assistant or junior instructor the experience of being observed, even if his lowly status forces him to accept it, is often painful. Faculties refuse to adopt observation of colleagues as a method of evaluating teaching not, I find, usually because of doubts about the accuracy of judgments made by qualified colleagues after direct observation of a teacher, but simply to avoid the destruction of teachers' self-respect and the consequent erosion of morale that they are sure would accompany introduction of any large-scale program for observing classes.

Though these arguments are powerfully urged by many teachers and some administrators, quite a few administrators obviously find the arguments favoring observation more weighty than the objections, especially in "apprenticeship" programs for young teachers. What do administrators learn about a teacher's performance from observing classes that justifies overruling these vigorous objections? What

characteristics of a teacher's work admit generalizations based on observations of a handful of his classes? What events or conditions do they notice when they look at a class in action? Chairmen (and deans) asserted that no generalized list could enumerate the events or acts that observers noted, and insisted further that they would never employ routine or mechanistic procedures for recording the events they witnessed. In observations at his school, one of my correspondents wrote, "There is absolutely no attempt to count the number of students who participate, or to keep track of how many minutes the teacher talks, or other such banal and inconsequential matters. There is an attempt to put words around what the teacher is doing that works or doesn't work." His statement, of course, clearly assumes that an observer can judge what "works," even as the writer insists that quantitatively describable characteristics of teaching are in no way of interest to the observers. From the comments of many chairmen and from sample observation reports that they pulled from their files and sent me, however, it is possible to draw some inferences about the specific characteristics of teachers' performances in the classroom that observers assess. Many of these characteristics, it will at once be clear, are those which students are asked to think about in completing their rating forms. The problem arises of who is best situated to offer cogent answers to these questions; there is not much evidence in my correspondence that chairmen have compared very fully answers gleaned from the two sources. It will be useful to set down here what chairmen think they can learn about teaching from observing it, though part of what follows restates chairmen's views on what they value in teaching generally, as summarized in Chapter ii.

Most observers of classes, as chairmen tell it, note the kind of interaction that takes place between the instructor and the students. That is, they try to determine whether students are attentive to and interested in what is going on in the class, and to what extent students participate in discussions. Observers appear to want to see as many students as possible participating. Furthermore, they note how the instructor responds to comments and questions from his students. Does he answer students' questions directly and to the student's satisfaction? Does he use students' comments to keep the discussion moving? Does he keep students' comments directly to the subject matter at hand? Many observers look favorably upon an instructor who positively and warmly encourages student response and tries to build the class discussion around students' observations. Conversely, extensive talking at the students, the playing of a game in which the student is asked to guess what answer to a question is in the instructor's head, or the rapid asking of a large number of questions so that no student has time to answer any question satisfactorily—in short, the absence of a genuine "dialogue" between instructor and students and between student and student—is looked upon with disfavor by several administrators who wrote to me about classroom observation. At the same time, the teacher should not, say other observers, be too permissive; one observer expects the teacher to make it clear when the students' comments are unclear, imprecise, or irrelevant.

Many of these observers' preferences in classroom teaching procedures are just that, preferences; they leave out of account, for example, the possibility that the quiet student who takes little or no part in class discussions may be learning just as much as the student who talks. But there is now some research evidence to support the interest of observers in the use teachers make of students' comments in class. Writing of the relationship between the teaching process and the product (student learning), in their article on "Teacher Effectiveness" in the latest *Encyclopedia of Educational Research*, Ned Flanders and Anita Simon report 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."[4]

Besides observing how the instructor advances discussion, many observers also try to sense the instructor's implied attitude toward his students. The instructor who tries to assess what is going on in the minds of students in his classroom wins more praise than the instructor who does not concern himself with what students' responses imply about their thinking. Evidence of a sympathetic understanding of students, of respect for them, is also valued by these observers. So is evidence that the students understand the goals of the instructor and the value of what is happening during a class meeting. Many observers expect instructors to exhibit a sense of humor—not to be confused with a penchant for sarcasm—and many praise instructors in whose classes an atmosphere of "warmth and freedom" prevails.

Many observers of classes also look at the substance of what the instructor is presenting and the style in which he presents it. Some observers expect a high "level" of questions to be discussed in the classroom; they want the subject matter probed to a depth "appropriate" to the supposed purposes of the course, and they expect the instructor to ask challenging and demanding questions about the

subject and to take a "balanced" view of that subject. "Sufficiency" of content—vague as such a criterion may be—is among the expectations of several observers. Some said, too, that they expect the subject matter to be of "inherent interest"; if the subject is not made interesting, the teacher's performance is thought to be at fault. Some observers also demand that the subject matter (particularly in English) be "relevant to general education." And some, too, insist that the instructor "cover" his subject thoroughly.

Talking of the planning a teacher makes evident

Talking of the planning a teacher makes evident

during a class period, observers said they like to see instructors demonstrate thorough knowledge of their subject and careful preparation for class. Some observers expect the instructor to show that he has a clear conceptual framework to organize his materials. Several pay attention to whether the material is being developed in an orderly sequence, with clear focusing of attention on important points. Some observers said they look for clarity and memorableness of presentation (the instructor must, of course, be fully audible), and for the use of examples to illustrate points being made. For some observers, the "timing" of climactic emphases and the "tempo" of the class as a whole are important. And some observers expect the instructor to display "poise" and at the same time "enthusiasm" in dealing with the material. The instructor is expected by several observers to avoid digressiveness and irrelevance. Indeed, many observers insist that the instructor always be "in control" of the class, that is, able to lead the discussion—supposedly without appearing to do so—to assure that students see what is fundamentally important in the material. The assignments made by the instructor, another observer noted, should be clear and relevant to the purposes of the course. And several observers said they look for "appropriateness" in the teaching techniques employed, without saying how they know when the techniques used *are* appropriate.

Some observers spoke of their emphases in observing classes in terms of what is happening to students. One said that he looks simply for signs that learning is taking place (he did not say what these signs might be), and that he regards any signs that the students are having positive learning experiences as favorable to the instructor, while looking upon the absence of learning (he did not say how he would detect its absence) as unfavorable. Indications that the students are considering useful and important ideas apparently signal a "positive" learning experience, though how one identifies for sure an important or useful idea is not easily explained. Other observers take an interest in whether the students appear to be thinking seriously about the subjects under discussion, assuming that the students who are thinking seriously about the subject can always be identified by an observer. For at least one other observer, an instructor's felixibility in using various classroom techniques and his willingness to experiment, as well as his ability to perceive what is going well and to abandon methods that aren't working, count heavily in the instructor's favor.[5] Flexibility in the use of teaching techniques, of course, is something an observer may be able to see; he no doubt finds it much harder to see, and judge, what is happening to, or inside of, the students in the class.

To illustrate how one course chairman directs attention of those who observe classes to important elements in a teacher's performance, I give the following extract from his instructions to the observers:

> I am inclined to consider the following questions or suggestions as guidelines: (1) Does the assistant, if he is leading a discussion, "spread" his conversation? That is, is he simply reciting a monologue, conducting a dialogue, or something more? (2) Does he treat his students with respect? (3) Are his questions challenging, clear, organized so that they lead ultimately to some point of illumination? Are they related to writing—student's or essayist's? (4) Does he know how to make use of so-called wrong or "dumb" remarks and answers? A generally dull bunch of students? (5) If he makes mistakes, is he willing to correct them? (The willingness to correct is, naturally, more important than the mistake itself.) (6) Does his class have a distinct tempo or rhythm to it? (7) Does he make a "proper" application of his scholarship? That is, is he merely reciting facts ostentatiously? Or is he relating the facts to the subject at hand, and in such a manner that shows he accurately gauges the level of intelligence in his audience?

Appendix B, a further illustration, is a document issued by a major liberal arts college to those responsible for classroom observations; it describes rather fully the procedures followed there.

In schools where observation of classes is practiced, the procedures used in arranging and conducting the observations are (as Appendix B helps illustrate) of great concern to administrators. In many schools, of course, the observer makes his visit (or maybe two visits) at the invitation of the teacher, or at least by previous arrangement with him. But in some schools unannounced visits are part of the observation program; even course chairmen who are outspokenly anxious not to damage teachers' morale sometimes make visits unannounced, in order to spare any teacher his fearful anticipation of the visitor's presence at a given class. (One wonders if the fear is not increased

by the expectation of unannounced visits; the teacher must always fear that a visitor will appear in his class.) Some administrators give teachers their choice of whether the visits will be agreed upon in advance or be unannounced. In some schools, two or three visits to each teacher under observation are provided for; they are widely scattered in time (even over as long as a full academic year). In some of these schools, one observation may suffice if the observer's report indicates that the teacher's performance is satisfactory or better (though what comprises "satisfactory" performance is often unclear, or left to the preferences of the observer). Some schools, especially anxious for fairness in the judging of classroom performance, send two or more observers to the same classes. These observers may consult with each other before writing the reports of their visits, but more often they submit independent reports of what they saw, for the chairman and dean to compare. Observers at a few schools compare observations of the same teacher made at different times, and compare observations of the same students made in different classes, to help assure reliability to the overall judgments of teachers that are based on these observations.

Some chairmen are particularly concerned with the atmosphere in classes being observed and with the relationship of the teacher and the observer during class. Most visitors deliberately keep well out of class discussion, remaining an inconspicuous as possible, but in one school the observers participate in discussions in the teaching assistant's class, "using any possible means," says the chairman, "to avoid the impression that the teacher is under inspection." One chairman regards the observation of classes not as a way of getting data for evaluation but as a way of helping teachers to get to know each other. And in quite a few schools the observer of a young teacher's class has already invited the young teacher to visit the observer's own class, or will do so shortly after observing the young teacher—not simply to reduce the tensions associated with observation, but also to offer the young teacher an example of the observer's teaching to consider and compare with his own. Several chairmen speak of doing everything they can to reduce the atmosphere of threat that almost inheres in the visit of an observer. One way of reducing that threat to a teacher's self-confidence is for the observer to discuss in advance with the teacher what the latter proposes to do in the class observed, and why. Another way is for the observer to discuss the class with the teacher before he writes his report, giving the teacher an opportunity to assess the class for himself, to talk about his style of teaching, his preferences in teaching techniques, and so on. One chairman talks over with each teacher, before he files a report, the substance of what he plans to say, and agrees that nothing will be said in the report with which the teacher himself does not concur. It is clear from my correspondence that the actual management of classroom observation is a very sensitive subject for teachers and administrators, however widely observation is practiced.

Like the use made of reports of student ratings, the use made of reports written by those who have observed classes varies considerably. In some schools, particularly in junior and community colleges where observation of classes is frequent, formal reports on classes observed—sometimes based on a carefully planned series of observations—are discussed with the teacher, then filed with the department chairman or the dean of the college (sometimes with both). These reports of observations are then deliberately used in determining whether a teacher's performance is satisfactory, whether he should be retained on the staff, and whether he should receive increases in salary. In some schools, a report of the class observed is filed by the observer—sometimes the observer completes a printed form, sometimes he writes an informal letter—with the course chairman or the chairman of the department; if a copy is not given to the teacher observed, he can at least inspect the report and discuss it with the course chairman or the department chairman. Such reports are often made available to personnel committees considering retention, tenure, and promotion. In some schools, the course chairman gets a report of the observation(s), but makes the report available to no one but himself (and the teacher observed); he draws upon the reports when writing recommendations, but makes almost no other use of them. In some departments, an informal record is made in the teacher's file that an observation took place, but no comments on the observation go into the file to help administrators decide on the status of the teacher.

One of my correspondents, discussing his reluctance to use against a teacher any negative information in reports of classroom observations, told how he used these reports:

> [The teaching assistants] know that, because of the unscientific nature of the visitation system, I will never put into the permanent file anything negative which was discovered in [the] visits. Rather, any negative comments are simply used as a springboard for conversation between the experienced visitor and the apprentice teacher. These conversations have proved extremely fruitful.

Still another correspondent implies that his department feels the same reluctance to make observation of classes a basis of formal evaluation of teaching; after the observations (which are optional) "The

observer discusses with the teacher his impressions and judgments of the class. He makes a written report *only* if the teacher asks him to. The report is filed in the graduate teacher's confidential file, kept in the Department's main office, and is accessible to regular faculty members." "It is a question again of 'guidance' more than 'evaluation,'" says another member of the same department. "It is a question of community." These comments sum up many administrators' (and many teachers') reticence about observation of classes as an instrument of evaluation. These administrators seek from observation of classes the possible benefits of improved teaching, but do so with as little emphasis as possible on "evaluation." In several schools, indeed, no record whatever is kept of observations completed, judgments made, and advice given to the teacher observed: the observation is looked upon as an informal, freewheeling effort by members of the department to help younger colleagues improve their teaching, rather than as a way of supplying a document for the information of a chairman or a personnel committee. Arthur Eastman summed up this spirit in a recent talk to members of the Modern Language Association by saying that, in his observation of classes, he sought to be a helper, not a judge; he added that if the teachers in a department work cooperatively with each other to help each realize his full potential as a teacher, the evaluation of teaching may "take care of itself."

One technique for assuring that experienced colleagues see a young teacher in action without invading his privacy, opening up the possibility that the "observer" will see an atypical class, or forcing a written report of observations, is the use of team teaching. If older and younger members of a department are teamed in the teaching of a course, so that all members of the team meet the class together, each member can observe the other, the younger teachers learning from their more experienced colleagues, the older teachers observing their colleagues in action—and perhaps learning from *them*, too. Observation in these circumstances is more palatable to the younger teachers because it is not the primary purpose of the collaboration; the student's increased learning is the primary goal. Despite its advantages, team teaching seems not to be much practiced in the departments with which I corresponded. One school, the University of Iowa, now employs the procedure regularly in its well-known "English semester" (a twelve-hour course in English literature that is virtually the student's entire load for the semester) with excellent results. John C. Gerber, Chairman at Iowa, reports that teaming an assistant professor with a member of the senior staff "is a natural way of involving senior faculty in the teaching of their junior colleagues, and is the best way of discovering how effective the new man is. By comparison class visitation and student reports are pale inadequacies. . . . The faculty teach better than ordinarily, and the students are correspondingly better motivated."[6] It goes without saying that in this sort of team teaching the "observers" are not intrusive outsiders, but collaborators, and that they see the teachers they are observing not once or twice or three times, but daily throughout the semester. If judgments must be made, they have firmer support, and they can be arrived at from observations that are likely to yield more reliable data than isolated visits to the classes of nervous, possibly resentful young teachers.

But team teaching, as noted, is rare in higher education today. For the most part, the generalized, subjective responses that observers make to what they see when they observe a class, and the looseness with which some administrators (often deliberately) verbalize what they look at when they visit a class, confirm that the issues of how to record the results of observation, and what to record, and what use to make of what is recorded, are justifiably delicate. More important, however, is the fact that since we are not yet in a position to generalize dependably about what actions or events comprise or contribute to "good teaching," general conclusions about a teacher's effectiveness are hard to reach, especially from one or two visits; for the same reason, detailed reports of what a teacher actually does in class (his actions, his way of proceeding) are hard to interpret. "No fallacy," say Donald Medley and Harold Mitzel, "is more widely believed than the one which says it is possible to judge a teacher's skill by watching him teach."[7] Judgments on a teacher's performance in the classroom are inevitably based on selective, potentially fallible perceptions of events in that classroom, as filtered (very often) through the observer's preferences about teaching style. The result may or may not be a reliable indication of a teacher's success; the judgment may place a premium on teaching styles that conform to established practices (or to the observer's practices), and may penalize distinctive teaching or unconventional techniques deliberately adopted by the teacher because they suit his personality, his subject, or his perception of students. In reports of observations, the relation between the observer's preferences or values in teaching, and his judgments of the teacher observed, may often be forgotten, as may be the relation between the teacher's distinctive goals and the details of his classroom procedures. Finally, the danger remains that observation of classes may produce judgments that are, simply, wrong (as of

course may polls of students taken immediately at the end of a course). "My favorite teacher," says one chairman, "broke every good teacher rule I ever heard: he was dogmatic; he read in a low monotone from his own book; he wanted back on tests what he said in class. But in fact I think of him as the best teacher I ever had. Others hated him"—a comment on the dangers of evaluation by student ratings as much as on evaluation by classroom observation—"and I can think back on deadly dull classes I hated at the time which proved of enormous value later on (sometimes years after I took them)." These possibilities of bias and error, then, are significant dangers, capable perhaps of being overcome by a skilled observer who is conscious of how he is making his judgments, but important for every user of observation as a basis for evaluation of teaching to bear in mind.

Thus, the hesitation of many chairmen about giving heavy weight to observation reports seems quite proper. All chairmen might hesitate particularly to use observer "rating sheets"—forms on which the observer simply checks off the degree of satisfaction he experiences in looking at elements of the teacher's work that the compilers of the form evidently think are associated with effective teaching. There is plenty of testimony from professional observers of teaching that these superficial, hastily planned rating forms are of very doubtful value.[8] Indeed, substantial use by college administrators of formal observation reports as the exclusive means for evaluating teaching in English (perhaps in any field) has little support and may be unwarranted.

And yet, because the classroom is where much, though by no means all, of a teacher's work is done, we can hardly resist the temptation to believe that an experienced observer familiar with the subject and the teacher's goals ought to be able to make *some* useful suggestions, even judgments, about the teaching performance he is witnessing.[9] If an expert, or at least an experienced teacher, looking at a major part of a teacher's activities, cannot arrive at any judgment at all concerning the quality of that performance, who can? Warner Rice, formerly Chairman of English at Michigan, after looking at some of the arguments against observation, affirms the value of observation wisely conducted, suggesting that young teachers be allowed to observe videotapes or films of their classes before being observed in person by a colleague, and that programs of observation be introduced gradually, with both junior and senior teachers among the observers. The emphasis in observation, Professor Warner Rice proposes, should be "less upon ... tion than upon the quick development of ... the early correction of faults. a concern about good teaching." Observation of classes, especially in large universities, should be conducted according to a well-planned program including:

> (1) a staff of visitors trained for their task by conferences in which they agree upon aims and procedures; (2) an adequate briefing, for everyone involved, about the plan; (3) an acquaintance, established in . . . seminars or by private meetings, between the visitors and the person to be visited; (4) the opportunity for the candidate to present, before the visitor attends his class, an outline of what he intends to accomplish; (5) a sufficient number of visits so that the candidate may overcome his initial embarrassment and accept the presence of the visitor in his classroom as a normal occurrence; (6) visits to each teacher, over a period of months, by two or more visitors; (7) a carefully planned pattern for recording visitors' impressions and suggestions; (8) conferences for a full exchange of views between the visitors and those visited; (9) opportunities for the teacher visited to attend the classes of his visitors and of other staff members who have shown a special talent for teaching.[10]

Evaluation still emerges from Professor Rice's program, and he seems convinced of the reasonableness of evaluation through observation. A less confident view, but one that still recognizes the intuitive conviction that an experienced teacher ought to be able to offer some useful judgments about the teaching he sees, comes from Professor Neill Megaw, currently Chairman of English at the University of Texas and, at the time he wrote the comments that follow, Chairman of the AAUP's Committee C. Professor Megaw suggests that observation, if conducted at all, should be conducted by a carefully selected committee of a college's entire faculty, operating under a very explicit set of guidelines:

> To insure faculty confidence in its competence, impartiality, and tact, such a committee would need to be faculty-elected and large enough to provide repeated observations by a number of visitors. [For a faculty of 300, the equivalent of nine full-time positions might be needed.] To insure administrative confidence that committee members would not become self-seeking or try to usurp authority for tenure and promotion decisions, such a committee should be clearly separated from the administrative and faculty body responsible for those decisions: that is, the committee on instruction would submit appropriate findings on teaching effectiveness, and these reports would then take their place among other relevant considerations. Such a committee would also have open lines of communication with students. Finally, though it is unwise to be too specific without knowledge of the local situation, probably such a committee would produce recommendations other than those directly involving salaries; its most important suggestions might be for

changes of teaching assignments, leaves for refresher studies or for the development of new courses, early or delayed retirement, and changes in the curriculum or in conditions or methods of study at the institution.[11]

Professor Megaw's diffidence about the gains from evaluation of classes, and my own reservations about the value of this technique for assessing teachers, are based on uncertainty about whether an observer, however well-intentioned, can be trusted to make objective, reliable judgments. If, among teachers and administrators in English and among scholars in Education, the possibility of learning anything about teaching by looking at teachers is in doubt, what progress, if any, is being made toward learning more about that possibility? For some answers, it may be useful to mention here at least some of the work being done by scholars in Education in search of ways to arrive at dependable judgments, based on direct observation, concerning teachers and teaching techniques. For scholars in Education are developing techniques for recording data about the actions of teachers and about events in the classroom—techniques that teachers and administrators in English may want to know about, even if they do not wish to adopt them.

One technique is that developed by Professors Medley and Mitzel (whose convictions about the limited value of observation as a means for assessment of teaching have already been cited): the use of an instrument called OScAR (Observation Schedule and Record). The OScAR record is a card listing the most common kinds of behavior exhibited by a teacher or student in the classroom—a list developed by consolidating information from earlier observations of teachers—designed in such a form that an observer can note on it quickly each kind of behavior that is observed within a predetermined brief sample time period during the class. The rating instrument is dense and complicated, but if well used by a skilled (i.e., trained) observer it can yield a valuably detailed description of how a particular class has gone; such descriptions can later, perhaps, be connected with measures of teachers' attitudes, supervisory ratings, and possibly with data about students' achievements to suggest causal or coincidental relationships between the acts of teachers and learning by students—relationships that might otherwise have gone unobserved.[12]

A different approach to representing what occurs in the classroom has been taken by Ned Flanders, whose procedure, called "interaction analysis," involves noting every few seconds which of ten categories of acts or conditions has just occurred in a class under observation. The ten acts or conditions are: the teacher accepts the student's feelings; the teacher gives praise; the teacher accepts, clarifies, or makes use of a student's ideas; the teacher asks a question; the teacher lectures, giving facts or opinions; the teacher gives directions; the teacher gives criticism; the student responds to a statement or question of the teacher; the student initiates a comment of his own; there is silence or confusion in the classroom. Flanders hypothesized that there would be a connection between the kinds of events observed most frequently in a classroom and the kinds of achievements made by the students. His investigations, he says, tend to support his hypotheses, and he has offered a general description of teachers whose students make the most progress. But he has not yet, to my knowledge, sought to offer any definitive statement of what acts by what kinds of teachers promote different kinds of achievements in different kinds of students. He does suggest that knowledge of what teaching is and what teachers do is increasing ("we are," he says, "approaching a theory of teaching"), and he asserts that "non-teachers [parents, administrators, taxpayers] have a right to expect the teaching profession to take an active part in the development of successful methods for evaluating teacher effectiveness."[13]

Still another model for describing the acts of teachers is that of Milton Meux and B. Othanel Smith, who try to identify the "logical operations" performed by teachers and students; these men suggest possible connections between skill in performing (and guiding) these operations, on the one hand, and increased knowledge and improved critical thinking in students, on the other hand. They also suggest that the effectiveness of a teacher might be measured by the precision with which he carries out each logical operation. Logical operations, as Meux and Smith use the phrase, include defining, describing, designing, reporting, evaluating, classifying, comparing, contrasting, inferring, and so on.[14] If these scholars' efforts to identify the components of the verbal behavior of teachers and students have any merit, they may furnish a conceptual framework by which one can describe with some precision elements in the performance of a teacher. But I have not yet seen any indications that a researcher using this technique has achieved a significant breakthrough in the evaluation of teaching.

An investigator who is more directly concerned with evaluation rather than just description of teaching is J. C. Flanagan, whose "critical incident" technique, while broadly applicable to many different kinds of professional behavior, appears to be of some special use in assessing teachers. Flanagan's technique is to determine with great care, in

interviews with large numbers of people who have observed members of a profession in action, what acts performed by a member of that profession are most frequently and conspicuously associated with the *successful* practice of his profession, as perceived by these observers. Having gathered listings of critical behavior from many observers, Flanagan selects those acts or events that appear most useful in helping to discriminate between a representative successful performance and an unsuccessful performance. He then gives a written enumeration of those acts or events to the persons who will judge the aptitude of candidates for a profession or will determine the effectiveness of members of that profession. The evaluator determines how many features of behavior that appear on the list are exhibited by the subject, and how many are not, and then records his findings on a tally sheet.[15] If indeed it is possible ever to enumerate critical items in the behavior of a teacher dealing with a class, it may be possible to apply Flanagan's procedures to teaching, as well as to other professions where the elements of successful performance are more readily and regularly identifiable. But to date Flanagan's technique, though evidently of interest to commentators on teaching (one of my correspondents spoke of it as an important technique for evaluation), has not been extensively applied to the work of teachers (particularly teachers of the Humanities), possibly because it assumes that the critical behaviors of a teacher are much more categorically describable than they really are (in the Humanities, at least), and because it allows little scope for innovation or inventiveness in teaching.

Using Flanagan's "critical incident" techniques and other procedures of his own invention for describing human personality and temperament, David Ryans has carried on (and indeed continues to work on) extensive research concerning the important characteristics of temperament and behavior found in classroom teachers, though few of his descriptions specifically characterize teachers of English. Ryans has developed instruments for identifying and describing some typical patterns of behavior among teachers (e.g., "understanding, friendly," "responsible, businesslike, systematic," "alert, initiating, responsible, confident," etc.) and for correlating these characteristics with the kind of training received by teachers studied, with their attitudes toward students, with their attitudes toward school administrators, with their verbal ability, with their age and marital status, with the size of the community in which they teach, and so on. Ryans' studies look less toward describing the ideal or the most effective teacher than toward discovering interconnections between elements in a teacher's background and personal characteristics, on the one hand, and his probable behavior as a teacher, on the other hand, that may be significant for trainers of teachers and for those responsible for selecting and assigning teachers. Ryans' work establishes a foundation for efforts to connect characteristics of teachers with progress of students—important work still to be done.[16]

The most recent effort that I have seen to describe with precision the acts of a teacher in his classroom (in this case, specifically, a teacher of English) so that judgments can be offered about the teacher's behavior is by George Henry of the University of Delaware. Dr. Henry, who has worked with videotape recordings of classroom performances by experienced teachers (and applies his technique to tapes made of inexperienced teachers, too), identifies six elements in classroom activity that he believes can be described with considerable accuracy. These features include the number of questions asked by the teacher and the pace of his questioning; the "cognitive level" of the questions; the number of subtopics under the main topic that are dealt with in the class period; the number of student responses to comments or questions by the teacher and to comments or questions by other students, and the length of the students' responses; the reaction of the teacher to students' comments; and methods (i.e., panel discussions, committees, written compositions, and so on) employed by the teacher to encourage students to give "structure" to their ideas. Henry suggests that we may be able to look forward to the day when, on the basis of several closely successive observations of a teacher's classroom, we can speak with confidence about a teacher's "style" and can make judgments about that style by referring to counts and other appropriate measures of the six elements of classroom teaching Henry has studied.[17] But that day is not approaching fast; only one of my correspondents uses in his observation of classes techniques even faintly similar to those described by Henry.

The work of the scholars listed here by no means exhausts recent research on ways of talking about and judging a teacher's work by means of intensive observation of events in the classroom. None of these researchers (or any others, so far as I am aware) promises an imminent breakthrough in the quest for valid techniques of evaluating classroom teaching. The search for such dependable techniques is continuing. Yet despite the care and professional competence with which trained researchers are carrying forward their inquiries, none of the chairmen and few of the administrators who wrote me showed any awareness of these investigations or any interest at all in scholarly investigations of

classroom teaching. The suggestion that an observer of a class tabulate events in the classroom and determine the proportion of class time devoted to specific kinds of activities, indeed, elicited expressions of contempt, in varying degrees of intensity, from several of my correspondents (for instance, the chairman quoted on p. 39). Those schools in which observation of classes is practiced as a way of evaluating and improving instruction, in short, seem satisfied to conduct these observations with the aid only of general conceptions of what is important in the classroom and broad guidelines concerning what are and are not effective teaching procedures. One respondent, speaking against the suggestion that teachers be closely observed, but probably recording the sentiments even of those who do observe classes, said, "many of my colleagues would agree . . . that to go beyond the implicit limits of professional tact in seeking to judge our assistants as teachers—or for that matter our professional colleagues—could well be like judging the development of a tree by pulling it up repeatedly to look at its roots." The analogy may not be exact, but it eloquently implies the conception of teaching and the attitude toward evaluators of teaching that seem typical of many college administrators.

And, it is important to add, these attitudes have support among those who write professionally about the art of teaching. The potential usefulness of studies such as those by Ryans, Henry, Flanders, and others is called into question by several authorities, notably by H. S. Broudy, editor of *Educational Forum*. Though he names no specific studies or researchers, Dr. Broudy suggests that research on the qualities of effective teaching is proceeding up a "blind alley." We have no way of knowing, he says, what variables in a teacher's behavior are relevant to his effectiveness until we have a notion of good teaching—and we do not yet have that.[18]

Among college teachers, therefore, the observation of classes, as practiced today, depends for its usefulness on the wisdom and perceptivity of the observer and, ultimately, on the reasonableness of his preferences in teaching styles and his conception of what constitutes good teaching. Few college teachers or administrators exhibit any interest in changing this state of affairs; there is no evidence of any widespread effort to train the observers or to improve either their perceptivity or the reliability (in the technical sense of similarity between what two observers looking at the same class might see in that class) of their observations. Although, as has been suggested, one rejects only with great difficulty the conviction that observing what goes on the classroom—considerations of professional "tact" aside—is a dependable source of data for the evaluation of a teacher, the casual way in which observation is practiced in many schools today inspires little confidence in judgments about a teacher's effectiveness based wholly or largely on data thus gathered. The profession needs to know a great deal more about the kinds of behavior in teachers that it can legitimately praise and about the reasons for preferring these kinds of behavior—even though no one teaching style, of course, will necessarily be more effective in most circumstances than any other style—before it can place much confidence in the results of observation as a substantial basis for judging its members. In short, the "facts" of what has happened in a classroom are frequently disputed by observers of that classroom; the "value" of what happened, also, is often honestly in doubt.

This is not to say, of course, that observation is useless and should be abandoned. On the contrary, observation of classes by skilled, experienced observers (it would help if the observers were also good teachers) can give the teacher useful feedback about how the observer(s) reacted to his performance. Observation, too, can help administrators to identify characteristics of a teacher's activities in class—characteristics that may permit a more informed assignment of teachers to courses. Furthermore, the data gathered by an observer can be compared informally to student reponses on rating forms, to help the teacher think more perceptively about the possible effects of things he does on students' reception of his courses. But unless several observations are made by trained observers, the reports of classroom observations will not, by themselves, dependably inform a chairman or dean about the *value* of a teacher's work; even reports of observations by skilled observers ought to be compared with judgments of the teacher arrived at from different sources before being given weight in any final assessment of a teacher. Certainly no dean, trustee, or citizen who is demanding that a department be accountable for the quality of its teaching will find in the reports of classroom observations alone anything to fully satisfy him, and he can ill afford to base negative judgments on such data alone.

[18] Duane Anderson, in "A Report to the Committee on Instruction," (Washington, D.C.: National Association of Junior Colleges, 1970), recommends taping or filming of classroom activities as a way to get a durable, objective record of what goes on in a class, for later discussion and analysis. (p. 7)

[2]See, for instance, George Mills Harper, " 'The Waste, Sad Time': Some Remarks on Class Visitation," *College English*, 27 (1965), 119-23, and Professor Harper's rebuttal to my reply, in *CE*, 28 (1967), 620-21.

[3]Donald Medley and Harold Mitzel, "Measuring Classroom Behavior by Systematic Observation," in *Handbook of Research on Teaching*, ed. N. L. Gage (Chicago: Rand McNally, 1963), pp. 248-49.

[4]Robert Ebel, ed., *Encyclopedia of Educational Research* (New York: Macmillan, 1969), p. 1426.

[5]Attention to this characteristic of teaching is also supported by literature on educational research. Ned Flanders reports, "There is some evidence to suggest that in classrooms in which pupils have more positive attitudes and seem to be learning more content there is also greater flexibility and variation in teacher behavior." ("Teacher Effectiveness," in *Encyclopedia of Educational Research*, p. 1429).

[6]"Response to Austerity," *ADE Bulletin*, No. 25 (May 1970), p. 52.

[7]"Measuring Classroom Behavior by Systematic Observation," p. 257.

[8]See, for instance, H. H. Remmers, "Rating Methods in Research on Teaching," pp. 372-73, and Medley and Mitzel, "Measuring Classroom Behavior by Systematic Observation," pp. 248-49, 257-58 (both in *Handbook of Research on Teaching*). Neither essay discusses directly these informal rating sheets used by departments, but their caveats about rating instruments in general are instructive.

[9]The other time of direct face-to-face communication between student and teacher, of course, comes in the personal conference or interview. Nothing in my correspondence suggests that observation of interviews is an important part of the evaluation of teaching by faculty in any rank. Obviously the presence of a third party at a student's personal conference with a teacher could intimidate any student, and observation rooms equipped with one-way glass appear not to be well-known tools for data-gathering by college administrators. Still, where younger teachers carry on interviews in offices they share with older faculty, the older faculty are not always averse to making judgments about what happens in these interviews, and one chairman faintly lamented his department's forthcoming move to new offices because it would make such "unintentional" observation of personal interviews between students and younger teachers more difficult. Very few chairmen indicated, however, that they gain much information about their staff from colleagues' reports about teachers' interviews with students.

[10]"How the Candidate Learns to Teach English," in *The Education of Teachers of English for American Schools and Colleges*, ed. Alfred Grommon (New York: Appleton-Century-Crofts, 1963), pp. 581-82.

[11]"The Dynamics of Evaluation," in *Improving College Teaching*, ed. Calvin B. T. Lee (Washington, D.C.: American Council on Education, 1967), p. 285.

[12]"Measuring Classroom Behavior by Systematic Observation," pp. 278-86.

[13]"Some Relationships among Teacher Influence, Pupil Attitudes, and Advancement," in *Contemporary Research on Teacher Effectiveness*, ed. Bruce Biddle and William Ellena (New York: Holt, Rinehart, and Winston, 1964), pp. 219-21, 231.

[14]"Logical Dimensions of Teaching Behavior," in *Contemporary Research on Teacher Effectiveness*, pp. 127-64. Pp. 142-48 and 163-64 are especially importa· '.

[15]"The Critical Incident Technique," *Psychological Bulletin*, 51 (1954), 327-58.

[16]See Dr. Ryans' book-length study, *Characteristics of Teachers* (Washington, D.C.: American Council on Education, 1960), especially pp. 271, 380-98. A shorter report is Dr. Ryans' "Research on Teacher Behavior in the Context of the Teacher Characteristics Study," in *Contemporary Research on Teacher Effectiveness*, pp. 67-101.

[17]"Style of Teaching and Teacher Evaluation," *English Journal*, 59 (1970), 925-27.

[18]"Can We Define Good Teaching?" *Teachers College Record*, 70 (1969), p. 584.

CHAPTER V
THE INSPECTION OF TEACHING MATERIALS AND ANNOTATED STUDENT PAPERS

To observe a class is to examine the act of teaching—or at least a substantial part of it—as it is performed. But a teacher's behavior in his classroom is not the only visible procedure by which he teaches students; the syllabi that he prepares for his courses, the duplicated materials with which he supplements lectures and class discussions, and the responses that he makes to written work submitted by students are three additional elements of a person's teaching—elements no less important than classroom activity as potential demonstrations of a teacher's effectiveness, despite the fact that written materials cannot be seen interacting with living students. These important elements of a teacher's work offer many chairmen important evidence about the quality of an instructor's performance. About one administrator in seven with whom I corresponded makes the examination of a teacher's handouts and comments on papers a part of his technique for evaluating instruction.

Duplicated materials are best treated separately from comments on students' papers, the writing of which requires special kinds of teaching skills. Such materials usually include the following: course syllabi, outlines of units, lesson plans, individual assignments for writing, quizzes, examinations, instructional notes (written statements of opinions or propositions that substitute for or supplement lectures), extracts from printed sources, diagrams or outlines of conceptual models, course proposals and prospectuses, and so on. As with rating instruments and reports of classroom observations, procedures for handling documents of these kinds are by no means the same among the various institutions that examine them. Some administrators require teachers to submit these instructional documents or to file them in a central place; some administrators do not ask to see any documents, but willingly examine those that are given to them. Practices differ, too, about which members of the faculty are asked or invited to present teaching materials. In some schools, the syllabi, course outlines, and examinations given by virtually every member of the staff are examined; in other schools only the syllabi, assignments, and examinations of graduate assistants, instructors, and probationary assistant professors are inspected. There is, also, little uniformity in the amount of material that a faculty member is asked to turn in; many chairmen and deans speak of looking only at syllabi, others exhibit a lively interest in teachers' examinations, while others show an interest in both syllabi and examinations, and still others require the regular submission of lesson plans along with syllabi and examinations—particularly lesson plans for classes that have been observed by the administrator. Not too many chairmen say that they pay deliberate attention to course prospectuses and proposals submitted by faculty members, but some chairmen who wrote perceptive letters said that they paid attention also to this kind of "teaching" performance. Generally, chairmen at junior and community colleges seem much more likely to insist that an instructor's teaching materials be examined than do chairmen at state universities and liberal arts colleges.

It is hard to determine exactly what administrators look at or for when they review teaching materials and course proposals. For all one can gather from the comments of many chairmen, they simply try to determine the general level of intelligence and clarity with which an instructor is organizing his courses, making his points, and testing his students' understanding of the course. One correspondent said that he pays attention to the level of complexity and sophistication in a teacher's test questions. Some chairmen and directors of programs for teaching assistants are more explicit. They try to find out whether the individual teacher's plans and assignments suit the overall goals of the course as established by the department, and are reasonable in light of the syllabus for the course as adopted by a steering committee. Other chairmen and program directors say that they look, in examining teachers' assignments, for variety and originality as well as clarity. Still others look deliberately and principally for inventiveness and freshness in a teacher's materials; these chairmen do not speak of concern for the aptness of materials to the department's plans for the course. There seems, in short, to be a sharp difference of interest in teaching materials; they may tell whether the faculty member is doing what the department has decided that it expects of him, or they may tell how freshly and independently the faculty member is carrying out his responsibilities. Of course, both ends could be served simultaneously by the same teaching materials and documents, but none of my correspondents spoke of being interested simultaneously in both these qualities of a teacher's materials.

There is also some division of opinion concerning the value of dittoed documents as indicators of the

quality of a teacher's performance. Some administrators depend heavily on the assumption that a well-planned course with an interesting syllabus and "suitable" examination questions indicates the overall competence of the teacher; the documents are taken as signs representing the quality of the man's entire performance. Other chairmen stress that syllabi, lesson plans, and examinations may be little more than "window dressing" that masks subtle ineptness in the classroom, or may sell short the teacher whose strength is in his work in the classroom rather than in documents that support or surround the classroom encounter. One chairman distrusts syllabi and teaching materials, he says, because they show "potential, not performance"--though such documents *can* be regarded as a part of performance. But if substantial thought has been given to the precise value of teaching materials as sources of data for evaluation, this thought did not emerge visibly in my correspondence with chairmen and deans, nor does it emerge conspicuously in published research on problems of evaluation. The reason may be that teaching materials are so diverse as to discourage efforts to find questions or topics that will permit administrators to compare them informatively.

Examination of the teacher's response to student writing is an explicitly admitted part of the evaluation of teaching in about the same number of schools as practice the inspection of teaching materials; it is often carried on more formally and rigorously than examination of dittoed materials, particularly in those schools that have formal training programs for teaching assistants or programs for helping the work of new instructors. As can be imagined, most of the interest in teachers' annotations of student papers is focused on classes in written composition; teachers in these classes are regularly asked to supply a more experienced colleague with a set of papers that they have annotated; the colleague reviews the papers, looks into the justness of the grading, and makes suggestions about what the teacher said or did not say about the paper. For those who believe that the central part of the course in composing is the teacher's responses to his students' papers, the comments are important data for evaluation. Inspection of comments on student work seems not to be considered of comparable value in courses other than composition or in courses taught by senior staff. Annotated papers are much less frequently inspected outside of introductory courses in writing. And no evidence has been turned up, to my knowledge, suggesting any close connection between an observer's assessment of the comments on student papers in any course except one in composition, and the students' perception of the effectiveness of the teaching in that course.

Most of those who do regularly look at teachers' annotations of papers know much more clearly what they are looking at in comments than do many of the administrators who examine teaching materials. The values sought in teachers' comments are not surprising, but they bear enumeration nonetheless. Fullness of annotation is a virtue in the eyes of many observers; in the judgment of these observers, the strong and weak features of the paper should all be pointed out. A corollary requirement is accuracy of annotation: what the teacher says is good or bad should be visibly good or bad, and the analysis offered by the teacher should describe the strength or weakness precisely. A good comment, in the eyes of many observers, is "relevant" to the assignment (it may speak to how well the student has fulfilled the assignment) and takes into consideration the student's evident purpose in writing the paper. Some administrators expect teachers to make instructive comparisons between work handed in by the student in a previous assignment and work handed in on the current assignment. Comments should be clearly stated, say other observers, but should not be trivial and picayune (the dividing line between thoroughness of annotation and triviality presumably differs depending on who is looking at the comment); important matters in the paper should be emphasized and the tone should be constructive, without sarcasm. The focus of comments on any paper should be clear and emphatic, say several observers, who evidently disapprove of scattered, disconnected notes on many different features of the same paper. Most observers of comments insist that the annotation be forward-looking: it should tell the student precisely in what way he can improve his work and, in the words of one chairman, it should make him *want* to do better. Although the need for acute perceptivity in deciding what to comment on was not conspicuous in many of the replies I received, some chairmen did say that they hoped teachers' comments would show sensitivity and intelligence in dealing with written prose, and incisiveness in the selection of features of the paper to receive extended comment. Fairness of judgment—a combination of rigor in the application of standards and flexibility in the recognition of difficulties students face in achieving these standards—is expected of comment writers by many chairmen. A balance of positive and negative comment on every paper is expected by some chairmen. Succinctness and grace in comments are even sought by a few chairmen.

Many chairmen assert that is is not only what the instructor says about the paper that counts, but also what he does with it after he has received and

annotated it. These chairmen pay attention to the kinds of revision required of the student and to how thoroughly the instructor checks to see whether the revision is adequately done. Some chairmen expect dittoed examples of student writing to be used in the classroom.

Even though the emphases of chairmen in their inspection of comments differ from institution to institution, most chairmen who do look at comments agree on the importance of doing it. So convinced are some program directors about the importance of training teachers to write effective comments that they arrange meetings at which staff members can compare comments on the same paper, or panels of teachers can together prepare comments on the same paper, or a mentor and his apprentice can separately comment on the same paper and compare their notations. Teachers' comments at such meetings occasionally become bases for the evaluation of those teachers. The questions that remain unanswered, despite this unanimity of feeling, are these: exactly how do various techniques in writing comments, and kinds of insight into the strong and weak points of students' work, contribute to the total effectiveness of the teacher's work? Is the writing of quite a few incisive comments an infallible sign of a teacher's overall effectiveness, or can a teacher be effective even if he annotates papers lightly or offers nothing but 'praise for students' work in the hope of encouraging them to continue writing? And what, after all, is a "good" comment—one that is thorough, full, and balanced in its appraisal of the paper, or one that asks a single significant question about the paper, or one that simply asks for additional information on particular points made by the student, or one that specifies features of the paper to be changed in revision? The answers to these questions, of course, depend upon the teacher's purpose in comment writing and on his relationships with his students. An instructor's comments have to be judged in reference to his overall teaching strategy and his perception of how his students can be motivated or assisted. It seems manifestly unfair for an observer to criticize a teacher for not annotating papers as the observer would have annotated them, if the teacher's deliberate strategy in annotating papers is different from that of the observer. Even sarcasm and anger, revealed by a particular kind of instructor to some kinds of students, *can* spur the student to expend greater effort on his work rather than driving him away from both the course and the instructor (as we might assume they would). It may be administrators' recognition of these points that deters some schools from asking students to answer on rating instruments detailed questions about the kinds of comments written by the instructor and about the subjects treated in those comments.

In the assessment of teaching documents and of annotated papers, the subjectivity and personal preferences of the rater may be even more sturdy barriers to reliable and consistent evaluation than in the observation of classes. The dangers of subjectivity may be greater in the assessment of teaching materials, I think, because subjectivity there may go more easily unrecognized than in discussions of classroom performance, where differences in teaching styles are admittedly enormous. Teaching documents and comments on students' themes are successful or unsuccessful only in reference to the whole teaching enterprise; what they accomplish is not necessarily well judged if the observer rates them on how well they correspond to his own presuppositions about effective teaching. Often simply looking at these documents is not enough to permit evaluation of them (as some chairmen *appear* to believe it is); searching conversations with the teacher about his techniques are important as background for evaluation of materials. The question that needs to be asked in any effort at evaluation of syllabi, examinations, and comments on papers, is: to what extent do these documents support the specific goals and strategies of the teacher using them? Those goals and strategies are themselves open to evaluation, of course, but that evaluation is quite another act than evaluating the usefulness of the instruments in the attaining of the teacher's goal. Observers should take care that their examination of teaching materials and comments does not turn out to be an effort at requiring of the teacher the techniques and strategies preferred by the evaluator.

If such care is taken, however, the inspection of teaching materials can be a valuable way of learning about any teacher's techniques and strategies in practicing his profession. It can help administrators and colleagues in their efforts to assist a teacher's development, and it can *contribute* to an overall assessment of a teacher's strengths and weaknesses. But it can by no means offer the sole basis for evaluating a teacher, and can give misleading results if not carried on in the context of other features in a teacher's total performance.

CHAPTER VI
ASSESSING THE RESULTS OF INSTRUCTION

The "ultimate criterion" of a teacher's effectiveness is the learning achieved by his students, or so Cohen and Brawer observe in their monograph on the evaluating of teachers in junior colleges.[1] But this ultimate test of effectiveness has proved extraordinarily difficult to apply. Professor N. L. Gage of Stanford University, summarizing in 1963 the efforts of various researchers to connect, in broad terms, acts performed by teachers with gains achieved by students, concludes that such efforts have not been productive because they oversimplify drastically the relationships between teacher and student, and misrepresent the complex forces that are working in an educational situation to bring about whatever finally happens to the student. "Research . . . has been abundant; hundreds of studies, yielding thousands of correlation coefficients have been made. In the large, these studies have yielded disappointing results: correlations that are nonsignificant, inconsistent from one study to the next, and usually lacking in psychological and educational meaning."[2] The alternative to an excessively simplistic view of the relationship between a teacher's acts and a student's learning has been to develop rather complex models of the process of interaction between teacher and student (Professor Gage's chapter illustrates a few of these complex models of the teaching process) and to devise research activities based on these models with the hope that, even if sweeping conclusions cannot quickly be reached, at least some bits of knowledge about the connections between a teacher's behavior and his students' progress may possibly come to light. But even these more complicated models have not been especially productive; Professor Gage reported (in 1963) that "research workers have looked away from criteria involved with effectiveness, defined as teachers' effects on pupils' achievement of educational objectives."[3] In higher education, if a review of the most recent volumes of major periodicals such as the *Journal of Higher Education* and *Improving College and University Teaching*, together with my letters from college administrators, furnish any reliable indication, there has been relatively little effort to develop models of the process of teaching that could become the basis for research on the effects of teaching in causing learning. Certainly in English and the Humanities administrators are far more prone to distrust what a researcher might attempt and what he might conclude than they are to consult professional researchers for data that might guide the evaluation of teaching techniques or individual teachers.

Professional researchers on education, of course, continue the search for ways of finding out what learning is caused by what teaching (particularly since many researchers conceive good teaching much as Cohen and Brawer do), and some modest statements about such connections are beginning to appear. In Chapter iv I cited some conclusions offered by Flanders in the 1969 edition of the *Encyclopedia of Educational Research* to support the preference of chairmen for classroom procedures in which teachers make use of student contributions in moving the class along; these conclusions were based on studies of the connection between teaching techniques and changes in students. In the same volume, Professor Gage himself (in an essay on teaching methods) says of "student-centered classes" that even though they have not yet been shown to produce consistent differences in "knowledge gained by students," they do "yield greater gains in higher cognitive processes" (what we might sum up in the phrase "critical thinking") and in "affective outcomes" (changes in attitude).[4] Donald Hoyt's work at Kansas State University on an instrument for student rating of faculty, referred to in Chaper iv, began by asking faculty members to specify what they considered to be important results of instruction that they desired. The rating instrument then asked students how well (on a numerical scale) they thought these results had been achieved in their instruction, and then called for true-false replies to questions about the instructor's teaching techniques. Dr. Hoyt's report of his experiment attempts to show what teaching procedures are most visibly correlated with students' sense of having achieved specific gains from the course.[5] It may be too early to comment on the significance of Dr. Hoyt's work (the faculty in English decided, as a group, not to participate in his study), but its attempt thus to correlate students' sense of their progress in a course with their judgments of the instructor appears to be original and potentially important.

Administrators' distrust of educational researchers does not imply, however, that they are indifferent to what happens to their students as a result of a teacher's efforts. Around 10 percent of the administrators who answered my questions acknowledged that at least one measure of student performance was employed in assessing the effec-

tiveness of their faculty members. The most common of these measures is the grades awarded by the instructor. Some of these administrators are concerned mainly with the fairness of teachers as graders, and others assume (not necessarily with much justification) that a particular distribution pattern of grades is more desirable than another. Whatever the reason for their interest, many administrators said quite explicitly that the examination of the pattern of an instructor's grades in each of his courses was an important part of the evaluation of his work. Some made a point of saying that one of their principal sources of data about a teacher's work was his grade sheets; over a period of years, one chairman thinks, these grade sheets reveal something about the instructor's teaching, although the chairman was not quite explicit in saying what was revealed. One chairman looked for "unusual variations" in a grading pattern as an important datum for use in evaluation. For another chairman the rate of attrition or withdrawal in an instructor's courses was significant. At least a few of these chairmen appear to believe that the achievement of his students, as measured by the teacher, somehow discloses that teacher's effectiveness. These convictions about grades are rarely made explicit and defended, and not much evidence from research studies or other sources is offered in support of them.

Not only grade distributions are looked upon as evidence about the learning "caused" by a teacher. Several chairmen spoke of paying attention to the performance of the instructor's students in subsequent courses. For instance, the effectiveness of a teacher of freshman English is measured in one department by the performance of students in courses to which freshman English is a prerequisite. In another department, tests given at the end of the sophomore year on grammar, punctuation, and reading comprehension are viewed as indicators of the ability of instructors with whom students have worked. Several chairmen are interested in students' performance on departmental comprehensive examinations, and even on the graduate record examination, as measures of instructional effectiveness. Some deans attempt to evaluate the progress made by a student after he leaves the university.

In a few schools, it appears, members of the faculty get together and discuss the work in upper-division courses of students who have worked with particular instructors in their lower-division courses, and sometimes the papers written in these later courses by different students who have worked with different instructors are compared in order to which students have achieved the greatest ess. In these schools it would appear that a principal instrument of measurement of the effectiveness of teaching is the general satisfaction or dissatisfaction of a man's colleagues after they have had one of his former students in their classes—not always the most accurate and dependable measure. On one campus, faculty are asked to reply to questionnaires concerning the achievements of current students, and evaluation of their previous instruction is evidently based in part on repiies to these questionnaires. On another campus such intangible characteristics of students as the care and zeal with which they perform their work are used as data for evaluation of their previous teachers.

The unevenness and subjectivity of these data-gathering procedures are probably obvious enough, though the purpose of using such data is equally clear. An administrator wants to know how a teacher's students fare, and he wants to find out the information without elaborate testing procedures that require expert statistical analysis for interpretation of results. (One can, indeed, sympathize with these administrators, considering the many problems faced by those who try to design tests of the effectiveness of different procedures or persons. Wilbert McKeachie describes a few of these problems in his essay "Research on Teaching at the College and University Level," in Gage's *Handbook of Research on Teaching*, pp. 1122-25.) So he turns to one or more measures of a student's progress that are readily available to him, and assumes a connection between what the teacher did in a course and what the student did thereafter. Distrust of the educational researcher, the man who wants to reduce data to quantitative terms before interpreting it, compounds the administrator's willingness to use data thus informally gathered as bases for judgment—usually they are judgments about teachers rather than about teaching techniques or materials. The presumption here, of course, is that an outside observer or someone looking at the student's performance sometime after the instruction that is being evaluated can say as reliably as an educational researcher, or the student himself, what has happened to the student as a result of the teacher's work.

But occasionally students are asked to comment directly on their own progress in reaching educational goals. In at least one school, for instance, the administration demanded of chairmen that they supply information about the effects of teachers on students, as part of the administrative evaluation of faculty. At that school the first question asked of each chairman on an experimental evaluation form invited detailed comment on the faculty member's effects on students (Appendix C). Part of the information on which the answer is based evidently

came from student replies to questions (inviting free responses) on a form returned by them directly to the chairman.

One way to assess the results of instruction, though increasingly well known today, has become highly controversial. This method is the determination of how well the objectives of a course or of a teacher have been met by his students. This method is based on the belief that educational effectiveness can be determined by testing a student to determine whether he acts differently in describable ways or performs under specified conditions a particular act a larger (or smaller) percentage of the time after he has taken a course than he did before the course. This proposition is based on the hypothesis, as articulated by a prominent researcher, W. James Popham, that "the quality of learning in a given instructional situation is the result of *particular* instructional procedures employed by a *particular* instructor of *particular* students with *particular* goals in mind."[6] Instead of seeking superior instructional procedures that can be employed successfully by *any* teacher, Popham's research focuses on the achievement of goals, regardless of teaching techniques. "Although the instructional *means* may vary considerably from teacher to teacher," Popham suggests, "both may accomplish identical *ends* with equal success."[7] Popham discusses an experiment in developing tests of teaching performance, in which "the teacher is given sets of explicit instructional objectives, asked to teach specifically to them, and has his competence assessed in terms of his ability to produce the pupil behavior changes described by those objectives."[8] The experiment Popham describes, however, compares the abilities of a group of experienced teachers and a group of inexperienced teachers to achieve stipulated objectives in teaching auto mechanics, where desired behavior can be much more easily specified and results of instruction can be much more easily observed than in the Humanities. While Popham finds the results of his experiment "encouraging," since it *can* help an evaluator to determine which teachers attained the objectives better, he recognizes that thus far his data will only support comparisons of effectiveness between groups of teachers, not between individual teachers.[9] Professor Gage also supports this technique of evaluating instruction. "How much students learn has obvious strength as a basis for appraising teaching. This approach is usable in large introductory courses consisting of many sections, one per teacher, under uniform conditions, such as the same objectives, the same textbooks, the same laboratories, and the same ...ze. Such a basis can be objective insofar as ...asures of student achievement are objective. Measures of student achievement can be statistically corrected for major relevant variables over which the teacher has no control, such as the average scholastic ability and motivation of the students in his section."[10]

This approach to evaluation of teaching—through precise measurement of students' attainment of specific objectives—though increasingly popular with some educational specialists today, has not won much support among teachers of the Humanities. There is no reason why it should, because the observable and measurable effects of instruction—if indeed they are effects—are seldom the important effects of instruction in the Humanities, and the kinds of gains that one wants most to see in a student of the Humanities rarely lend themselves to visual observation, let alone quantitative measurement. A further objection, as set forth in a publication of the National Council of Teachers of English, is that "It's nonsense to talk about writing behavioral objectives for the school years when we don't know the consequences of instruction until the student has been away from us for years."[11] Important kinds of behavior that all educators seek—"patterns which describe good citizens, free men, cultured gentlemen," what James Hoetker calls "will-do behaviors"—"are by definition exhibited in times and places far removed from the training situation, so teachers seldom know whether their efforts have borne fruit."[12] (Recall the chairman who speaks of the teacher whose classes, seemingly dull and useless during his course, proved unusually illuminating years later.)

Still, if teaching is thought of as an instrumental art—an art that accomplishes some results—it might be useful for students and administrators to seek some way of deciding, at least in broad terms, what acts students might well be expected to perform differently or better after receiving instruction, and then to find ways (not objective or semiobjective tests that can be graded in neat percentages, but perhaps essays, brief notes, oral contributions to class discussion) that would display whether at the end of instruction the student had increased his ability to perform these acts. In such efforts administrators would naturally have to guard against committing the *post hoc* fallacy; there is no assurance whatever that some of the new skills or abilities demonstrated by the student might not have emerged as a part of his natural growth, whether or not he took the particular course. Still, a pattern of improved writing, or of more articulate and observant comments on poems, novels, or films, might serve at least as *prima facie* evidence that something had happened to the student that might be called "learning," and might justify some

tentatively favorable assessments of the instruction received by these students. Some highly sophisticated teachers of writing, though they would be among the first to deny that students' learning of what they teach can be measured and expressed in percentages of "successes" and "failures" per hundred words of writing, support their claims for the effectiveness of their teaching by challenging their hearers to "look at the papers" that students produce after taking their courses. And many teachers of literature no doubt hope that their students, after completing these teachers' courses, will read better—or at least differently—than when they began the courses.

For most teachers in the Humanities, informal and unspecific estimates of how well the student has learned what the instructor hoped he would learn may constitute the best we can do, in the present state of our knowledge, toward judging instruction by its effects. Until we know more about how teaching works and what kinds of causal assertions can be made about techniques and practices in teaching, we may have to be content with information that an educational psychologist trained in statistics would reject as inconclusive. On the other hand, a fair and deliberate effort to see whether students are performing some jobs better at the end of the course than at the start may be at least as well directed as the inspection of grade sheets, the charting of grade distributions, the drawing of inferences from students' work in later courses, and the assuming of connections between students' performance on departmental examinations and the effectiveness of their instructors. Without in any way climbing on the bandwagon now ridden by advocates of "behavioral objectives," we might still use the correspondence between what was sought and what was achieved, between what was promised and what was performed, as one possible indicator of instructional quality—beneficial at least to the teacher himself and possibly to trainers of teachers, though hardly apt for administrators or citizens as a basis for judging an individual, a department, or an institution. Indeed Professor McKeachie has high hopes for research on teaching, as he shows in his review of the many fruits of such research in Calvin Lee's *Improving College Teaching*—a review intended for teachers and administrators, not for researchers. McKeachie concludes that "we do know more from theory and research on classroom teaching than we are usually given credit for. We have . . . fairly convincing evidence that differing teaching methods do make a difference in learning if one analyzes the different goals of education."[13] Nonetheless, the "ultimate criterion" of whether instruction is effective continues at the moment to elude serious effort to apply it in the teaching of English in college.

[1] *Measuring Faculty Performance* (Washington, D.C.: American Association of Junior Colleges, 1969), pp. 57-58.

[2] "Paradigms for Research on Teaching," in his *Handbook of Research on Teaching* (Chicago: Rand McNally, 1963), p. 118.

[3] "Paradigms for Research on Teaching," p. 120.

[4] "Teaching Methods," in *Encyclopedia of Educational Research*, ed. Robert Ebel (New York: Macmillan, 1969), p. 1455.

[5] Donald P. Hoyt, *Instructional Effectiveness: I. Measurement of Effectiveness*, Kansas State Univ., Office of Educational Research, Research Report No. 6 (Manhattan, Kan.: Kansas State Univ., 1969). Pp. 1-5 review the research and assumptions on which the research is based.

[6] "The Performance Test: A New Approach to the Assessment of Teaching Proficiency," *Journal of Teacher Education*, 19 (1968), 217.

[7] "The Performance Test," p. 217.

[8] "The Performance Test," p. 218.

[9] "The Performance Test," p. 222.

[10] "The Appraisal of College Teaching: An Analysis of Ends and Means," *Journal of Higher Education*, 32 (1961), 18.

[11] John Maxwell and Anthony Tovatt, eds., *On Writing Behavioral Objectives for English* (Champaign, Ill.: National Council of Teachers of English, 1970), p. 22.

[12] "Limitations and Advantages of Behavioral Objectives in the Arts and Humanities," in *On Writing Behavioral Objectives for English*, p. 50.

[13] "Research in Teaching," in *Improving College Teaching*, ed. Calvin B. T. Lee (Washington, D.C.: American Council on Education, 1967), p. 230.

CHAPTER VII
CIRCUMSTANTIAL EVIDENCE

In judicial proceedings, a distinction is commonly made between direct evidence—which places the defendant irrefutably at the place where events in question occurred—and circumstantial evidence—data concerning the defendant's behavior near but not quite at the time of those events, and also miscellaneous information about conditions at the time of the act that imply but do not prove his responsibility. Direct evidence can establish rather conclusively the facts about the defendant's behavior; circumstantial evidence creates a presumption that the defendant may or may not have committed the acts he is charged with. In the assessment of a man's teaching effectiveness, the distinction between direct and circumstantial evidence seems also to be useful. Direct evidence might be thought to include comments by colleagues or peers who have watched the teacher in the act of teaching (either in the classroom or in conference with a student), responses to student writing, the written materials drawn up for distribution to students, and the results of his teaching; circumstantial evidence might include those elements in a teacher's behavior that, while they do not involve direct communication with students, may facilitate a guess or prediction about how the teacher fares when indeed he does communicate in writing or orally with students. For administrators who refuse to engage in direct observation of the act of teaching (in any of its component parts) and who refuse to take testimony from those persons (students) who have observed, indeed experienced, the act of teaching with a particular instructor over several weeks—and even for administrators who do gather these kinds of data—the taking of circumstantial evidence is a significant part of the process of evaluating a teacher. Over a quarter of the administrators who wrote to me about their techniques in the evaluation of teaching admitted relying, in some cases relying almost exclusively, on circumstantial evidence (though they did not call it that) as the basis for appraising members of their faculty.

In the evaluation of teaching, as in judicial proceedings, circumstantial evidence is often easier to gather than direct evidence. No need to alter the human relationships in a teacher's classroom, or to demand copies of his teaching materials, or to eavesdrop on what are assumed to be private conversations, and no need to evaluate the testimony of onlookers (students) who are in a position to observe but may not be qualified (as some administrators believe) to evaluate or describe what they are seeing. Instead, the administrator simply assumes the existence of a plausible analogy: between a faculty member's behavior under a given set of conditions outside the classroom and his behavior within the classroom. Behavior that impresses the chairman outside the classroom leads the chairman to predict impressive behavior within the classroom; behavior that is unimpressive outside the classroom may implant the suggestion (though no one who wrote to me mentioned this possibility specifically) that the faculty member's performance in the classroom is likewise undistinguished. Whether or not the assumed analogy between performance outside the classroom and that inside the classroom holds may admit no testing, because the administrator who observes behavior outside the classroom may get only hearsay evidence that supports the assumed analogy. Then again, if one knows and likes a colleague, one prefers to believe that he is a valuable member of the profession, and one resists admitting that one's likable colleague can be anything less than a valued and effective teacher. "To know a man as a professional colleague," said one of my correspondents, "is to know him as a teacher."

What then are the circumstances that furnish data from which one can infer the characteristics of a teacher's performance? Evidently, any circumstances in which a teacher can form an impression of one of his associates. Several administrators who wrote to me refer to comments by colleagues or members of the department as evidence regarding the effectiveness of some faculty member as a teacher, just as deans, in earlier studies of how faculty are judged, referred to the evaluations by department chairmen as their principal method of evaluating department members. One department is explicitly concerned, said its chairman, with "the way a new staff member impresses the permanent staff: his ideas, his attitudes, his enthusiasms, what he says about how he views his teaching, how he seems to handle his class problems, how he reacts to others' ideas of how they handle their problems, and the like."

From what sources do the department members draw the information about colleagues that is later passed on to chairmen? Administrators who discuss this use of circumstantial evidence are often unconcerned about those sources, but some are quite candid, and the sources they mention might

not exactly promote the peace of mind of younger faculty members. "We listen everywhere," says one chairman. Important occasions for listening, apparently, are corridor conversations, exchanges of ideas over lunch, chats at cocktail parties or other social functions where senior men in the department get together with those whose work is being evaluated. What the senior men expect, evidently, is that form of academic endeavor one might call (it has not been named by others, to my knowledge), "teaching one's colleagues in a social setting"—a kind of teaching that perhaps we can indeed value. It may hardly be consoling for a teacher concerned about his prospects in a department to learn that those prospects may be as much affected by what he says about Faulkner to a senior colleague over cocktails on Saturday evening as by what he does in his classes in the American novel on weekday mornings; but in at least some institutions younger faculty members need evidently to be aware of this possibility. Teaching, said one chairman, is little more than talking about literature; if the teacher talks well about literature with his colleagues, the presumption is that he teaches literature well to his students. Other chairmen emphasized the importance of impressions gained in corridor conversations, and one specifically suggested that he looked with favor upon any colleague who could teach [*him*] *something over a drink.*" (Some chairmen, indeed, take pride in knowing enough about some subjects to find out from staff members' talk how well read those staff members are.) From reading some of these comments, one begins to suspect that lunch with one's colleagues and cocktails in the afternoon after classes are really disguised seminars, in which a young teacher is almost in the position of a student making a report. And one chairman reports that members of the freshman-course committee, concerned more, perhaps, with observing their younger colleagues than with learning from them, regularly look up new teachers of the course for "a casual chat about students, grading, and the relative effectiveness of different assignments," and one wonders whether in other schools, too, conversations about teaching between senior and junior teachers are as casual as they may appear.

Social and other "informal" occasions, of course, are not the only opportunities for young teachers to show how much they know. Another opportunity is the department meeting. Several administrators reported observing the performance of colleagues in department meetings and using the impressions gained there as bases for estimating their colleagues' competence as classroom teachers. A younger teacher's performance at oral examinations is also an opportunity for colleagues to form an impression of the quality of his mind. Another opportunity is offered by committee meetings, where again tenured faculty observe the performances of their younger associates—performances they remember when the time comes to make an estimate of their associates' effectiveness as teachers. A teacher's curriculum proposals, his suggestions for new courses, his thoughts about the revising of established courses, are significant pieces of evidence for many chairmen concerning the proponent's skills in teaching. There is obviously considerable justification for giving weight in evaluation of a teacher to the kinds of courses that a teacher proposes; the danger is that the man who can talk well about his plans or make fresh suggestions about curriculum may not necessarily be able to teach well the courses he proposes or to carry out effectively the imaginative suggestions that he has offered. Still another opportunity is offered by the research in which a teacher engages, and presumably the articles and books that he has published; these are looked at by many administrators as indices of what the author is like as a teacher. One chairman looks in particular at the "critical questions" colleagues feel they have to answer as guides to the kinds of research they engage in and to the ways they disseminate information, hence as guides to the probable emphases of their teaching. Again, there is probably more justification for assuming such a connection than there is for assuming an analogy between a teacher's luncheon conversation and his classroom performance. And publications are, as many chairmen insist, a significant form of teaching. Still, the danger exists that a highly skilled researcher, a man capable of lucidly expounding to his colleagues a scholarly argument, may not necessarily be equally well qualified to appraise the needs and interest of his students and to assist these students in developing an interest in a subject that he is teaching. In his recent report of research at Kansas State University (entitled *Instructional Effectiveness: Interrelationships with Publication Record and Monetary Reward*), Donald Hoyt finds that "Teaching effectiveness and scholarly publication record were unrelated to each other. . . . The assessment of one does not provide information about the other."[1] The cogency of some circumstantial evidence, then, may be in dispute.

Some circumstances reported by administrators to be sources of evidence are probably more reliable than those we have looked at. One such circumstance is the conversation between the instructor or teaching assistant and his "mentor" or advisor. Since the "mentee" is in effect reporting upon his plans and progress, it is probably fair to assume that what he says in such conversations furnishes a clue

to what goes on in his classroom. Another potentially reliable circumstance is the meeting in which junior and senior faculty together plan large courses, especially multisection courses. These kinds of meetings, in which the young teacher contributes specific suggestions about the teaching of a course in which he will be directly involved, or discusses ways of evaluating work turned in by students in the course, may give quite a few clues about the emphases that the teacher establishes, the assumptions that he makes about his students, the procedures that he follows in the classroom, and the clarity and order with which he can discuss points germane to the course. The performance of younger colleagues, particularly teaching assistants, in courses specifically designed to help prepare them to teach—for example, to teach the freshman course in expository writing—seems also a potentially reasonable indicator of how the young teacher may approach the course when he faces his own class. The responses made by a younger teacher to student writing, the sample assignments that he perhaps produces in connection with such a course, the sample lesson or unit that he may teach before his colleagues in order to secure practice in teaching—all these are in effect simulations or microteaching exercises in doing what the young teacher will be doing for real in the future; they make possible some predictions of what the young person may be like as a classroom teacher. These judgments may be all the more valuable if made by persons who themselves have recent experience in the kind of course for which the young person is preparing. A few schools also consider the progress a teaching assistant is making in his own course work, particularly in seminars where reports and discussions are required, in assessing his probable effectiveness as a classroom teacher.

Still another and perhaps equally useful circumstance that may require performance analogous to that in the classroom is the prepared lecture or departmental seminar. Many chairmen say that they try to arrange to bring young teachers before their colleagues for seminar-style discussions or presentation of prepared papers during their probationary period so that the young teachers can establish their credentials before their colleagues, and at least a few chairmen who do not now follow the practice said they would like to install it. One said he would like to see instructors draw up plans for units that might be included in their courses, and present these units before the department for discussion. Although a public lecture or seminar before an audience of colleagues, many of whom are familiar with the materials in question, is not precisely comparable to the classroom filled with young students, it does constitute a quasi-teaching situation in which the new teacher must adjust his presentation and his techniques for leading a meeting to the interests and knowledge of those with whom he is working. It remains possible, of course, for the young teacher to succeed brilliantly with older colleagues or colleagues of his own age and fail abysmally with younger and less motivated students; but at least a limited analogy between the faculty seminar and the student classroom can be argued.

We know, then, what kinds of occasions regularly furnish circumstantial evidence; we know what kinds of evidence these occasions furnish. The large questions remain: who selects the data, who decides whether to transmit it to the chairman, and on what grounds? Few of my correspondents had many comments on these questions. Those who did comment used inevitably subjective designations (preferring to take note of "clear and lively" discussions, "perceptive and skillful" observations). The response to the teacher's performance evidently boils down to whether the listener likes or is interested by the ideas offered by the teacher. Some observers say they look for signs of openmindedness and loyalty, and try to determine whether the teacher is respected by his colleagues. Others ask: Is he stimulating? Does he appear to be happy? Does he sound like an "exciting" teacher? Little evidence came to my attention that the members of any department had discussed or agreed upon characteristics they would look for in the performance of staff members in department meetings or committee meetings, let alone in social conversation (how, indeed, can one "evaluate" social conversation?), though many senior faculty members do consider whether they "learn something" from the conversation of junior colleagues. Even where it is assumed that a senior faculty member or mentor will know or should know the younger instructor well, there is no sign whatever that senior faculty report to the chairman anything more than what the spirit moves them to report. Since the interests and teaching styles and probably the teaching abilities of those who report such encounters are almost inevitably diverse, the young teacher is in danger of being judged on whether his casual remarks and contributions to democratic discussion strike the fancy or engage the interest of the man who has elected to comment. The wisdom, fairness, perceptivity, and reasonableness of the faculty member making the report appear to be taken for granted—although some chairmen do check out the information they get by hearsay from various members of the department. Yet it is difficult to assume that the accumulated bits of hearsay and circumstantial evidence add up to a rounded and accurate

representation of the strong points and weak points of a teacher's performance, even where a chairman's visceral reaction and spontaneous comments from students may more or less corroborate the evidence offered by other faculty members. Thus, when a department asserts, either modestly or proudly, that its faculty are evaluated by colleagues and professional peers, a member of that department or a member of the university administration or an outsider is entitled to wonder on how substantial a basis the evaluation is being offered, even where peer evaluation is solicited on a seemingly precise rating form, such as the one reproduced in Appendix D. In the absence of direct evidence from observation of one or more acts of teaching, it may not be unreasonable to fear that the "professional" evaluation amounts to little more than a summary of circumstantial evidence, gathered perhaps by colleagues who know the teacher well and are trying hard to be fair, but whose tastes in teaching style may be quite different from those of the teacher who is being evaluated. A few letters in my file attest to the danger that these so-called "professional" evaluations may be one-sided, incomplete, and unjust.

Circumstantial evaluation, then, though widely used and influential, is at best only approximate and at worst may be misleading or unworkably vague or both. It can hardly serve well the needs of deans and presidents or alumni or legislators in assessing instruction within a department or a university. At best it should be regarded, like circumstantial evidence in difficult judicial proceedings, as suggestive but often inconclusive. And two important differences between judicial proceedings and the evaluation of teaching need to be underscored; first, the teacher may be totally unaware of the evidence adduced against him, hence unable to rebut it; second, the act over which litigation is occurring usually is completed and cannot be recaptured, but the teacher is always giving new performances that can be observed live, if necessary, and rated by students, so that he need not be judged solely on circumstantial evidence, however appealing and plausible that evidence may be. It is always possible to supplement circumstantial evidence in the evaluation of a teacher, and it is almost always highly desirable to do so.

[1]Kansas State Univ., Office of Educational Research, Research Report No. 10 (Manhattan, Kan.: Kansas State Univ., 1970), p. 13.

CHAPTER VIII
THE TEACHER EVALUATING HIMSELF

Thus far we have looked at ways in which teaching can be evaluated by the teacher's clientele, by observers of the act of teaching, by those who look at the results of the teaching, and by colleagues who neither see the teaching performed nor examine its results, but instead combine impressions of the teacher from informal conversations and departmental meetings to create an image of what the teacher may be like as he teaches. There is, however, one party to the transaction of teaching who is present continuously when the teaching is planned and when it is going on, but who is rarely encouraged to participate in the evaluating of that teacher. This party is, of course, the teacher himself. Granted that he sees what is going on from a distinctive point of view, and granted further that—especially if his rank or tenure or salary hinges on the judgment finally rendered about his teaching—he is a biased observer of the transaction, the teacher himself is nonetheless in a position to supply valuable firsthand information about what has happened in his course and about the results of what has happened. Sometimes, indeed, the teacher's own reports are virtually the sole source of information with which to apply such criteria as whether or not the teacher uses the newest teaching techniques available (such as instructional media) or devises techniques of his own—criteria that seem important in the guidelines for evaluation of teaching at some schools. If he is invited to express an opinion, he can also supply valuable information indirectly by the items that he chooses to talk about, the emphases he establishes, the kinds of questions he asks of his own teaching. Some chairmen, to be sure, believe that a faculty member's appraisal of his own teaching is likely to be little more than self-glorification. But others expect faculty to be reasonably accurate in evaluating themselves and they think self-evaluation may help teachers to change their behavior. If there is any chance that the latter expectation can be realized, it is to everyone's advantage to ask the teacher's opinion about the success of his own teaching, as in fact is done in a small minority of colleges and universities with which I corresponded.

How extensively chairmen encourage members of their faculty to evaluate their own teaching, however, is not altogether clear. Something over 10 percent of the administrators who answered my questions indicated that one means by which they arrive at judgments about teachers and teaching is to talk with the teachers, particularly teaching fellows or instructors. If these conversations provide opportunities for the teachers to describe in detail the distinctive characteristics of their teaching as they see them, and to tell in what ways they think their work has been effective and in what ways perhaps less effective, the conversations may give a real chance for older faculty members to assist with evaluation of their work. Too often, unfortunately, the conversations between administrators and younger faculty members apparently are devoted only to discussion of complaints from students or to reactions from the teacher's peers, rather than to a searching exploration by the teacher of his techniques and emphases. Of course, teachers have a right to know the criticisms being made of their work, so that they can answer the criticisms or try to alter the disturbing behavior. But relatively little self-evaluation takes place in a confrontation in which the teacher acts more or less as defendant in an inquisition. A statement by his chairman that he talks with members of his faculty concerning reports of their teaching is not necessarily evidence that each faculty member is encouraged to engage in self-evaluation.

On the other hand, it is evident that some administrators do provide for conversations with their colleagues, particularly younger faculty, in order to find out exactly what their colleagues are doing in their classrooms. A few chairmen apparently expend considerable effort to find out exactly what members of their faculty are doing and how successfully they think they are doing it. One chairman spoke of conversations in which he asked his staff members to reveal their "convictions, plans, and successes." Another said that he regularly requested accounts of practices employed by the staff members in their teaching, and of the courses that they had designed for themselves. Others spoke of conversing with staff members concerning their special strengths as teachers, problems they perceived in their teaching, and weaknesses they significantly needed to eliminate. One chairman said that he placed a great deal of weight in his overall evaluation on statements by the teachers regarding their own work. And some of the chairmen whose remarks are summarized here evidently gather syllabi and statements about the content and form of the teachers' courses in order to gauge the consistency between what the faculty member says he does and what he does in fact do in his classes.

Still, the number of chairmen who spoke of giving members of their departments an opportunity to talk with them about the goals and emphases of their teaching, and their success in reaching those goals, is but a handful out of the group that attempted to answer my questions. There are no doubt many possible explanations for the small number; some chairmen may preside over departments that are far too large to permit the chairman annually to confer with each member of his staff; some chairmen may work with faculties that would resist what might appear as an annual confrontation between themselves and staff members (particularly senior staff); some chairmen may honestly doubt, as a few said they did, that the comments by an instructor or professor offer any reliable information about the effectiveness of their teaching, and many chairmen may simply not have thought about the possibility of talking with their colleagues about teaching.

The practice of asking teachers to comment formally in writing on their teaching is even less frequent among my correspondents than that of inviting an informal, oral self-evaluation. Fewer than 10 percent—several of them in junior colleges—said they made a point of collecting written data that I could in some way identify as self-evaluations by teachers. A few chairmen did say that they asked their associates for statements of the strengths and weaknesses of their teaching and for assessments of how well their teaching in that semester had gone, or that they asked staff members to describe the areas of teaching in which they thought they wanted especially to improve. A few chairmen ask faculty members to complete self-evaluation questionnaires, some quite detailed, others (like the one reproduced in Appendix E) less detailed. A very few of these administrators attempt to discover how well their faculty members are doing by asking for a statement of the goals sought by the faculty member and of the methods he uses to attain those goals. But in some of these schools forms for such statements are made available to faculty only when issues of tenure or promotion arise; rarely are teachers urged to make an annual written assessment of their teaching. Some of the forms employed by administrators ask one or two questions about the teacher's teaching and devote the remainder of what looks like a self-evaluation form to a request for a list of the publications, professional meetings attended, papers read, and so on, during the last year.

Thus, the following request for a self-evaluation, made annually of all faculty members at a midwestern university in which the evaluation of teaching has become an important concern of all members of the faculty, is of distinctive interest, even though self-evaluation is optional with each faculty member:

> Evaluate your effectiveness as a teacher this term. To do this, you might comment on such aspects as the following: your teaching methods; the kinds of examinations, papers, projects, etc., you use or require; how effectively you believe you have communicated with your classes; the extent to which you have achieved your goals in your classes. As part of your response, discuss the kinds of changes, if any, you would make if you taught the same classes again.

We may teach our students that from writing down data about their experiences—from composing and interpreting experiences by selecting words to represent them—we learn a good deal about the experiences and about ourselves. But few administrators in higher education evidently are asking members of their faculty to practice what they preach. The opportunity for self-discovery and self-knowledge through written evaluation of one's own teaching is not often given to teachers of any rank in our colleges, universities, and junior colleges today. But self-evaluation forms may become increasingly prominent among our data-gathering instruments: the Muscatine Committee in its report, *Education at Berkeley*, includes as a part of its recommendations the proposal that "a statement by the candidate describing the rationale of his teaching efforts" be a part of the dossier that every candidate must help the chairman compile when he is to be considered for tenure or promotion.[1]

The possibilities of self-evaluation are not new to writers on the evaluation of teaching; articles and comments on the desirability of self-evaluation have appeared occasionally over the last decade. Their approaches and their theses vary. Some writers have typically dealt with the methods for gathering the data for self-evaluation (e.g., meeting with colleagues, taping class sessions), as Drs. Cohen and Brawer do in their monograph on *Measuring Faculty Performance*.[2] Bruce Biddle views self-evaluation statements as analogous to the self-reports used in sociological studies—as means of getting data that are otherwise impossible to gather concerning classroom activities. Because the comments in self-evaluation reports will not reflect "adequate" categories for discussing classroom situations, Professor Biddle is in doubt about the reliability and value of such reports and discourages their use where more objective data are available.[3] Some articles enumerate questions (with implied right answers) that the teacher may put to himself in thinking back over his teaching. In an article published in the fall of 1962, Professors Herman

Estrin and Ruth Goodwin enumerated a series of topics for the teacher to use in reviewing his performance, and concluded their statement by citing three questions recommended for consideration by teachers (in this case elementary and secondary teachers) by Professor Paul Farmer:

> What did I do this semester to help pupils find satisfaction in using their strong points and in improving their weak points in reading, writing, speaking, and listening?
>
> As a person who believes that literature is a living, vital force in one's life, what books did *I* read for my own satisfaction?
>
> What creative work of my own, factual or imaginative, caught the interest of my pupils and helped to build mutual confidence and respect?[4]

A year earlier in the same publication Ordway Tead had also urged upon teachers the practice of self-evaluation, and had enumerated several questions that teachers might put to themselves concerning their work.[5] Several years later (in 1966) Gertrude Lewis offered her own series of questions for use in self-evaluation; she was writing for elementary and intermediate teachers primarily, but some of her questions, too, might be applicable to college teaching.[6]

The difficulty with these questions is that the teacher is asked to judge his own performance by someone else's model, rather than making an independent judgment of what is important to him and what criteria he feels he should stand or fall on. The practice followed by some chairmen and by the midwestern university whose evaluation form I have cited above seems much fairer: it encourages the faculty member to determine what *he* hopes and wishes to see happen in his students as a result of his teaching, and then asks him to say whether or not he thinks he has accomplished what he sought to do and to what extent he has accomplished it. Questions of this sort allow the teacher to define his own ground; further, they encourage the administrator to ask of the teacher in what ways *he* thinks he is especially effective as a teacher, in what ways less effective. Data-gathering for self-evaluation, of course, is as difficult as for other kinds of evaluation. Unless the teacher has a dependable measuring device by which to determine how well he has accomplished his objectives, he has to rely on an intuitive reaction to students' behavior, to their participation in classroom activities, and to their performance on examinations in order to make his evaluation. He can, of course, gain a little perspective on his own teaching by watching one or more of his colleagues in action (with their permission), and such observations of colleagues may be very useful, but the teacher still relies on intuitive perception when he compares his colleague's work and his own. Yet, that intuitive reaction *may* be as dependable as the results of everyday techniques for measuring learning (in most classrooms hardly an exact science today); the behavior of students may give a teacher clues concerning the success of his work, even though it cannot inform him conclusively of whether his intuitions are sound or unsound.

Self-evaluation, then, substitutes the dangers of biased and one-sided perception for the limitations on data-gathering powers of other agents: the intrusiveness of outside observers and the limited scope for their reflections; the limited perspective and possibly different biases of students. But whatever else it may gain or cost, it clearly places upon the instructor himself the responsibility for some decisions concerning his own professional work, and in so doing contributes perhaps both to the teacher's independence and to his sense of responsibility; it may also help greatly in improving his own performance. It may be impossible to allow the evaluation of a teacher by himself to stand as the only basis for determining a teacher's effectiveness (one chairman wrote that at his school the teaching of a faculty member is accepted as "effective" even if the only proof of his effectiveness is his word), but if a teacher's evaluation of his own work is considered alongside data from other sources accessible to administrators, it may help to complete and indeed make vivid a picture that would otherwise be shadowy and inconclusive. Furthermore, a teacher may be able to attain objectivity in talking about his teaching techniques, and he may be able to give information about the values of particular teaching procedures that is accessible in no other way. Self-evaluation, too, can help the faculty member and the administrator to understand the faculty member's work and his potential for unique contributions to the department, as well as to identify the courses and activities to which the faculty member might well not be assigned. For these reasons, self-evaluation may deserve much more widespread use than it currently is getting.

[1](Berkeley: Univ. of California Press, 1968). p. 44.

[2](Washington, D.C.: American Association of Junior Colleges, 1959), pp. 10-11.

[3]*Contemporary Research on Teacher Effectiveness* (New York: Holt, Rinehart, and Winston, 1964), pp. 27-28.

[4]"Evaluating College Teaching," *Improving College and University Teaching*, 10 (1962), pp. 195-96.

[5]"Teacher Self-Evaluation," *Improving College and University Teaching*, 9 (1961), p. 148.

[6]*The Evaluation of Teaching* (Washington, D.C.: National Education Association, 1966), pp. 84-86.

CHAPTER IX
SOME CONCLUSIONS AND QUESTIONS

Today, in response to current interest in evaluation of teaching, many chairmen argue that their faculty members regularly and wisely evaluate their own teaching techniques (and thus maintain the quality of teaching throughout the university) and that any attempt systematically to gather data in an organized evaluation procedure is an expression of no confidence in the professional qualifications, indeed in the integrity, of people in their departments. Such treatment is simply intolerable; those with tenure in upper academic ranks have—so the argument runs—demonstrated the ability not only to teach but to make continuous improvements in their teaching. After all, say chairmen, is not the award of tenure and promotion based upon their demonstrated superiority as teachers? Their universities or colleges assert that effective teaching is an essential criterion for promotion; since the promotion and the award of tenure have come, the criterion must once and for all have been met.

And even if one sets aside momentarily the argument that evaluating the performance of a teacher or the efficacy of a teaching style is unprofessional rudeness, these chairmen assert that the adoption of a systematic process for gathering data, let alone for interpreting it, implies a regularization, a formalization, of an act that by its very nature does not admit of generalized description, a humane act that ought under no circumstances to be degraded by mechanical probes or investigations. These chairmen implicitly accept the model of a department as a community of equals, each always ready to help the others but under almost no circumstances willing to sit threateningly in judgment on them. Besides being inhumane, these chairmen hold, evaluation procedures are misleading: they are, in the words of one chairman, subject to "human fallibility, prejudice, and bias." And they threaten the teachers, who must, of course, be allowed to carry on their work in an atmosphere as nearly as possible devoid of threat. To avoid creating "an atmosphere of uneasiness and mistrust," says another chairman, his department seeks "an atmosphere in which young teachers know that they are being trusted to do well and that a wide variety of *help* [italics mine] is available if they will ask for it." Paralyzing fear of being rude to colleagues, resistance to infringing academic privacy, denial of the possibility of developing general criteria for effectiveness in teaching, and a determination to protect teachers from the threat to self-confidence that evaluation might bring—all these feelings are writ large in the letters of many adminstrators who wrote to me. If teaching is an art, there are many practitioners of the art who insist that no standards of judgment can be developed for it, even though the search for such standards is a source of lively discussion in other arts. I suspect that many administrators, even those who are concerned with improvement of teaching in their departments, wish that the demand for evaluation could be made to go away. One chairman, speaking no doubt the sentiments of many, concluded his comments by telling of his efforts "to encourage continuing engagement in examination of teaching problems and innovative ideas through constant dialogue within the department and inter-collegiate dialogue by means of articulation and other conferences. This method has the advantage of respecting the ideas and methods of individual teachers and avoiding the antagonisms aroused by investigative procedures. If pursued zealously and creatively enough, it might even do away with the necessity of most evaluation."

But if the respondents who hold these views are set to one side, and respondents who view systematic evaluation as a desirable instrument for their departments' professional improvement become our major sources of information, it remains clear that the art of evaluation, though it has been with us for many years, has by no means come to maturity. What is undeveloped is our understanding of what "teaching" embraces, and which methods of gathering data will enable us to say how well that act, whatever it may embrace, is being performed. There is no dearth of notions about what "teaching" may include, but the notions are sharply divergent, and rarely address the question of what is supposed to happen within those, our students, who have been the objects of teaching, the clients of the teacher. And even among those who hold fairly sophisticated notions of "teaching," the wisest and most dependable ways for getting data that will disclose whether effective teaching has taken place are hotly in dispute. The task of gathering these data is complicated, even for chairmen who shoulder it willingly, by the necessity to maintain good manners in the treatment of one's colleagues, to assure that absolute fairness and justice are observed, and to avoid interfering with and thus altering the very act that is to be monitored. Though quite a few chairmen were fully satisfied—

some aggressively and some complacently satisfied—with their evaluation of their faculties, many chairmen with whom I corresponded spoke diffidently of the procedures they used in evaluation (some said their procedures worked as well as any could), and some admitted embarrassment at their inability to give crisper reports about their gathering and interpretation of data about teaching.

Embarrassment, of course, is by no means in order among chairmen who see weaknesses in their data-gathering procedures. Most chairmen and administrators in departments of English and the Humanities have given little of their time to the formal study of teaching. The study of teaching is the professional domain of members of the faculty of education in most universities, and in most universities there is a wide difference in interest (as well as a substantial difference in prestige) between the teachers of humanistic disciplines and the professional investigators of education. Hence the lack of information about educational theories among many chairmen of humanistic departments. (Among administrators who oppose the idea of evaluating teaching are many who insist that they do not wish to have the "educationists" invading their territory.) Then, when one turns to the work that educational theorists and researchers are doing, one finds varied models of the act of teaching, varied analyses of its logic, complicated ways of studying interactions in the classroom, and much analysis of alternative teaching techniques and strategies. But one gets comparatively little help in gathering reliable data for evaluation of teaching, despite the increasingly large numbers of articles and books on evaluation emerging from schools of education and centers for research today. No one yet knows how to establish a dependable connection between an act or acts performed by an agent whom we call a teacher, and important changes in the persons we call them "students"—who interact in some mysterious way with the teacher. Some data suggesting correspondences and correlations between kinds of teaching behavior and kinds of student response turn up in the literature (the number of such studies is increasing), but however painstakingly accumulated and sensitively interpreted the supporting data may be, most of the discoveries of educational researchers are limited in scope, subject to qualification, and open to challenge from other studies. Not surprisingly, few administrators of colleges and departments of arts and letters have had the time to familiarize themselves with the detailed, and sometimes doubtfully rewarding, articles that report educational research, particularly since understanding of that research often requires familiarity with mathematical and logical techniques well beyond the grasp of other than highly skilled professional researchers.

And so a chairman or dean who would evaluate teaching must work to accumulate impressions, opinions, intuitions, and feelings, and must hope that somehow these impressions and feelings may produce a judgment that roughly corresponds to reality. In this discussion, I examined most of the available sources of these impressions: the judgments of students about their increases in skill, their relationship to their teachers, their understanding of the subject; the assessments by colleagues who have observed the teacher in action and who appraise it according to their values; the impressions colleagues get of a faculty member's behavior that suggest what he might be like in dealing with students; the faculty member's perceptions of the effects of his work on his students; and so on.

The value of these data appears not to be much in doubt among chairmen at two-year colleges. The administrators of two-year colleges take the teaching responsibilities of their schools so seriously that they adopt complex data-gathering programs for evaluation of faculty. Perhaps inevitably, a counterreaction sets in at many two-year colleges: the faculty, in some schools unionized, negotiates with the administration a forbidding set of agreements on procedures for review of salary and qualifications for tenure, in order to insure that the faculty member is judged fairly, that he has every opportunity to know and answer the evidence gathered against him, and that he is kept fully informed of his status so that he may make plans to move elsewhere if the school at which he is working appears likely to vote against him. The most intricate procedures for evaluating faculty are those in some two-year colleges, where money for carrying on these procedures apparently is allocated without much question; compared to some data-gathering and appeals procedures in use at two-year colleges, the practices for determining teaching effectiveness during deliberations about salary, rank, and tenure at four-year colleges and universities seem casual and informal, however many committees must review departmental recommendations before final decisions are made.

Despite, or maybe because of, the informality of many data-gathering techniques, many chairmen at four-year colleges and universities who are sympathetic to evaluation are doubtful of their success in carrying it on. Many said that they would like additional data for that purpose; these chairmen observed that they were not doing enough to assure sound judgments, and expressed the wish for more money, more time, more released time for full-time faculty members, to improve data-gathering tech-

niques. The improvement most often wished for by such chairmen was additional time for observing classes and for talking with faculty members; a few chairmen said they wanted better instruments for student rating of faculty, improved procedures for determining the results of instruction, and time to encourage their colleagues to undertake more thoughtful evaluations of their own work. Those who expressed confidence in the possibility of evaluation, that is, tended to believe that sound judgments came from observation of the act of teaching and observation of the mind of the teacher as revealed by his talk.

Given the knowledge about teaching accumulated thus far by those who study it, what exactly is the value of sophisticated procedures for gathering data? Would additional time for observing classes and talking to faculty members give more dependable information about teachers or teaching techniques than is given by less intensive procedures already available and already practiced? Will any amount of such data, given our present knowledge about teaching, help an observer, however well trained, to say of a teacher that he is good or mediocre or inferior, or to say of a method that it is effective or ineffective? However conscientious we are today in gathering data, can we fairly judge every teacher? Can we responsibly assert that one teaching method is to be cultivated and another to be abandoned? Can we urge that some conditions for instruction are beneficial, and others are harmful and always to be avoided if at all possible?

Since the answers to these questions are honestly in doubt, we may not need merely new procedures for gathering information about teaching so much as a new spirit and outlook in attempting evaluation. We may, of course, need much more than a new attitude. Edward Shoben, Jr., of the American Council on Education urges some rigorous conceptual thinking about teaching:

> The lack of any significant conceptual framework within which to set the problem of teaching makes teaching hard to evaluate, difficult to improve, and a little dull to discuss. . . . our goals of improved teaching may . . . be attained more readily if they are thought of as by-products of an effort to reformulate teaching as an intellectually interesting problem, worth the time of men who devote themselves professionally to difficult but significant conceptual affairs.[1]

Lacking a fresh conceptual framework to set forth, I attempt in my last chapter the more modest task of suggesting a fresh outlook on evaluation.

[1] "Gimmicks and Concepts in the Assessment of Teaching," in *Improving College Teaching*, ed. Calvin B. T. Lee (Washington, D.C.: American Council on Education, 1967), p. 295.

CHAPTER X
SOME RECOMMENDATIONS

The teaching profession, then, at this moment confronts a troublesome paradox. It is pressed by students, citizens, legislators, and private donors of funds to make an accounting of how well its members carry out their professional responsibilities. Yet the art of demonstrating success in teaching, despite the attention it has received and the numerous instruments for practicing it, is still not greatly advanced—and the art is one that some teachers do not like to see practiced, let alone mastered. In view of this sharp opposition of desires, and of our present difficulties in accounting with any confidence for a teacher's effectiveness, what steps can be recommended to department chairmen, deans, and other college and university administrators who must make judgments about the quality of educational programs, the quality of teaching techniques employed by members of their faculty, and—perhaps most important of all—the quality of professional service rendered by individual members of those faculties?

Without assuming that what follows is in any sense a panacea or a comprehensive solution to this dilemma (or to the problems in evaluation elaborated earlier), I offer here some suggestions for administrators and faculty alike. Many of my suggestions urge new attitudes toward evaluating rather than recommending techniques of evaluation. But suggestions about attitudes seem particularly in order, since attitudes to such a large extent *create* the problem of how to carry on evaluations. Perhaps these new attitudes will emerge more readily if I begin modestly, by recording first some assumptions on which most teachers and administrators might agree, before going on with suggestions about ways of approaching evaluation.

First, I propose that the problem we face be redefined. As I have suggested in this essay, the question to be considered is that of how to gather *specific* and *reliable* data on which to base judgments. The making of judgments is an act we cannot escape; when we promote or deny promotion, when we support or oppose the introduction of innovative teaching techniques, when we say yes or no to new curriculum plans, when we increase salaries, we are making evaluations. We are saying (if effective teaching is one of our criteria for academic advancement) that the man is a sufficiently successful teacher, or that the new technique will or will not bring benefits to justify the cost, or that the new curriculum is or is not superior to the previous one—all of them evaluative judgments. Unless we want to abandon any claim that we consider the effectiveness of a teacher, of a teaching technique, or a curriculum, in the making of decisions about those subjects, we cannot escape the act of evaluation. The issue at stake is really how to gather sufficient data, accurately and fairly, to permit a sensible evaluation, and get those data without disrupting or fundamentally altering the activities that we are trying to evaluate. The issue is whether we want to evaluate on assumptions and guesswork, or on evidence. The absence of evidence is not proof of success—or of failure.

Second, I suggest that we accept openly the soundness of the comment made by many administrators: that no one teacher, teaching technique, or curriculum works exactly the same way or with the same degree of effectiveness at all times and in all circumstances. Teachers have their good days and their less good days; they often work better with some kinds of students than with other kinds, on some subjects than on others, under some conditions than under others. Teachers and curricula differ in the kinds of goodness that they achieve and in the weaknesses that they exhibit. Teacher A may be especially successful in using certain techniques and unsuccessful with other techniques; teacher B may succeed where teacher A failed and be less successful where teacher A is strongest. These conclusions are by no means startling, and enlightened administrators have long insisted upon them vigorously, but many of our procedures for rating teachers ignore these differences and try unjustifiably to generalize about the processes of "good teaching." Teaching may have to be evaluated by its products, not its processes.

Third, it follows that a search for absolute comparisons among teachers of different subjects, or indeed among teachers of the same subject, is doomed to be misleading. To ignore the differences in ways in which teachers' techniques are effective and in the circumstances where they are most effective is to blind ourselves to essential human discriminations. To say, as some student rating forms encourage interpreters to say, that in overall effectiveness teacher A scores 8.1 on a ten-point scale and teacher B scores 7.5 on the same scale, is in effect to provide very little information that is of use to an administrator or to a citizen or to a student. To say that teacher X can be rated B- in "motivating students to do their best work," while

66

teacher Y rates C+ on the same question, is again to provide no useful information about the teachers involved. What is needed is an art of evaluation that will make emphatically clear the ways in which each teacher is effective, the circumstances in which he is effective, the kinds of students with whom he works best, and that will also make clear in what ways and in what circumstances he is less effective. If no teacher or instructional technique or curriculum designed is always bad or always good, the only sensible form of evaluation is the one that says wherein each is superior and wherein each is of only average effectiveness, or less. Rather than codify simplistically some criteria for effective teaching, rather than set absolute, numerically measurable goals for a teacher to attain, we need to discriminate degrees and kinds of excellence in teaching.

Fourth, it follows further that our instruments for data-collection concerning teaching may need drastically to be revised so that they answer questions that are worth asking. The instruments should help us *describe* precisely what the teacher or the method or the course does and does not do, *then* show how well each thing it does is *accomplished*. At a minimum, our student rating forms should be redesigned to disclose what acts of the teacher students thought especially successful and what acts they thought less successful or of no special value, what materials the teacher covered with special effectiveness and what materials he dealt with less satisfactorily, what kinds of responses from students he handled most effectively, what kinds less effectively, and so on. It may be possible to continue using some items currently prominent in our evaluation forms so that computers can give us some information. But it is equally possible that, unless long and intricate rating instruments covering all possible kinds of teaching activities are to be developed, we shall have to retire the computer from an active role in interpreting student ratings and assign the responsibility for collating students' assessment of teachers to a group of wise human readers who know what differentiates teaching techniques and strategies and who can bring together students' responses in an informative way. Where self-evaluation by instructors is encouraged, the evaluation questions asked of the teacher might well discriminate between ways in which he believes he is especially effective and ways in which he may fall short of his goals.

It is likely also that those who observe classes will have to be asked (perhaps trained) to report specifically what went particularly well, what materials the teacher seemed best equipped to deal with, what techniques the teacher employed most effectively in the classes observed, and, if necessary, the contraries of these items as well. If observation of classes is carried on, moreover, it seems unfair to departments and students for the results of such observations not to be available to administrators responsible for the teaching assignments and for the instructional support given each member of the faculty. If the role of the observer is identified as the setting down of what is effective and what is less effective in a teacher's performance, rather than as the rendering of an absolute judgment, the threat implied in a procedure that reports the results of observations to an administrator may be diminished.

Fifth, just as in the assessment of an event, a person, or a nonacademic problem, the availability of several kinds of data can help insure fairness of judgment and soundness of decision, so in the evaluation of a teacher, an instructional technique, or a curriculum, data recorded from different perspectives may give a more complete picture than would data from only one source. An administrator is more likely to know well what a teacher or a technique can and cannot accomplish if he has the views of colleagues who have observed the act of teaching (both as performed in the classroom and as performed outside the classroom—in the preparation of syllabi, examinations, and so on), of students who have tried to learn from the teacher, of the teacher himself, and—possibly—of observers who have noted the performance of the teacher in noninstructional or semiinstructional settings. These complementary perspectives, perhaps supplemented by some assessment of the changes that have taken place in students after exposure to the teacher, may supply a sufficiently rounded view of the teacher for use in decision-making. At the same time those whose responsibility is to engage in evaluation must guard against the tendency—especially noticeable in some two-year colleges—toward overemphasis on evaluation to the point where the teacher may be more concerned with his evaluation than with finding the best ways for instructing students. Overevaluation is a real danger—one that we should not minimize, but one that should not deter us from a flexible and balanced program of evaluative procedures, judiciously selected to fit whatever is being evaluated.

Sixth, those who are responsible for selecting instruments for evaluation, and those whose work is to be evaluated, should recognize the legitimate interest of students in the quality of their instruction, and should give students a chance to be heard by those who make decisions about teachers, curricula, and teaching assignments. The sense of responsibility exhibited by students in rating teachers is amply documented. If students tend to be lenient in their ratings, if they react more favorably

to elective than to required courses, if they judge faculty in the upper ranks more favorably than junior faculty, these trends can be taken into account, but they should not take away from the group that is, after all, most vitally affected by a teacher's work, and is well situated to discriminate the kinds of success enjoyed by their teachers, the opportunity to affect in turn the quality of that work. And if teaching is to be judged by what it achieves (however achievement is identified), students' perceptions of what has happened to them in the course of, and as a result of, their interaction with a given teacher may be as valuable as any data we can now get about that teacher's achievements.

Seventh, we need not confine ourselves, I think, to the traditional data-gathering devices of observation in the classroom and the taking of student opinion. Any procedure that enables the administrator to learn about the performance of a teacher or the usefulness of a technique, or to make reliable estimates about his probable performance of the act of teaching, may contribute to our efforts at evaluation. Thus team teaching, while giving the students the benefit of varied perspectives on the subjects under study and of arguments by different faculty members in support of conflicting views, also gives faculty members the opportunity to appraise strong and less strong features of each other's work, so that the profile of each teacher becomes clearer and sharper. Team teaching need not be adopted simply to provide a fresh approach to the evaluation of teaching, but this incidental value of the technique need not be overlooked. Also, the providing of occasions on which a young or inexperienced teacher can help to teach his colleagues, by leading seminars or giving public lectures or discussing in faculty groups instructional techniques he uses or wants to use, can furnish information that administrators may use in predicting how the instructor will teach students.

And self-evaluation, though infrequently used, can give both the administrator and the teacher himself some valuable insights into the teacher's effectiveness. Not only what the faculty member says he has accomplished or not accomplished, but also the topics he selects for emphasis and the evident honesty with which he confronts his teaching will assist in the making of judgments concerning that faculty member. Furthermore, if to put fleeting glimpses, half-hidden perceptions, and previously unverbalized intuitions into language for contemplation is a way of making discoveries about oneself and one's world, the very act of self-evaluation can be an important means for self-discovery for each faculty member. Nor need there be fear that a faculty member will glorify himself insupportably in the hope of winning recognition to which he may not reasonably be entitled. If complementary ways of viewing a man's teaching are used, and if there is appreciable difference between his assessment of himself and the assessment suggested by other data, the difference may justify concern about the teacher's perceptiveness and lead to useful discussions with the teacher about the disparity between the way he sees himself and the way others see him.

Eighth, administrators and teachers in departments of English might well give some thought, and some open discussion, to what the teaching of English in college might comprise. The question admits of no one answer, of course; for different teachers with different styles and different interests, the answers will diverge widely. But comparing the divergent answers may be itself a healthy activity. It will be good, for example, for a teacher who views his job as training students in the application of mid-century critical techniques to literary works—if there is a teacher who would describe his goals thus—and a teacher who seeks mainly to give his students a lively understanding of human nature in all its diversity, it joys, and its troubles—probably many teachers, after the Dartmouth Conference, see their jobs thus—to share their assumptions with each other, which they may anyway do regularly in social conversation, and to consider quite specifically how their work may be evaluated, which they are much less likely to do in ordinary conversation. If teachers of English are to be judged by their clients and their employers, they might prefer to tell their judges by what standards they should be judged, rather than allow outsiders, unfamiliar with the unique features of their field, to set the standards of judgment for them.

Ninth, administrators and teachers alike should remember that what is to be sought in evaluation is not simply or mainly *judgments* on people, but an understanding of how people learn, what they want to learn, what they value in learning, and thus how they can best be taught. Instead of setting out evaluation instruments as tools for threatening and humiliating colleagues, we should take care that these instruments appear as aids in the practice of our profession—as our way of finding out how well we are doing, like the verdict rendered after lawyers have pleaded their case—and as some of the tools by which we learn what our profession requires of us and how we may meet its challenges. From feedback comes a better understanding of how a process works; feedback from teaching, if carefully observed, may clarify how the process of teaching works, too.

Tenth and finally, all who are concerned with

education might remember that evaluation, broadly defined, is a human act—indeed it is one of the principal acts for which teachers are preparing their students. It is a complex act, requiring rigor in the examination of one's values, fairness in the accumulation of data, and wisdom in the interpretation of those data in light of the examined values. It may even be, indeed, an art. But certainly it is an act not unworthy of a professional person to practice, and to profit from the results of evaluation ought not to be unworthy of a professional person, either. As members of a profession, we need to be concerned for the effectiveness of that profession; we need to monitor our own work.

If evaluation humbles, humility is no despicable virtue in any profession, certainly not in the teaching of the Humanities. Indeed the humility engendered by seeing ourselves as others see us may help to save us as members of a profession. Humility from self-knowledge among its members helps to keep the entire profession alive, better equipped to identify and defend its genuine accomplishments before those who would degrade it. Evaluation, then, as a means to self-knowledge, is an instrument for both preservation and reform.

I do not believe that the evaluation of teaching needs to be the feared instrument for compelling academic conformity, wielded by a reactionary and bureaucratic administration, that some chairmen and teachers envisage it to be. Not only is it a humane way of achieving valuable self-knowledge and an essential step toward professional self-improvement, but it can even be, as the director of one program for use of student ratings said, a form of protection for the faculty member against rumors and gossip. And I do not believe that at the present moment of radical reexamination of universities' educational programs, we can afford to indulge a timid, defensive view of evaluation. Teachers are responsible for putting forth their best professional efforts, and they are equally responsible for assuring that the efforts they expend are the best of which they and their colleagues are capable. Moreover, as one director of an evaluation program puts it, "If we have confidence and take pride in our workmanship as teachers, as we do, ought we not to feel strong enough to submit our work to inspection and take satisfaction in that fact?"

At the same time, new attitudes and ways of looking at the goals of evaluation are overdue. We will not be able to assure ourselves that we are doing our best work if we adopt only those techniques for evaluation that place faculty members, teaching procedures, and courses on numerical rating scales or that reach summary judgments on criteria that may differ appreciably from observer to observer. A department needs an inventory of its most powerful resources, so that those resources may be deployed where they will be of greatest benefit to students, university, and community at large. That kind of deployment of resources should be the goal of administrators, chairmen, teachers, and students alike. But the art of evaluating the teaching of English, if my investigations are reliable, is today not equipping us to take a precise inventory of our resources. If the art of evaluation can possibly be improved, we should try to make it tell us exactly where each of us can do the best teaching of which he is capable. With that kind of information, a chairman can assure his students, his dean, and the citizens who support him that each course in his department is taught as well as the school's resources will allow.

SELECTED BIBLIOGRAPHY

Amidon, Edmund, and Elizabeth Hunter. *Improving Teaching*. New York: Holt, Rinehart, and Winston, 1966.

Anderson, C. C., and S. M. Hunka. "Teacher Evaluation: Some Problems and a Proposal." *Harvard Educational Review*, 33 (1963), 74-95.

Biddle, Bruce, and William Ellena. *Contemporary Research on Teacher Effectiveness*. New York: Holt, Rinehart, and Winston, 1964.

Brudy, H. S. "Can We Define Good Teaching?" *Teachers College Record*, 70 (1969), 583-92.

Bryan, Roy C. "Student Ratings of Teachers." *Improving College and University Teaching*, 16 (1968), 200-02.

Busacca, Basil. "A Postscript to 'The Department Awards a Certificate of Excellence.'" *Association of Departments of English Bulletin*, No. 25 (May 1970), 30.

Coffman, William E. "Determining Students' Concepts of Effective Teaching from Their Ratings of Instructors." *Journal of Educational Psychology*, 45 (1954), 277-86.

Cogan, Morris. "Research on the Behavior of Teachers: A New Phase." *Journal of Teacher Education*, 14 (1963), 238-43.

Cohen, Arthur, and Florence Brawer. *Measuring Faculty Performance*. Washington, D.C.: American Association of Junior Colleges, 1969.

Daugherty, Hope. "Appraising the College Teacher." *Improving College and University Teaching*. 16 (1968), 203-06.

Dressel, Paul, et al. *Evaluation in Higher Education*. Boston: Houghton Mifflin, 1961.

Eastman, Arthur. "Evaluation as Helping." Unpublished address delivered at the meeting of the Modern Language Association (Dec 1970).

——"How Visitation Came to Carnegie-Mellon University," *Association of Departments of English Bulletin*, No. 21 (May 1969), 10-20.

Ebel, Robert. *Measuring Educational Achievement*. Englewood Cliffs, N.J.: Prentice-Hall, 1965.

——,ed. *Encyclopedia of Educational Research*. New York: Macmillan, 1969.

Eble, Kenneth. *The Recognition and Evaluation of Teaching*. Salt Lake City: Project to Improve College Teaching, 1970

Elbow, Peter. "The Definition of Teaching." *College English*, 30 (1968), 187-201.

Estrin, Herman, and Ruth Goodwin. "Evaluating College Teaching." *Improving College and University Teaching*, 10 (1962), 193-96.

Farrell, Edmund, and Leo Ruth. "The Scene: The Coming of the Wasteland in Education." *English Journal*, 59 (1970), 713-20.

Flanagan, J. C. "The Critical Incident Technique." *Psychological Bulletin*, 51 (1954), 327-58.

Freedman, Morris. *Chaos in Our Colleges*. New York: McKay. 1963.

Gage, N. L. "The Appraisal of College Teaching: An Analysis of Ends and Means." *Journal of Higher Education*, 32 (1961), 17-22.

——*Handbook of Research on Teaching*. Chicago: Rand McNally, 1963.

Gerber, John C. "Response to Austerity." *Association of Department. of English Bulletin*, No. 25 (May 1970), 49-52.

Gray, Charles S. "The Teaching Model and Evaluation of Teaching Performance." *Journal of Higher Education*, 40 (1969), 636-42.

Grommon, Alfred, ed. *The Education of Teachers of English for American Schools and Colleges*. New York: Appleton-Century-Crofts, 1963.

Gustad, John W. *Policies and Practices in Faculty Evaluation*. Washington, D.C.: American Council on Education, 1961.

Guth, Hans. "The Monkey on the Bicycle: Behavioral Objectives in the Teaching of English." *English Journal*, 59 (1970), 785-92.

Harper, George M. "'The Waste, Sad Time': Some Remarks on Class Visitation." *College English*, 27 (1965), 119-23.

Hazen Foundation. *The Importance of Teaching*. New Haven, Conn.: Hazen Foundation, n. d.

Henry, George. "Style of Teaching and Teacher Evaluation." *English Journal*, 59 (1970), 921-27.

Hoyt, Donald P. *Improving Instruction through Student Feedback*. Manhattan: Kansas State Univ., 1969.

—*Instructional Effectiveness: Interrelationships with Publication Record and Monetary Reward.* Kansas State Univ. Research Report No. 10. Manhattan: Kansas State Univ., 1970.

—*Instructional Effectiveness: Measurement of Effectiveness.* Kansas State Univ. Research Report No. 6. Manhattan: Kansas State Univ., 1969.

Hunter, John O. "Faculty Evaluation as a Liberal Persuasion." *Improving College and University Teaching*, 17 (1969), 90-92.

Hutchisson, Elmer. "Measuring Teaching Performance," *Improving College and University Teaching*, 5 (1957), 42-45.

Koen, Frank. "The Training of Graduate Teaching Assistants." *Educational Record*, 49 (1968), 97-104.

Langen, Thomas D. F. "Student Assessment of Teaching Effectiveness." *Improving College and University Teaching*, 14 (1966), 22-25.

Larson, Richard L. "Observing and Guiding the Work of New Teachers." *College English*, 28 (1967), 617-21.

Lee, Calvin B. T., ed. *Improving College Teaching.* Washington, D.C.: American Council on Education, 1967.

Lewis, Gertrude. *The Evaluation of Teaching.* Washington, D.C.: National Education Association, 1966.

Lovell, George D., and Charles F. Haner. "Forced Choice Applied to College Faculty Rating." *Educational and Psychological Measurement*, 15 (1955), 291-304.

Lundstedt, Sven. "Criteria for Effective Teaching." *Improving College and University Teaching*, 14 (1966), 27-31.

Marshall, Max. "Is Teaching?" *Improving College and University Teaching*, 17 (1969), 93-96.

Maxwell, John, and Anthony Tovatt. *On Writing Behavioral Objectives for English.* Champaign, Ill.: National Council of Teachers of English, 1970.

McGrath, Earl J. "Characteristics of Outstanding College Teachers." *Journal of Higher Education*, 33 (1962), 148-52.

McKeachie, Wilbert J. "Student Ratings of Faculty." *AAUP Bulletin*, 55 (1969), 439-44.

Meredith, Gerald. "Dimensions of Faculty-Course Evaluation." *Journal of Psychology*, 73 (1969), 27-32.

Morris, William, ed. *Effective College Teaching.* Washington, D.C.: American Council on Education, 1970.

Morton, Richard K. "Evaluating College Teaching." *Improving College and University Teaching*, 9 (1961), 122-23.

Perry, Richard R. "Institutional Research." *College and University Business*, 47 (Oct 1969), 18, 22.

Popham, W. James. "The Performance Test: A New Approach to the Assessment of Teaching Proficiency." *Journal of Teacher Education*, 19 (1968), 216-22.

Purves, Alan. "Of Behaviors, Objectives, and English." *English Journal*, 59 (1970), 793-97.

Ryans, David. *Characteristics of Teachers.* Washington, D.C.: American Council on Education, 1960.

Sanford, Nevitt, ed. *The American College.* New York: Wiley, 1962.

Select Committee on Education. *Education at Berkeley.* Berkeley: Univ. of California Press, 1968.

Soar, Robert. *Multivariate Statistical Procedures in Predicting Teacher-Pupil Classroom Behavior.* Columbia: Univ. of South Carolina, 1962.

Soloman, D., L. Rosenberg, and W. E. Bezdek. "Teacher Behavior and Student Learning." *Journal of Educational Psychology*, 55 (1964), 23-30.

Stewart, J., and K. Malpuss. "Estimates of Achievement and Ratings of Instructors." *Journal of Educational Research*, 59 (1966), 341-50.

Tead, Ordway. "Teacher Self-Evaluation." *Improving College and University Teaching*, 9 (1961), 148.

Tuckman, B. W., and W. F. Oliver. "Effectiveness of Feedback to Teachers as a Function of Source." *Journal of Educational Psychology*, 59 (1968), 297-301.

Yamamoto, Kaoru. "Evaluating Teacher Effectiveness: A Report of Research." *Journal of School Psychology*, 2 (1963-64), 60-71.

APPENDIX A

SAMPLE FORMS FOR STUDENT RATING OF FACULTY

UNIVERSITY OF WASHINGTON Instructor _____
SURVEY OF STUDENT OPINION Course _____ sect _____ date _____
OF TEACHING

STUDENT CHARACTERISTICS

Major
— This department
— Social science
— Natural science
— Humanities
— Other (incl pre)

Class
— Freshman
— Sophomore
— Junior
— Senior
— Graduate
— other

Age
— 18 or under
— 19-20
— 21-22
— 23-24
— over 24

Cum GPA
— Below 2.5
— 2.5-2.9
— 3.0-3.4
— Above 3.4

Sex
— Female
— Male

Course
— Required
— Elective

Grade expected
— A
— B
— C
— D
— Fail or E
— Pass

USE PENCIL ONLY. Mark one space for each item. Please do not make stray marks on this answer sheet. Additional comments may be made on the back of this sheet. Many instructors make the statistical report of these surveys available to students in the department and to their chairman and dean.

A-1

	Of less value	Only fair	Competent	Superior	Outstanding
1. Abstract ideas and theories were clearly interpreted	—	—	—	—	—
2. Takes an active, personal interest in the class	—	—	—	—	—
3. My skills in thinking were increased	—	—	—	—	—
4. Helped broaden my interests	—	—	—	—	—
5. Stressed important material	—	—	—	—	—
6. Made good use of examples and illustrations	—	—	—	—	—
7. Motivated me to do my best	—	—	—	—	—
8. Inspired class confidence in his knowledge of subject	—	—	—	—	—
9. Gave me new viewpoints or appreciations	—	—	—	—	—
10. Clear and understandable in explanations	—	—	—	—	—
11. Lectures gave viewpoints and info text did not contain	—	—	—	—	—
12. Material enthusiastically presented in lectures	—	—	—	—	—
13. Material presented in a well-organized fashion	—	—	—	—	—
14. Helpful to individual students	—	—	—	—	—
15. Integration of material into coherent whole was	—	—	—	—	—
16. Text clear in presentation of concepts	—	—	—	—	—
17. Text's overall rating	—	—	—	—	—

18. How much was your interest in the subject changed by this course? More interested — — — — — Less interested

19. What level of student sophistication was assumed in lectures? Very high — — — — — Very low

20. Were students free to ask questions, disagree, express their ideas, etc.? Encouraged — — — — — Discouraged

21. Has improved my problem-solving methods Very much — — — — — Not at all

22. Did test questions cover the material emphasized in the text and lectures? Very well — — — — — Very poorly

23. Would you recommend this course by this instructor to majors in this dept? Very highly — — — — — Never

24. Would you recommend this course by this instructor to non-majors? Very highly ___ ___ ___ ___ ___ Never

25. (25.) _____ ___ ___ ___ ___ _____

QUESTIONS 25-27 WILL BE ON THE CHALKBOARD IF THE INSTRUCTOR WANTS TO USE THEM.

26. (26.) _____ ___ ___ ___ ___ _____

27. (27.) _____ ___ ___ ___ ___ _____

Your instructor would like to know if there is something you believe he has done especially well in his teaching of this course _____.

_____

_____

_____

_____

_____

_____

Your instructor would also like to know what specific things you believe might be done to improve his teaching of this course _____

_____

_____

_____

_____

_____

_____

Thank you.

A-2

SURVEY OF STUDENT OPINION OF TEACHING

NAME OF INSTRUCTOR | COURSE AND NUMBER | CREDIT

YOUR FIELD | YEAR IN COLLEGE | YOUR SIGNATURE

The main task of the college is teaching. It is of first importance that the college be continuously informed of the quality of its teaching and the respects in which that teaching can be improved. Students are in a position to judge the quality of teaching from direct experience.

You are being asked to indicate your opinion of this instructor, with whom you have studied, in regard to five characteristics which students have indicated are important to them.

Read the description of each quality listed below. Draw a circle around the number that indicates your rating of this instructor compared with other instructors you have had.

Circle
1. if this instructor is the best you have had
2. if he is one of the best you have had (but not *the* best)
3. if he is above average but not one of the best you have had
4. if he is generally average compared with others you have had
5. if he is below average compared with others you have had
6. if he is one of the poorest you have had

Do this for each of the five qualities, making each answer a separate judgment. Obviously only in extremely rare cases will the circled number be the same for all qualities.

	High					Low
1. Makes learning active for you, as by stimulating thinking, encouraging participation, guiding discussion	1	2	3	4	5	6
2. Knows subject thoroughly enough to organize material and relate it to other fields, integrate, answer questions	1	2	3	4	5	6
3. Displays an active, personal interest in you *as* by being easy to approach, patient, willing to help	1	2	3	4	5	6
4. Presents what he has to say clearly, at your level of understanding	1	2	3	4	5	6
5. Gets you interested in his subject	1	2	3	4	5	6

Write in your own words your general comment on his teaching. (Use the back of the sheet also, if you wish.)

FORM A

A-3

STUDENT APPRAISAL OF COURSE AND INSTRUCTOR

Professor's name _____

Course number and short title _____.

Quarter _____

Student's major _____

Class standing (Fr., Soph., Jr., Sr., Grad.) _____

Grade point average _____

This course appraisal form is intended to help your professor assess his teaching and his courses. At the present time, he is neither asked nor required to show the results to anyone else. (At a later date, the results of surveys like this one may become part of an overall evaluation of the teaching effectiveness of the Department.)

We invite your candor in two ways: (1) you do not sign these forms; (2) the completed forms will not be made available to the professor until after the deadline for the submission of grades for this quarter.

There are three parts to this form. The first two are conventional and familiar; the third is unconventional and unfamiliar. We hope you will answer all parts with equal care.

PART ONE

1. Did the course fulfill your expectations? __ Yes; __ No. If *No*, in what respect(s)? _____

2. Has the course stimulated you to continue study in the same or a related field?
   __ Yes; __ No. Comment: _____

3. Check the teaching method which you feel predominated in this course:
   __ Lecture __ Discussion __ Seminar __ Other (specify): _____

   Did you feel the dominant method appropriate to the instructor's skills?
   __ Yes; __ No. To the course subject matter __ Yes; __ No.

4. To what extent did the text(s) command your attention? __ Considerably
   __ Moderately __ Very little

5. In relation to other college courses, this course was: __ One of the best __ Average
   __ Below average

PART TWO

Listed below are several qualities which help to define and describe an instructor's task in almost any course. Rate your instructor on each of these items by encircling the number that best indicates his position in comparison with other instructors you have had.

	Outstanding	Superior	Adequate	Only fair	Of little value
1. Gets me interested in the subject	1	2	3	4	5
2. Inspires my confidence in his knowledge of the subject	1	2	3	4	5
3. Interprets abstract ideas and theories clearly	1	2	3	4	5
4. Clarifies the objectives of the course	1	2	3	4	5
5. Places proper emphasis on major points	1	2	3	4	5
6. Provides me with new viewpoints	1	2	3	4	5
7. Motivates me to think for myself	1	2	3	4	5
8. Composes clear writing and test assignments	1	2	3	4	5
9. Grades thoughtfully and fairly	1	2	3	4	5
10. Is available for help and advice	1	2	3	4	5

11. Your instructor would like to know if there is something you believe he has done especially well in this course: _____

_____

_____

_____

12. Your instructor would like to know what specific thing(s) you believe he might do to improve his teaching of this course: _____

_____

_____

PART THREE

Forms like the one you have just filled out have a limited use and a limited appeal. This section is an attempt to extend the usefulness and enhance the appeal of this questionnaire. Here you are invited to make some responses that are, perhaps, like the ones you would write in the margins if you were given a chance. Moreover, the special purpose of PART THREE is to get at some of those qualities which are peculiarly appropriate to an English instructor.

Use the following code to answer the questions below: *O* (Often); *S* (Sometimes); *N* (Never); *?* (Don't know)

Circle the extent to which you feel the instructor:

1. Makes learning in this class a pleasure, even though you might be hard put to it to explain just how he does it	O	S	N	?
2. Displays in his own use of language some of the qualities which we admire in the language of literature	O	S	N	?
3. Is capable of an insight which is really astonishing	O	S	N	?
4. Is capable of "thinking on his feet," not invariably delivering preconceived ideas	O	S	N	?
5. Can catch fire and depart from his preconceived plan to exploit an unforeseen opportunity in a way that makes learning a delight	O	S	N	?
6. Conveys his own opinions concerning controverted issues without being either too slack or too dogmatic	O	S	N	?
7. Is not afraid to depart from the more "popular" aspects of his subject to explore difficult, "dry," but necessary aspects	O	S	N	?
8. Demonstrates real ingenuity in finding ways to make difficult aspects of his subject easier to grasp	O	S	N	?
9. Appears to assume that it is part of a teacher's responsibility to try to make his subject interesting	O	S	N	?
10. Seems able to elicit responses from the less able students without being either patronizing or sentimental	O	S	N	?
11. Demonstrates an awareness of his own fallibility	O	S	N	?
12. Makes reasonable allowances for the customary frailties of students	O	S	N	?
13. Seems genuinely capable of recognizing and responding to an idea originating with a student	O	S	N	?
14. Appears reasonably aware of his own strengths and limitations and does not try to be all things to all men	O	S	N	?
15. Displays some of his own idiosyncrasies without converting these into tiresome mannerisms	O	S	N	?
16. Seems capable of handling those occasionally awkward or embarrassing situations which can crop up in any class	O	S	N	?

FORM B

STUDENT'S EVALUATION

Course: _____

Quarter: _____

Instructor: _____

This course appraisal form is intended to help your professor assess his teaching and his courses. At the present time, he is neither asked nor required to show the results to anyone else. (At a later date, the results of surveys like this one may become part of an overall evaluation of the teaching effectiveness of the Department.)

We invite your candor in two ways: (1) you do not sign these forms; (2) the completed forms will not be made available to the professor until after the deadline for the submission of grades for this quarter.

(Note: In answering Questions 1 and 2, please draw a circle around the number which most closely reflects your rating on the scale where 1 is the lowest and 10 is the highest rating.)

1. I rate my *degree of interest* in the subject of this course as it was presented:

Low 1 2 3 4 5 6 7 8 9 10 High

2. I rate the *value received* from this course as follows:

Low 1 2 3 4 5 6 7 8 9 10 High

3. I received the following *benefits* from this course:

4. The course's *weaknesses* seemed to be:

5. Suggested *improvements* would be:

6. Additional comments:

A-4

COURSE EVALUATION

This course evaluation questionnaire has been designed to benefit students and faculty. The responses provided by you and other students will be analyzed by the _____. Please follow these directions:

1. You are being provided with both a questionnaire booklet and an answer sheet. Please respond to all questions which call for a rating by indicating your answer, *in pencil*, on the answer sheet. All written comments, however, should appear in the spaces provided in the questionnaire booklet. *Do not write your name on either the booklet or the answer sheet.*

2. First, turn the answer sheet to a horizontal position and indicate the department number and course number for this course. A list of department numbers can be found on the back page of this booklet. For example, Religion 204 would have department number 72 and course number 204. After printing each department and course number at the top of the appropriate column, blacken the smaller spaces below which correspond to that number. Be sure to fill in each space completely. Use No. 2 pencils only.

3. Indicate your Class and the answer sheet serial number.

4. Returning the answer sheet to the vertical position, please respond *only* to the sections of the questionnaire that are applicable to this course. Since this same questionnaire is being used for all courses, some sections will be applicable to this course while others will not. Most items involve a mark on the answer sheet, others a brief comment in the booklet itself.

5. If you wish to change a response on the answer sheet, be sure to *erase* the incorrect mark. Do not cross out responses you want deleted. Please remember to use a No. 2 pencil on the answer sheet and to darken each answer space completely.

Your frank and thoughtful answers to these questions will be appreciated.

PART 1: LECTURES

1. Applicable ___ Inapplicable ___

		Excellent 1		*Good* 2	3	*Fair* 4	5	*Poor* 6	*Not Applicable* 0
2-11.	Rate the quality of the *lectures* in terms of the degree to which they:								
	2. Held your attention and interest	1		2	3	4	5	6	0
	3. Covered the material at an appropriate intellectual level — neither too complicated nor too simple	1		2	3	4	5	6	0
	4. Clearly presented the relevant subject matter	1		2	3	4	5	6	0
	5. Covered diverse points of view and helped to expand your awareness of alternatives	1		2	3	4	5	6	0
	6. Emphasized principles and generalizations	1		2	3	4	5	6	0
	7. Stimulated your intellectual curiosity and provoked independent thinking	1		2	3	4	5	6	0
	8. Stimulated student discussion outside of class	1		2	3	4	5	6	0
	9. Were related to one another and followed a coherent sequence	1		2	3	4	5	6	0
	10. Were coordinated with other parts of the course	1		2	3	4	5	6	0
	11. Rate the general quality of the lectures as a whole	1		2	3	4	5	6	
12.	In comparison to all other lecturers you have had, how would you rate the lecturer in this course?	1		2	3	4	5	6	
13.	About how many lectures did you miss?	1) 1 or none 2) 2-4					3) 5-7 4) 8 or more		

14-22. Often some very concrete characteristics of the lecturer reduce the value of a course or interfere with the achievement of its objectives. For each of the following characteristics indicate whether or not you felt that it applied to the lecturer:

	YES	NO		YES	NO
14. Inaudible	___	___	19. Speaks too fast	___	___
15. Writing illegible	___	___	20. Disorganized presentations	___	___
16. Too abstract	___	___	21. Repetitious	___	___
17. Too superficial	___	___	22. Distracting mannerisms	___	___
18. Covers too much material	___	___	(Specify: _____ _____)		

Which lectures (or lectures on what subjects) did you find most valuable? Why?

Which lectures (or lectures on what subjects) did you find least valuable? Why?

PART 2: READINGS

		Excellent 1	2	Good 3	4	Fair 5	Poor 6	Not Applicable 0
23-30.	Speaking generally, rate the quality of the readings in terms of the degree to which they:							
23.	Were interesting to read	1	2	3	4	5	6	0
24.	Were of the right level of difficulty — neither too complicated nor too simple	1	2	3	4	5	6	0
25.	Clearly presented the relevant subject matter	1	2	3	4	5	6	0
26.	Covered diverse points of view and helped to expand your awareness of alternatives	1	2	3	4	5	6	0
27.	Helped you to integrate facts and develop generalizations	1	2	3	4	5	6	0
28.	Stimulated your intellectual curiosity and provoked independent thinking	1	2	3	4	5	6	0
29.	Balanced one another and formed a coherent whole	1	2	3	4	5	6	0
30.	Were relevant and supplementary to the lectures	1	2	3	4	5	6	0
31.	Rate the general quality of the readings as a whole	1	2	3	4	5	6	
32.	About what proportion of the readings did you do before reading period?	1) All 2) ¾			3) ½ 4) ¼			
33.	About what proportion of the readings did you end up doing?	1) All 2) ¾			3) ½ 4) ¼			

Which readings did you particularly find to be of *most* value? Why?

Which readings did you particularly find to be of *least* value? Why?

PART 3: PRECEPTS OR CLASSES

34. Applicable ___ Inapplicable ___

		Excellent 1	2	Good 3	4	Fair 5	Poor 6
35-44.	Rate the general quality of the preceptor classes on each of the following. (The word instructor is used to refer to the person responsible for your section regardless of whether it was a class or a precept.)						
35.	Interest of the instructor in the precept (class)	1	2	3	4	5	6
36.	Instructor's ability to raise challenging questions	1	2	3	4	5	6
37.	Instructor's ability to help clarify readings and lectures	1	2	3	4	5	6
38.	Instructor's ability to encourage broad student participation	1	2	3	4	5	6
39.	Instructor's ability to conduct discussions	1	2	3	4	5	6
40.	Instructor's responsiveness to students' comments and questions	1	2	3	4	5	6
41.	Integration with other parts of the course — relevant and supplementary	1	2	3	4	5	6
42.	General attitude and preparedness of fellow class members	1	2	3	4	5	6
43.	Your own interest, preparation and participation	1	2	3	4	5	6
44.	Value of the preceptor classes as a whole to this course	1	2	3	4	5	6
45.	How many preceptor classes did you miss?	1) 1 or none 2) 2-3			3) 4-5 4) 6 or more		

Were there characteristics of the preceptors (classes) which you found particularly valuable or not very useful? Please explain and, if possible, suggest modifications.

Name of Instructor: _____

PART 4: PAPERS, REPORTS, PROBLEM SETS

46. Applicable ___ Inapplicable ___

How would you rate each of the following aspects of the papers, reports or problem sets in this course?

		Excellent 1	2	*Good* 3	*Fair* 4	5	*Poor* 6
47-49.	Appropriateness of level of demand in terms of:						
47.	Number of papers, reports or problem sets	1	2	3	4	5	6
48.	Size or length of paper(s), report, or problem sets	1	2	3	4	5	6
49.	Difficulty of subject(s)	1	2	3	4	5	6
50.	Degree of guidance given by the instructor to choice of topics and suggestions for relevant research	1	2	3	4	5	6
51.	Freedom given to develop topics in a creative, imaginative way	1	2	3	4	5	6
52.	Relevance to the rest of the course	1	2	3	4	5	6
53.	Timing of papers, reports or problem sets to occur at the most beneficial time in the course	1	2	3	4	5	6
54.	Overall value of the papers, reports or problem sets to this course	1	2	3	4	5	6

Please explain ratings of *Poor* given above and add any further relevant comments you might wish to make.

PART 5: LABORATORY

55. Applicable ___ Inapplicable ___

		Excellent 1	2	*Good* 3	*Fair* 4	5	*Poor* 6
56-60.	Rate the general quality of the laboratories in terms of the degree to which they:						
56.	Were interesting and stimulating to you	1	2	3	4	5	6
57.	Contained the right amount of structure and guidance by the instructor	1	2	3	4	5	6
58.	Gave you an opportunity for creative and imaginative work	1	2	3	4	5	6
59.	Were useful to you as a supplement to the lectures and reading – added significantly to the course	1	2	3	4	5	6
60.	Contained adequate equipment	1	2	3	4	5	6
61.	Rate your own general interest in and enthusiasm for experiments or experimental procedures	1	2	3	4	5	6
62.	Rate the quality of the laboratories as a whole	1	2	3	4	5	6
63.	How many lab sessions did you miss?		1) None 2) 1-2		3) 3-4 4) 5 or more		

Are there any characteristics of the labs which you found particularly valuable or not very useful? Please explain and, if possible, suggest modifications.

of Lab Instructor: ___

PART 6: SEMINARS

64. Applicable ___ Inapplicable ___

		Excellent 1	Good 2	3	Fair 4	5	Poor 6
65-75.	Please rate the general quality of the seminars on each of the following:						
65.	Interest to you of the topics covered	1	2	3	4	5	6
66.	Professor's ability to raise stimulating, provocative questions	1	2	3	4	5	6
67.	Professor's ability to encourage broad student participation	1	2	3	4	5	6
68.	Professor's ability to conduct discussions	1	2	3	4	5	6
69.	Degree to which the topics covered were related to one another	1	2	3	4	5	6
70.	Degree to which the topics emphasized the fundamentals of the course	1	2	3	4	5	6
71.	Degree to which you felt a sense of challenge, insight and discovery	1	2	3	4	5	6
72.	Degree to which you felt a part of a continuing scholarly discussion	1	2	3	4	5	6
73.	General attitude and preparedness of fellow class members	1	2	3	4	5	6
74.	Your own interest, preparation and participation	1	2	3	4	5	6
75.	Quality of the seminars as a whole	1	2	3	4	5	6

76. How many seminars did you miss? 1) None 2) 1-2 3) 3-4 4) 5 or more

Were there characteristics of the seminars which you found particularly valuable or not very useful? Please explain and, if possible, suggest modifications.

PART 7: LANGUAGE DRILLS

77. Applicable ___ Inapplicable ___

		Excellent 1	Good 2	3	Fair 4	5	Poor 6
78-81.	Please rate the general quality of the language drills in terms of the degree to which you found that they:						
78.	Were helpful to you with your grammar	1	2	3	4	5	6
79.	Helped you to understand the language	1	2	3	4	5	6
80.	Helped you to speak the language	1	2	3	4	5	6
81.	Helped you to write the language	1	2	3	4	5	6
82-87.	Rate the utility of each of the following in assisting you to master the language.						
82.	Classes	1	2	3	4	5	6
83.	Drills	1	2	3	4	5	6
84.	Textbook	1	2	3	4	5	6
85.	Tests	1	2	3	4	5	6
86.	Laboratories	1	2	3	4	5	6
87.	Course as a whole	1	2	3	4	5	6

88. Is this course a requirement either for your degree or for your admission to graduate school? Yes ___ No ___

89. Regardless of whether it is a requirement, did you generally find the course to be pleasurable and/or interesting? Yes ___ No ___

90. Taking the course as a whole, including both classes and drills, do you think that you learned enough to make the time spent on this course worthwhile? Did the results match the effort? Yes ___ No ___

Do you think that the methods used in this course were generally well designed to achieve the goals of the course? Are there other methods or other goals you might suggest? (Please be specific)

PART 8: OVERALL RATINGS
(Applicable To All Courses)

In relation to your own objectives in this course and compared to other courses you have taken at Princeton, how would you rate this course in terms of how much it contributed to each of the following:

	Excellent 1	2	*Good* 3	4	*Fair* 5	*Poor* 6	*Not Applicable* 0
91. Your mastery of the relevant content or subject matter	1	2	3	4	5	6	0
92. Your mastery of the relevant skills or methods	1	2	3	4	5	6	0
93. Your ability to see more alternatives and have more insight into the complexity of the relevant subject matter	1	2	3	4	5	6	0
94. Your ability to formulate general principles in the relevant subject matter	1	2	3	4	5	6	0
95. Your abilities for critical evaluation in the relevant subject matter	1	2	3	4	5	6	0
96. An increase in your interest in the field so as to take further related courses or do reading on your own	1	2	3	4	5	6	0
97. An impact upon your emotional sensitivity to the relevant phenomena or upon your values and attitudes toward parts of life or your self	1	2	3	4	5	6	0
98. Your total educational growth and development	1	2	3	4	5	6	0

PART 9: REASONS FOR TAKING COURSE
(Applicable To All Courses)

99-109. Check whether each of the following was a reason for your taking this course.

		Yes	No
99.	Reputation of the instructor	—	—
100.	Recommended by faculty advisor	—	—
101.	Recommended by fellow students	—	—
102.	Desire to learn about an area you previously knew little about	—	—
103.	Desire to pursue an area already of interest to you	—	—
104.	Related to career interests	—	—
105.	Thought course would be easy in terms of amount of work	—	—
106.	Thought course would be easy in terms of a grade	—	—
107.	To fill a field requirement	—	—
108.	Meets at a convenient time for your schedule	—	—
109.	Other (specify)		

Please write in below any generally descriptive comments you would like to make about the course or specific points which were not brought out in the questionnaire.

A-5

This and forms A-6, A-7, and A-8 are from the same campus. Forms A-5, A-6, and A-7 are sponsored by faculty and administration; the form in A-8 is student-sponsored.

STUDENT OPINION OF COURSES AND TEACHING

TO THE STUDENT:

The act of evaluating the educational process is not a simple one for either the teacher or the student. However, the faculty has found that both teacher and student benefit from the careful and honest opinions given by our students. It is, therefore, the policy of the College to conduct this inventory of course objectives and teaching procedures every other year. Your thoughtful responses to this questionnaire will assist the College in improving the methods and objectives of our common educational endeavors.

FILL IN:

Department and Course Number _____

Section Number _____

Name of Teacher _____

School or College in which Enrolled _____

Class (Circle one): Fresh., Soph., Junior, Senior, Grad., Special

Field of Major interest or Concentration _____

Overall Grade Point Average _____

PLEASE DO NOT SIGN YOUR NAME

THIS FORM WILL NOT BE RETURNED TO THE INSTRUCTOR UNTIL AFTER GRADES HAVE BEEN REPORTED.

TO SAVE YOU TIME, READ THE INSTRUCTIONS FOR ALL SECTIONS of this form BEFORE you begin to answer any one. This will help you avoid unnecessary or inappropriate answers.

1. What do you think are the objectives of this course AS EMPHASIZED BY THE INSTRUCTOR? Here is a list of statements which can be used to identify this emphasis as given in most of our College courses. First read through the entire list and then underline as many phrases as you believe represent the main emphasis of this course. Use a double underline for the one, two, or three statements that are especially applicable.

   a) Learning new terminology or vocabulary

   b) Acquiring specific and factual information

   c) Learning rules, procedures, techniques, or methodology

   d) Learning concepts, principles, or theories

   e) Applying facts, procedures, principles, or other knowledge and skills

   f) Analytic or critical thinking; that is, learning to analyze or make evaluative judgments about data, ideas, arguments, or theories

   g) Creative thinking; that is, learning to combine facts, ideas, and procedures, or produce original material

   h) Changing or developing your interests in this field

   i) Changing or developing your attitudes or values

2. Are you satisfied with these course objectives; if not, how would you wish them to be changed?

3. Summarize briefly one of the more specific ways that this course has influenced or changed your interests, attitudes, or values.

4. Ten attributes of instruction have been listed below. For each attribute, circle the word or phrase which is nearest to your impression of this course, i.e., which best describes that aspect of the course FOR YOU. Where appropriate, give reasons or examples to support your opinion. Not all of the attributes apply equally to each course in the College, so you may wish to make some qualifying comments in the space near each item.

   a) The use of class time was:

      very effective satisfactory unsatisfactory at times

   b) The pace of classroom presentation of material, for the most part, was:

      too slow too fast about right

   c) Individual help or further discussion outside of class:

      was encouraged by instructor — was normally available — should have been more available

   d) The integration of lectures with other course material was:

      somewhat lacking good excellent

   e) The assigned material was on the whole:

      too difficult fine for me too easy

   f) In my opinion the class procedure was:

      well organized moderately well organized poorly planned

   g) The instructor stimulated my interest in the subject matter:

      a great deal somewhat very little

h) The instructor's enthusiasm for the subject matter was:

strong and sincere | adequate | somewhat lacking

i) The feeling between the instructor and the student was:

somewhat antagonistic | cordial | especially close and friendly

j) The instructor's description, explanation, or analysis of the subject matter was:

seldom clear | sometimes clear | consistently clear

5. Keeping in mind that the returns from this questionnaire will be used by the instructor in the process of improving his teaching, please mention ANY OTHER ASPECTS OF THE COURSE OR INSTRUCTOR not covered in previous questions which you consider to be especially good or poor. For example, consider any of the following list which are relevant:

text and outside readings	papers, projects, and examinations
lectures	course procedure
recitation or discussion	instructor
laboratory	

Offer any suggestions that you have for improving this course.

A-6

Indicate your evaluation of characteristics below, using numbers based on the following scale:

1. Superior 2. Very Good 3. Good 4. Fair 5. Poor

A. Rate the instructor's general teaching effectiveness _____

B. Rate the value of the course as a whole _____

Course _____ Instructor _____ Term _____ 19_____

GPA _____

COMMENTS BELOW:

A-7

STUDENT OPINION OF TEACHING AND COURSE

CHARACTERISTICS OF THE *TEACHER* AND THE *TEACHING*

Each of the items below deals with a characteristic of instructors which students feel to be important. Indicate your rating of your instructor by a check at the appropriate point on the scale. The exact point at which you rate is less important than the general impression.

Write in after the question any additional comments that you wish to make. Give examples wherever possible.

1. Is he actively helpful when students have difficulty?

NOT HELPFUL ACTIVELY HELPFUL

Example or Comments:

2. Does he appear sensitive to students' feelings and problems?

UNAWARE RESPONSIVE

Example or Comments:

3. Was he flexible?

RIGID FLEXIBLE

Example or Comments:

4. Does he make students feel free to ask questions, disagree, express their ideas, etc.?

INTOLERANT ENCOURAGES STUDENT IDEAS

Example or Comments:

5. Is he fair and impartial in his dealings with the students?

FAVORS SOME FAIR

Example or Comments:

6. Is his speech adequate for teaching?

UNINTELLIGIBLE — GOOD

Example or Comments (Volume, Tone, Enunciation, Rate, Vocabulary, etc.):

7. Does he belittle students?

BELITTLES — RESPECTS

Example or Comments:

8. Does he tell students when they have done particularly well?

NEVER — ALWAYS

Example or Comments:

9. Does he dwell upon the obvious?

DWELLS ON OBVIOUS — INTRODUCES INTERESTING IDEAS

Example or Comments:

10. Is he interested in the subject?

SEEMS UNINTERESTED — SEEMS INTERESTED

Example or Comments:

11. Does he use enough examples or illustrations to clarify the material?

NONE — MANY

Example or Comments:

12. Does he present material in a well-organized fashion?

DISORGANIZED — WELL-ORGANIZED

Example or Comments:

13. Did he follow an outline?

NOT AT ALL _____ VERY CLOSELY

Example or Comments:

14. Does he stimulate thinking?

DULL _____ STIMULATING

Example or Comments:

15. Does he put his material across in an interesting way?

DULL _____ VERY INTERESTING

Example or Comments:

16. Other important characteristics – Please specify:

_____.

Considering all of the above qualities which are applicable (including those that you added), how would you rate this teacher? (Circle your rating.)

EXCELLENT VERY GOOD GOOD FAIR POOR VERY BAD

Now go back over the list and place a check (X) before the five items which were most important to you in making your judgment.

1. Are the objectives of the course clear?

UNCLEAR _____ CLEAR

Example or Comments:

2. Is the amount of work required appropriate for the credit received?

TOO MUCH _____ TOO LITTLE

Example or Comments:

3. Was the assigned reading difficult?

TOO EASY TOO DIFFICULT

Example or Comments:

4. Are the tests fair?

UNFAIR FAIR

Example or Comments:

5. Are the grades assigned fairly?

UNFAIR FAIR

Example or Comments:

6. How would you rate the contribution of the textbook to the course?

POOR EXCELLENT

Example or Comments:

Considering all of the above qualities which are applicable (including others that you added), how would you rate this course? (Circle your rating.)

EXCELLENT VERY GOOD GOOD FAIR POOR VERY BAD

1, 2, 3, 4, 5, 6, Items loading on corresponding factors of Isaacson, et al. (1964).

If you have any additional comments to make about the course or the teacher, please make them at the bottom of this page.

A-8

This form is from the same campus as A-5.
Association for Academic Evaluation

Instructor's Name_____

STUDENT COURSE EVALUATION QUESTIONNAIRE

INSTRUCTIONS: Put all answers on this sheet, using No. 2 *pencil* ONLY.
In some cases, the question is NOT APPLICABLE; so indicate.
Please fill in Course Title and Instructor's Name above. Do not sign this questionnaire.

For the first three sections, give your reaction to the statement, using:

a) STRONGLY AGREE b) AGREE c) NEUTRAL d) DISAGREE e) STRONGLY DISAGREE

COURSE	a	b	c	d	e
1. This course has been challenging in the sense of demanding concentration and intelligence.	—	—	—	—	—
2. This course was stimulating, requiring independent and creative thinking.	—	—	—	—	—
3. Too much material was presented to be adequately covered in the time allotted.	—	—	—	—	—
4. More credit should be given for the amount of time spent in this course.	—	—	—	—	—
5. The major objectives set for the course were adequately achieved.	—	—	—	—	—
6. There was unnecessary repetition in the lectures, recitations and readings.	—	—	—	—	—
7. This course made a significant contribution to my education.	—	—	—	—	—
INSTRUCTOR	**a**	**b**	**c**	**d**	**e**
8. The instructor was enthusiastic and interested in his subject.	—	—	—	—	—
9. The instructor was well organized.	—	—	—	—	—
10. The instructor had poor speaking ability.	—	—	—	—	—
11. The instructor gave clear explanations.	—	—	—	—	—
12. The instructor should use a greater variety of classroom techniques.	—	—	—	—	—
13. The instructor shared the values and ideals of the field with his students.	—	—	—	—	—
14. The instructor produced an appreciation of scientific and scholarly research.	—	—	—	—	—
15. The instructor presented mainly a one-sided view of the field.	—	—	—	—	—
16. The instructor was sensitive to the level of student comprehension.	—	—	—	—	—
17. The instructor got to know, and be known by, his students.	—	—	—	—	—
18. The instructor conveyed general perspectives as well as specific facts.	—	—	—	—	—
ASSIGNMENTS/EXAMS	**a**	**b**	**c**	**d**	**e**
19. Standards for student performance were reasonable.	—	—	—	—	—
20. The grading system was fair.	—	—	—	—	—
21. Comments and criticisms of student work were instructive.	—	—	—	—	—
22. The exams and quizzes were a good test of students' mastery of the course material.	—	—	—	—	—
23. The exams emphasized original thought rather than memorization of facts.	—	—	—	—	—
24. The exams concentrated on: a) lectures b) readings/text c) lectures & readings d) other	—	—	—	—	—
25. In comparison to other courses, I spent_____ time on this course. a) much more b) somewhat more c) about the same d) somewhat less e) much less	—	—	—	—	—

GENERAL

Rate the following items on a scale between OUTSTANDING and POOR (or NOT APPLICABLE – N.A.)

	NA	O			P
26. The text	—	—	—	—	—
27. The assigned problems	—	—	—	—	—
28. The reading assignments	—	—	—	—	—
29. The papers/written projects	—	—	—	—	—
30. The exams and quizzes	—	—	—	—	—
31. The recitations/discussions	—	—	—	—	—
32. The laboratories	—	—	—	—	—
33. The lectures	—	—	—	—	—
34. The over-all quality of the course	—	—	—	—	—

PERSONAL	a	b	c	d	e
35. I am a: a) Fr. b) Soph. c) Jr. d) Sr. e) Grad.	—	—	—	—	—
36. This course is: a) in my major b) in a cognate field c) for distribution d) and elective	—	—	—	—	—
37. My grade I expect in this course is p) Pass a) A b) B c) C d) D e) E/Fail	—	—	—	—	—
38. My cumulative average is: a) below 2.0 b) 2.0-2.49 c) 2.5-2.99 d) 3.0-3.49 e) 3.5 or above	—	—	—	—	—

OPEN-END

Feel free to make any appropriate comments (or none) on the following topics. The sub-topics are only suggestions. Be CONCISE.

CLASS SIZE: Were the lectures and recitations too big, too small? Would the lecture have been more effective as recitations or vice versa? Was there opportunity for student participation, suggestions and criticisms?

RECITATION: Name of recitation LEADER _____

Did the recitation leader lecture or encourage discussions? Did he clarify the lecturer or add other worthwhile material? Was he responsive to questions and criticisms? committed to one point of view? sensitive to the difficulties students were encountering in the course? Consider also items from INSTRUCTOR which may apply.

LABORATORIES: Did they add to understanding of principles being taught? Was the lab instructor useful as a guide in using the equipment, running the experiments?

LECTURER: Any outstanding – good or bad – characteristics of the lecturer not covered under INSTRUCTOR. State reasons for giving ratings there of a) or e).

ASSIGNMENTS/EXAMS: Were they unclear, too difficult, trivial, never discussed in class? Were the readings useful, readily available, suited to the course and the class level?

COURSE IN GENERAL: Any comments or criticisms? things you would like to see changed, things that were good and should be kept. Should it move faster or slower, be more in depth or more general, have different prerequisites . . .

APPENDIX B

A PROCEDURE FOR HANDLING CLASSROOM OBSERVATION

SUGGESTED GUIDELINES FOR THE DIVISIONAL EVALUATOR

I. Allocation of *Responsibility*

A. It shall be the responsibility of the instructor being evaluated to initiate all arrangements with the evaluator and provide all the necessary materials and information required or requested.

B. It shall be the responsibility of the evaluator to:

1. Critically and cogently comment *within* the framework set by the instructor, unless the framework, goals, structure, and method are manifestly effecting the conduct of the course either from an innovative, inappropriate or disjunctive nature.
2. Once a visit has been made, the evaluator should speak with the instructor about the session as soon as possible.
3. From his notes and after the conclusion of a second visit, he should write a brief description on the characteristics of the instruction and make an overall judgment of the instructor's effectiveness as a guide for FPC. He should communicate such, either orally or in written form, back to the instructor.
4. In writing an evaluation, perspective should be the course as a whole and not only the classroom experience.

II. *Structure and Materials*

Before he makes his class visit, the instructor and evaluator should have an informal discussion concerning the following:

A. An examination of the explicit goals and structure of the course.

B. An examination of the materials and methods used (programmed texts, tape recorders, phonograph, films, overhead projectors, etc.).

C. An examination of outlines, syllabi, etc., that would indicate how the course is organized.

D. Examine the means by which students are evaluated and informed of their progress (tests, papers, counseling, etc.) and the results which students achieve in the course.

E. Discuss the make-up of the class (e.g., Is the course required of all? of majors? What level is the course? What specific problems face the instructor?).

III. *In-class Visitation*

The first session described above should provide the evaluator with some indication of the organization of the course as a whole. He is aware that classroom performance is only a part of the total structure of the course instruction. However, in his class visitations, the evaluator should take into consideration the following:

A. *Organization:* the extent to which what the instructor has said were his goals is being achieved in the classroom. Consider organization of class time, materials, etc.

B. *Communication:* basically, this asks how effectively the instructor communicates with his class. One might consider clarity of presentation, the giving of assignments, the handling of questions and discussion, the effectiveness of the teaching methods employed, etc.

C. *Awareness:* is the instructor aware of and sensitive to the class as a whole and to the individual members: Is there a positive interaction between instructor and student? What kind of atmosphere is generated in the classroom? Has the instructor been able to interest the class in the subject and motivate them to do good work?

One might also note whether any aspects of the instruction detract from the effectiveness of the teaching (e.g., voice, attitude, mannerisms, etc.).

At the end of the class, it would be well to ask the instructor to indicate how he felt the class went.

(2-25-70)

APPENDIX C

PART OF A DEPARTMENT CHAIRMAN'S EVALUATION FORM, EMPHASIZING RESULTS OF INSTRUCTION

DEPARTMENT CHAIRMAN'S EVALUATION

NAME OF FACULTY MEMBER _____

RANK _____ DEPARTMENT _____

CHAIRMAN _____ DATE _____

Please answer narrative parts of numbers 1-4 on separate pages. Answer number 5 in the space provided on page 2.

The questions included below are suggestive of relevant considerations, but respondents may find some of them inappropriate in some cases. Appropriate and thoughtful evaluations are needed, and respondent should feel free to include appropriate considerations not included here and to omit those which he feels clearly do not apply. In all cases, concrete examples should be included wherever possible.

1. *Effectiveness in teaching.*

(*Note*: The following questions pertain to the *ends* of instruction too often overlooked in teacher evaluation. But the respondent may find that he needs to include comments on how a faculty member works with the *means* of instruction. In all cases specific information is more useful than only broad generalizations.)

Are students better informed after working with this faculty member? more motivated for intellectual endeavors? more interested in his field? more aware of the place of his field among other fields? better disciplined in critical and analytic skills? better accomplished in other appropriate skills? better able to use abstract concepts? better able to work independently? more involved in joint projects related to the faculty member's research or creative interests? more interested in professional affairs?

APPENDIX D

A FORM FOR PEERS TO USE IN MAKING AN EVALUATION (NOT NECESSARILY BASED ON CLASSROOM OBSERVATION)

Faculty Member _____

Improvement of Instruction Program

Department _____

EVALUATION BLANK FOR USE BY PEER EVALUATORS

Evaluation Year _____

Write an evaluation of this faculty member with respect to each of the following areas. In each case give as much concrete evidence as you can to support your evaluation.

A. Instructional Services (course and individual instruction, scholarship, etc.).

B. Services in the Discipline (research, departmental service, professional activity, etc.).

C. Institutional Services (counseling, faculty and committee work, student organizations, services to faculty colleagues, special programs, etc.).

D. Community and Public Services.

(use reverse side for additional comments)

Do you know of any particularly significant publication or research accomplished, or of any significant event, project, award or honor which the faculty member participated in or received during the last year? If so, describe and evaluate.

The Faculty Personnel Committee is particularly interested in your evaluation of the faculty member as a scholar and teacher. Keeping this in mind, and focusing on the last year, rate the faculty member using the rating scale below. Use the faculty members at the same rank as the comparison group.

Scholarship

Bottom 30%	Middle 40%	Next 15%	Next 10%	Top 5%

Teaching

Bottom 30%	Middle 40%	Next 15%	Next 10%	Top 5%

(5-13-70)

APPENDIX E

A FORM FOR FACULTY SELF-EVALUATION

FACULTY SELF-EXAMINATION

Excellence in teaching, as defined in our Faculty Handbook, is that which will inspire and convey the *excitement* of learning. This would seem to imply a necessity for the instructor to maintain a high level of enthusiasm and excitement about the positive value of learning. With this requirement in mind the following questions seem appropriate to ask of ourselves.

1. AL = ALWAYS 2. US = USUALLY 3. SM = SELDOM
4. NV = NEVER 5. DK = DON'T KNOW

Circle the appropriate response.

A. Preparation – Do I

1. Keep up with the current literature in my field?	AL	US	SM	NV	DK
2. Attend workshops and conventions?	AL	US	SM	NV	DK
3. Update class presentations with new information?	AL	US	SM	NV	DK
4. Seek better textbooks and materials often?	AL	US	SM	NV	DK

B. Class – Do I

1. Know, to the best of my ability, who my students are – names, ability, and background?	AL	US	SM	NV	DK
2. Try to design class to meet needs of all student levels of experience?	AL	US	SM	NV	DK
3. Leave the class anticipating with pleasure the next meeting?	AL	US	SM	NV	DK
4. Let students know clearly my expectations of them?	AL	US	SM	NV	DK
5. Encourage and/or insist on questions and discussion?	AL	US	SM	NV	DK
6. Give tests which evaluate how well my stated goals have been reached?	AL	US	SM	NV	DK
7. Remain an impartial and fair evaluator?	AL	US	SM	NV	DK
8. Invite a failing student in for conferences?	AL	US	SM	NV	DK
9. Change my methods when students seem not to be learning?	AL	US	SM	NV	DK
10. Explain exactly how final grades are derived?	AL	US	SM	NV	DK
11. Keep accurate and efficient records to show students who wish grade reviewed?	AL	US	SM	NV	DK
12. Return papers promptly and marked so students can see what they did right?	AL	US	SM	NV	DK
13. Use tests as a teaching device?	AL	US	SM	NV	DK
14. Avoid using the same tests from year to year?	AL	US	SM	NV	DK
15. Make different tests of equal difficulty for each section of the same course?	AL	US	SM	NV	DK
16. Maintain office hours regularly during which students may drop in or schedule conferences?	AL	US	SM	NV	DK
17. Perceive students as individuals instead of stereotyping?	AL	US	SM	NV	DK

18. Try to find loopholes in my teaching and correct them?	AL	US	SM	NV	DK
19. Use books and library assignments to enrich classroom instruction?	AL	US	SM	NV	DK
20. Encourage independent reading as a lifetime habit for my students?	AL	US	SM	NV	DK
21. Set an example for my students by referring to books in my class lectures and discussions?	AL	US	SM	NV	DK