

1978

A developed and field-tested experiment to test the effect of increased student feedback on specific teacher performance behaviors

Carl Robert Bennett
Iowa State University

Follow this and additional works at: <https://lib.dr.iastate.edu/rtd>

 Part of the [Educational Administration and Supervision Commons](#)

Recommended Citation

Bennett, Carl Robert, "A developed and field-tested experiment to test the effect of increased student feedback on specific teacher performance behaviors " (1978). *Retrospective Theses and Dissertations*. 6446.
<https://lib.dr.iastate.edu/rtd/6446>

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.

INFORMATION TO USERS

This material was produced from a microfilm copy of the original document. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the original submitted.

The following explanation of techniques is provided to help you understand markings or patterns which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting thru an image and duplicating adjacent pages to insure you complete continuity.
2. When an image on the film is obliterated with a large round black mark, it is an indication that the photographer suspected that the copy may have moved during exposure and thus cause a blurred image. You will find a good image of the page in the adjacent frame.
3. When a map, drawing or chart, etc., was part of the material being photographed the photographer followed a definite method in "sectioning" the material. It is customary to begin photoing at the upper left hand corner of a large sheet and to continue photoing from left to right in equal sections with a small overlap. If necessary, sectioning is continued again — beginning below the first row and continuing on until complete.
4. The majority of users indicate that the textual content is of greatest value, however, a somewhat higher quality reproduction could be made from "photographs" if essential to the understanding of the dissertation. Silver prints of "photographs" may be ordered at additional charge by writing the Order Department, giving the catalog number, title, author and specific pages you wish reproduced.
5. PLEASE NOTE: Some pages may have indistinct print. Filmed as received.

University Microfilms International

300 North Zeeb Road
Ann Arbor, Michigan 48106 USA
St. John's Road, Tyler's Green
High Wycombe, Bucks, England HP10 8HR

7513219

BENNETT, CARL ROBERT
A DEVELOPED AND FIELD-TESTED EXPERIMENT TO
TEST THE EFFECT OF INCREASED STUDENT FEEDBACK
ON SPECIFIC TEACHER PERFORMANCE BEHAVIORS.

IOWA STATE UNIVERSITY, PH.D., 1978

University
Microfilms
International 300 N. ZEEB ROAD, ANN ARBOR, MI 48106

A developed and field-tested experiment to test the
effect of increased student feedback on specific
teacher performance behaviors

by

Carl Robert Bennett

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of
The Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Department: Professional Studies
Major: Education
(Educational Administration)

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa

1978

TABLE OF CONTENTS

	Page
CHAPTER I. INTRODUCTION	1
Statement of the Problem	4
Hypotheses Tested	5
Potential Value of this Investigation	5
Delimitations of the Study	6
Definition of Terms	8
CHAPTER II. REVIEW OF LITERATURE	10
Historical Background	10
Current Evaluation Methods	15
Support for Student Observers of Teaching Methodology	18
Importance of Multiple Appraisers	25
Experimental Teacher Performance Instruments	27
Summary	29
CHAPTER III. METHODS AND PROCEDURES	31
Introduction	31
Delimitation	31
Preparation of Instrument	32
Selection of Teachers	34
Collection of Data	35
Treatment	37
Statistical Methods	39
Generation of Means	40
Pearson Product-moment Correlation	41

	Page
t-tests	41
CHAPTER IV. FINDINGS	42
Hypotheses	42
CHAPTER V. SUMMARY, CONCLUSIONS AND RECOMMENDATIONS	58
Summary	58
Conclusions	60
Limitations	63
Recommendations for Practice	64
Recommendations for Further Research	67
Discussion	69
BIBLIOGRAPHY	73
ACKNOWLEDGMENTS	77
APPENDIX	78

LIST OF TABLES

	Page
Table 1. Analysis of student mean responses for control teacher behaviors (correlations and t-tests)	43
Table 2. Analysis of student mean responses for experimental teacher behaviors (correlations and t-tests)	45
Table 3. Analysis of differences of student mean responses between teachers not receiving feedback (control group) and teachers receiving feedback (experimental group). Data taken from fourth (last) data collection	47
Table 4. Variable (mean of 5 items selected by the investigator and mean of items selected by teachers), means, coefficients of correlation, discriminating value of t-test, and probability for between the two groups. Data taken from last collection. . . .	51
Table 5. Analysis of student mean responses for each item on the instrument as reported by males and females on the second data collection.	52
Table 6. Analysis of teacher responses to questions about the validity of student assessments of teaching methodologies.	57

CHAPTER I. INTRODUCTION

In recent years the public has been bombarded with an assortment of critics who have been taking potshots at public schools. Writers look at drug abuse and changing moral standards of today's young people and lay blame on the school's doorstep. Other critics see formal education to be unimaginative and too much like that of yesterday. They point to the regimentation imposed on youth by dictatorial administrators and shake their heads in disbelief at callous teachers reminiscent of Charles Dicken's Mr. Creakle, the scurrilous school master who badgered poor David Copperfield. The latest attack comes because results are declining on standardized basic skills tests. With today's schools operating on the biggest budgets in history and having the latest teaching gadgetry, the critics chastize teachers and administrators for graduating students who collectively score lower on standardized test norms than did their counterparts twenty years earlier.

Obviously, if one examines any of these accusations carefully, other reasons can be given that would deflect some of the criticism from schools, but pointing the finger at another cause does not enhance the educator's respectability. Thus, educators are faced with the challenge of simultaneously making schools more humane and more productive. Inherent in this quest is the improvement of instructional techniques of teachers.

The improvement of instruction is an admirable goal but when pursued presents many problems. A simple solution might be to purge poor

instructors and frill courses, but how to decide who is a poor teacher and what is a frill is not as easy as it would first appear. A better solution is to improve the quality of all instruction now being offered. Releasing a poor instructor does not improve the teaching of others. Neither does it guarantee the replacement will be any better.

For years there have been efforts to evaluate teachers, but in the past two decades the process has been taken more seriously. At recent conventions of the American Association of School Administrators, the "standing room only" seminars have been those dealing with teacher evaluation. It is no wonder. Each year more states are mandating teachers evaluation programs. Both teachers and administrators are searching for methods which will accurately assess teaching performance and point to ways for improving performances judged to be inadequate.

Because it is difficult to measure teacher effectiveness on the basis of student output or teacher personality characteristics, it appears more reasonable to focus upon the various forms of pedagogical behaviors exhibited in the classroom. Teaching, in this sense, can be viewed as a set of logical operations carried out by the instructor in a particular social setting. Theoretically, the more proficient a teacher is when engaging in such operations, the better is the teacher. Menne explains that measures of teaching performances can be obtained by one of several groups of people. He believes, however, that no matter who observes teachers, there are three conditions necessary to show that measurement has occurred:

- a. There must be more than one rater;

- b. the ratings must closely agree for the individual being observed;
- c. the ratings must indicate differences in teachers.
(25, p. 4)

While there has been considerable attention given to teacher performance, there is little evidence that such assessments have caused teachers to change or improve their behavior in the classroom. If the purpose of evaluation is to improve instruction, it seems imperative that improvement occur after appraisals are made.

Current practices dictate that teachers are to be observed and formally evaluated two to three times a year. Often a teacher is observed even less. It is difficult to believe that accurate assessments of a teacher's ability can be obtained with so few visits to a classroom. It is also understandable why little, if any change, in pedagogical behavior can be noticed when evaluation and feedback occur on such an irregular basis. Worse yet, only one evaluator sees a teacher. If the instructor is evaluated by more than one person, the visits occur at different times, so an opportunity for comparing ratings is invalidated because of the different circumstances observed.

Chatterjee, Daw and Gage believe educators value assessment if the process will help them improve their skills. They also assert that teaching performance can be modified when the assessment is based on techniques thought to be important by the instructor and the pupils. In the "equilibrium" theory, these researchers propose that if one increases the feedback to teachers regarding teaching strategies, that teachers will change their behavior in classrooms (In Gage et al.,

14, p. 174).

Studies by Heider, Newcomb, Osgood, Tannenbaum, and Festinger show that teachers strive to maintain a harmonious relationships with their students, in other words, a "state of equilibrium". If feedback to the teacher from students indicates that the teacher's performance is not adequate, the desired harmonious state no longer exists. When such a condition is known by the teacher, the researchers believe the instructor will try to change or improve the behavior being criticized (In Gage et al., 14, p. 174).

There is still reluctance to allow students opportunities for giving teachers feedback about teaching performances, but there is increasing evidence that students can and do make valid assessments about teaching behavior.

Is it possible, then, to devise an improved process by which teaching performances can be measured and changed or improved? The development and testing of such a model is the problem to which this study will be directed.

Statement of the Problem

The problem of this study was to develop and test a student-centered feedback system that would be used by teachers to assess their teaching performances. More specifically, the instrument and system was to be used to seek answers to the following questions:

- a. Can teacher behavior be modified when receiving repeated feedback about selected performance objectives?

- b. Can selected teaching behaviors relating to improvement of teaching performances be identified for students to observe?
- c. Can students give accurate feedback to teachers about their performances in the classroom?
- d. Is there any difference in the kind of feedback provided by male and female students?

Hypotheses Tested

The pilot experiment with this instrument offered the opportunity to test the following hypotheses with thirty-seven high school teachers in four Iowa school districts:

- a. Teachers receiving feedback from student observations will show a more significant positive change in student rating scores than teachers not receiving a feedback.
- b. There will be more significant positive rating changes for feedback items chosen by the teachers than those items fixed by the research design.
- c. The kind of teacher feedback will not vary significantly by the sex of the student offering feedback.
- d. Teachers will have positive reactions to student feedback regarding their teaching methodology.

Potential Value of this Investigation

The primary purpose of this study was to investigate the effect of increased feedback to teachers about selected teaching behaviors. It was reasoned that if feedback about teaching methodology was necessary to improve teaching, the more often it occurred, the better the chances for teacher improvement would become. The results of this study may stimulate teachers and evaluators to ask for feedback more often

about teaching performances.

Objectivity about teaching performances is always difficult to obtain. Students may be one of the best and most efficient sources for such information. Students were used to give feedback in this experiment. Perhaps the influence of their comments on teacher behavior will encourage other instructors to seek students' help when they are looking for such information.

Administrators charged with improvement of instruction may look to this kind of procedure as a method for helping teachers take a new look at themselves. The instrument used in this experiment allowed the teachers being rated to have some input as to the kind of items about which they would be given feedback. The inclusion of teachers' input to rating instruments may be the key for genuine teacher acceptance of wanting feedback from students about their teaching performances.

Delimitations of the Study

The study was delimited to investigating what effect increased feedback from multiple appraisers about specific teaching behaviors would make on educator performance behavior. Several evaluation schemes for rating teachers were examined and a search of the literature was conducted in the area of attempts made on modifying performances of teachers in the classroom.

Because of the large amount of attention to the subject in recent years, reading was restricted to materials dealing with multiple assessments of teaching behavior, specifically those experiments where

students have been used to give instructors feedback about their teaching methodologies.

Measurement of educator performances was limited to feedback from students of forty-three Iowa English and social studies teachers. These high school teachers were full-time educators at Algona Public, Estherville Public, Fort Dodge Public and Spencer Public schools. Each teacher was asked to select one class and the students in this class were asked to provide feedback about the teachers' methodology every two weeks for a period of eight weeks.

Assessment of behaviors was measured by an instrument which included items selected by the investigator and items selected by the instructor. Items selected were from a list of teaching behaviors tested for having discriminating qualities by Madeline Hunter and a team of researchers at the University of California at Los Angeles.

Treatment for the teachers receiving feedback was limited to the receipt of data provided by the evaluative instrument designed by each instructor. Computer printouts summarizing data recorded by students were given to the teachers in the experimental group every two weeks during the eight weeks experiment. Teachers in the control group were rated every two weeks also but did not receive feedback until the end of the eight-week period. Mean scores for each of the ten items selected by investigator were obtained from all the teachers. Scores of those receiving feedback were compared with those who did not. The mean score of items selected by the investigator and those selected by the teacher were also compared.

Definition of Terms

1. Accountability - holding the schools (and professionals) responsible for results in terms of student learning rather than solely for the use of input resources.
2. Appraisal Feedback - the rating of criteria selected to determine effective teaching performance.
3. Assessment - the process of making a value judgment about how well an instructor performs a specific teaching procedure.
4. Behavior - an educator's actions when attempting a specific teaching procedure.
5. Behavioral Objectives - a statement of terminal student behavior or instructional outcomes that can be measured.
6. Evaluation - the process of ascertaining the effectiveness of teaching methodologies for the purpose of improving instruction.
7. Feedback - the return of compiled data about teaching behaviors to the instructor being observed.
8. Inservice - a group activity in which students and teachers are taught how to use the instrument devised to give teachers feedback about teaching behaviors.
9. Job Targets - a statement of desired terminal teacher behavior that can be measured.
10. Multiple Appraisers - more than one person making an appraisal of teaching behaviors at the same time.
11. Student Feedback - data given to teachers about how well their

students thought they performed specific teaching acts.

12. Teaching Performance - the execution of specific behaviors in the classroom.

CHAPTER II. REVIEW OF LITERATURE

Historical Background

Evaluation of teachers has received much attention in light of the growing demands for accountability of educational institutions. The issue of accountability has reflected a deep and genuine apprehension on the part of the public that the schools are not doing the job they should be doing. Insistence on accountability is perhaps a logical consequence of mounting costs and rising teacher militancy. As far back as 1970 the Iowa Governor's Educational Advisory Board stated that citizens were concerned about such things as increasing education costs and the need for accountability of teachers (12, p. 1).

"Asking for accountability is a legitimate response of government, which wants to know how efficiently its money is being spent," said Theodore B. Dolmatch, president of the Pitman Publishing Corporation, at a conference of educational marketers and investors. According to Dolmatch it is a legitimate response of taxpayers to want to know how their money is being spent. He also believes it is a legitimate response of parents to want to know how well their children are being educated (38, p. 41).

It appears that the key issue in the broad area of accountability in education has centered around who is doing a good job in the classroom and how that is being determined. The major problem in answering those questions is determining a way to ascertain quality teaching and how to improve the performances of those instructors whose classroom

behavior needs modification.

To the classroom teacher, this clamor for accountability from individuals and groups who have had little or no recent experience in the day-to-day business of teaching in elementary and secondary schools may sound like an indictment and even a threat. Actually, these clarification calls for better education are not so new. At the turn of the century serious empirical attempts to assess educator performance were being made. In 1896 master educators were selected to observe other teachers and then submit their findings to appropriate authorities. The major purpose for this procedure was to determine whether or not the teachers being observed deserved an increment in their salaries or should be dismissed for being incompetent (23, pp. 413-18).

Educational literature is filled with discussions of investigations that sought ways of assessing teacher performance. Two excellent reviews of the literature in this area are the doctoral dissertations of Everett Hidlebaugh (16) and Robert Eickhoff (11) both of which were completed at Iowa State University.

Perhaps the latest piece of research regarding assessment of teachers is that done by Richard Shavelson and Nancy Dempsey-Atwood. Their research at the University of California in Los Angeles gives an excellent review of studies about teaching effectiveness. Unfortunately their conclusion entertains the possibility that it is impossible to generalize about teaching behavior because most studies are methodologically inadequate at this point in time to resolve the issue. (36, pp. 553-611).

Ironically, with all the research that has been done to assess the quality of good teaching, as yet there is still not a clearly defined criterion for the excellent instructor. Biddle and Ellena stated it this way:

Recent summaries have revealed that literally thousands of studies have been conducted on teacher excellence since the beginning of the century. Investigators have looked at teacher training, traits, behaviors, attitudes, values, abilities, sex, weight, voice quality, and many other characteristics. . . . And yet, with all this research, results have been modest and often contradictory. Few, if any, facts are now deemed established about teacher effectiveness, and many former "findings" have been repudiated. (2, p. 6)

Historically, teaching has been considered an art and those people labeled teachers presumably had some kind of mysterious personal power to inculcate youth with the knowledge and personal growth to which schools have been held responsible. In other words, if one went to school and took the right kind of courses and received a diploma and teaching certificate, he/she could teach!

Of late, however, there are some who have begun to doubt the "inherent abilities" of those people certified to be called teachers. They have begun to demand a more rigorous examination of teaching as an activity rather than rely upon a belief in certificated titles. Following this lead, in the spring of 1976, the National Institute of Education called for a new approach to the definition of teachers. It suggested licensing teachers on the basis of competence/performance demonstrated in the classroom rather than on the basis of satisfactory completion of college and university programs (33, p. 1).

Unfortunately, until a universal accepted definition of what is a

competent teaching performance can be agreed upon, the above mentioned approach for ascertaining teachers will be impossible. However, how well an instructor performs whatever methodology he chooses to use in the classroom might be rated with some degree of objectivity.

So far, most colleges and universities and state departments have left the issue of teacher competency to the whims of local school systems. Thus, teachers and administrators have struggled with local school boards to create acceptable evaluation instruments that will help identify effective instruction.

As has been stated earlier, evaluation of teachers is not a new process. For years some school systems have been evaluating their teaching staffs. In the last decade this process has been given new attention because of state laws that now require evaluation of teachers to occur in all school systems. In Iowa, for example, the Iowa General Assembly modified procedures for terminating the contract of an Iowa teacher in May of 1976. The intent of the bill labeled SF 205 was to establish fair dismissal practices for all certified employees of a school district employed by the school district with the exception of superintendents, assistant superintendents, principals and assistant principals. Inherent in such a bill was a key provision that stated, "The board shall establish evaluation criteria and shall implement evaluation procedures." Moreover, "If an exclusive bargaining representative was certified, the board must negotiate in good faith with respect to the procedures for the evaluation (35, p. 262a).

Unfortunately, in some districts the attempts at constructing

evaluation programs have bogged down with the negotiations procedures that have occurred in the past several years. Ted Davidson, executive officer of the Iowa Association of School Boards, wrote that negotiations in many parts of Iowa are bogged down as a result of excessive demands by teachers' associations and their insistence on negotiating issues which would undercut school management (19, p. 1).

As negotiations between teachers and their respective school boards continue, more and more issues cloud the area of determining effective teaching. Of late, school boards have been confronted with proposals from teachers which would limit the boards' rights to decide how many pupils to put in a classroom. The teachers are presenting this position because they feel there is a direct relationship between the number of students in the class and how effective they are as instructors.

Such proposals have produced abrasive relations between teachers and administrators and school boards. In December of 1976, only three months before school budgets were to be in final form, it was reported that only 46 school districts had negotiated salary settlements and that 108 school districts were still at impasse (30, p. 1).

Certainly, this process has taken many valuable hours of both teachers and administrators, and there is no evidence to suggest that teachers' performances have improved in the classroom because of the long negotiations procedures. Nor is there evidence that such efforts have improved the process of evaluating teachers more accurately and fairly.

Current Evaluation Methods

In an assessment of teaching, Johnson explained that there does not appear to be a logical, theoretical framework for analyzing teaching and the phenomena associated with the profession. He believes that at best, when any practical attempt to deal with teaching is begun, one must know what teachers are supposed to be doing if a judgment about their ability is to be made. Johnson has found that analysis of pedagogical behaviors fails to distinguish critical teaching acts from more general teacher characteristics interpretable in terms of teacher personality (21, p. 173).

Herman contends that any program of evaluation is incomplete without the addition of an inservice or a job upgrading phase. He believes that it is grossly unfair to the person being evaluated if areas of weakness are identified and no program of assistance is provided which will enable the employee to overcome his weaknesses and improve his performance. To get a clear picture of what the instructor is attempting requires a pre- and postconference between the evaluator and the teacher. However, Herman believes that even with all of these necessary procedures taken, the anxiety level of many instructors raises when a formal evaluation occurs. Too often instead of presenting a program to meet the needs of students, the teacher behaves in such a way that is thought will please the biases of the evaluator (15, pp. 4-14).

Contemporary evaluation procedures used to evaluate teachers was the cause of great concern for the Iowa Governor's Educational Advisory Board. The Board found that in many schools the process was far from

systematic. The time necessary to make classroom visits and have the recommended pre- and postconferences with each teacher visited plays havoc with most evaluator's schedules. Unfortunately, then, there are fewer formal evaluations made than desired by either teacher or administrator. The Advisory Board found that commonly, evaluation took place sporadically during the first year and subsequently only after the teacher had a problem (12, p. 2).

The fact that evaluative information is usually gathered at isolated points in time raises some problems. First, some care should be taken to ascertain how representative the information is of a faculty member's typical performance rather than that of an atypical routine to please an administrator.

In most public schools the person most often held responsible for evaluating a teacher's performance is an administrator who also fulfills the role of chief disciplinarian of the school. Students are not oblivious to this situation. Thus their behavior changes as well as the teacher's when such a person enters the classroom. Information gathered at isolated points in time by building administrators can be very misleading under such circumstances. Too often, the evidence gathered only supports a conclusion already formed before the observation occurred and those conclusions are based more on personality judgments rather than on systematic judgments of what the teacher normally does in the classroom.

The second time-related problem in faculty evaluations has to do with evaluations and individual faculty growth. Because information

gathered at isolated points in time may lack the context necessary for a meaningful interpretation, evaluation should be a repeated if not continuous undertaking. Doyle believes that the most useful evaluative data are those that reveal patterns of effectiveness over time, particular in the case of younger teachers. He has concluded that the more likely a person is to be changing, the more important it is that his/her evaluations reflect such change (9, p. 16). According to Doyle, then, if one is not seen on a regular basis, it would be difficult to note any changes in the teaching performance. Then an opportunity for influencing a teacher's behavior would be lost.

Evaluations that occur infrequently and that are done by the building administrator may be the most typical found in most schools and may meet negotiated and state law requirements, but one can hardly contend that a valid assessment of a teaching performance is obtained with such procedures.

With the present financial straits facing school systems, perhaps the present method employed by administrators is the best that can be provided to give administrators the kind of information they need to make decisions about teachers' salaries, tenure and rank in the department. However, for the teacher who is looking for ways of getting objective feedback about his/her performance in the classroom with the intention of becoming a more effective instructor, there is another source besides administrators that may be able to provide that service. That source is the students who face the instructor every day he/she appears in the classroom.

Support for Student Observers of Teaching Methodology

No one would propose that student ratings of teachers will insure better teaching, but evidence exists that feedback from students about teaching methodology can be a useful tool for teaching improvement if used properly. When investigating effective teacher evaluation programs, Hidlebaugh stated that the bulk of empirical evidence indicated that students offered the best feedback toward assessing a teacher's true performance. He concluded that secondary student ratings of teachers could provide valid input into the evaluation process (16, p. 96).

The practice of collecting feedback about teachers from their pupils has had a moderate amount of acceptance for about the last thirty years. Normally, this feedback has been presented in the form of a rating of specific behavior or personality traits of the instructor. Ratings are a complicated form of measurement because they involve three variables: the rater, the task, and the ratee.

Perhaps no ratings are so controversial as those done by students about their teachers. Student ratings constitute at least a perceived threat to the self-esteem, reputation, perhaps even a threat to a teacher's professional career. Skeptics argue that students will only be able to rate that which is superficial and merely entertainment. There is some justification for this kind of thinking. Two studies by Bush have shown that one of the most powerful factors relating to positive feedback about an instructor is that of the personality of the teacher. Those teachers projecting an image liked by students received the highest ratings by their students according to the Bush experiments

(5, p. 51).

Another study about the importance of student feedback was that done by Morsh and Wilder in 1954. After searching the literature about the importance of student feedback, these two investigators concluded that there was not much evidence to support any claim that students' feedback about their teachers' behavior would have an effect on the teachers' performances (27).

The arguments for and against the use of student ratings as a basis for improving teaching have been going on for some time, now, and the question most often asked from teachers is how good is the data gathered from the students?

Before one looks at whether or not the data collected will make a change in the behavior of the instructor, a determination should be made as to whether or not students can do at least as effective a job at collecting data as their adult counterparts who are currently involved in this process.

Studies by Wise (42) and Moss (28) in 1971 and 1973 indicate that the criteria used by professionals (supervisors, administrators, department chairpersons, etc.) varied little, if any, from those used by students in rating teacher performances.

As to whether or not students will be able to determine what are desirable teaching performances of instructors, the young people seem to do at least as well as the administrators who are given the responsibility for determining the goals and philosophies of our schools. In a study by the Ohio Commission on Public School Personnel Policies, the

Commission concluded that students and administrators did not differ significantly as to the behavioral traits desired of instructors (10, pp. 48-54).

When reviewing the literature about the importance of student feedback about teaching performances, it has been noted that researchers seldom discriminate between the terms educator performance and educator effectiveness. As far as this study goes, the distinction must be made. Such was done by Menne when he stated:

If the behaviors of the teachers are measured in some way (e.g., by observations made by administrators, peers, or students), then the teacher's performance is being evaluated. However, if the incremental knowledge gained by the students as a consequence of the contact with a particular teacher is measured, then the teacher's effectiveness is being evaluated. (25, pp. 26-27)

Swanson and Sisson investigated the use of a theoretical model for the appraisal of teachers which identified three dimensions of an instructor's activities. The model was designed to allow for assessment of teaching performance, scholarly productivity, and service to the educational community. Their conclusion was that students are best qualified to rate the performance of the teachers but are not able to assess the research and service dimensions of an educator's contributions (39, pp. 64-79).

As one begins to refine what is to be found out about the teacher and narrows this information to the teaching performance, the feedback from students about this part of a teacher's activities takes on new importance. Costin, Greenough and Menges stated that the burden of evidence indicates that students' ratings can provide reliable and valid

information. They noted that where criteria for educator performance exists, i.e., supervisor and peer ratings and measures of postinstruction student performance, student ratings tended to show a low positive correlation, suggesting that such assessment does make its contribution. They also claim that student feedback in the form of student ratings may improve an educator's performance (7, pp. 511-535).

Frey made the distinction between students as evaluators and students as information sources. He stated that: "When a student makes an evaluative judgment about his teacher, he is likely to weight the specific teaching traits somewhat differently than would a faculty member or an administrator." He goes on to explain that when care is taken to develop a sound measuring instrument, instructional ratings from students can become particularly valuable to the teacher who wants an unbiased account of his/her teaching performance (13, pp. 83-85).

As far as students having the ability to assess the effectiveness of a teacher's performance on what is learned by the students, the evidence is pretty spotty. But McKeachie points out that students do indeed make judgments about their teachers, and even though these judgments may be based on false or questionable evidence, the feedback may be of great value to the instructor. When large numbers of students share a particular judgment about a teacher's behavior, McKeachie believes that judgment should not be ignored. If nothing else, he believes the feedback forces the instructor to take another look at what he/she is attempting with the students (24, pp. 439-44).

Some teachers are hesitant to ask certain kinds of students to

provide feedback about their performances because they feel the grade the student is getting in the course will affect or bias the students' opinions about the teaching performance being assessed. Blum investigated this area and concluded that whether students received an A, B, C, or D in the class, the estimation of the instructor's performance remained essentially the same and closely resembled the average estimation of the group (4, pp. 217-21).

Voeks and French studied the same question and their findings were that grades and student ratings had no reliable relationship. The teachers with the highest student ratings seldom had given higher grades than teachers with the lowest ratings (41, pp. 330-34).

One of the most interesting discoveries to come from research concerning student feedback about teacher performance is the work done by Zelenak and Snider. They have found that teachers are more receptive to the idea of collecting feedback from students when they viewed the students as data collectors and not evaluators. These investigators suggested that evaluation philosophies are usually separated into two distinct areas as far as teachers are concerned. Whenever teachers are assessed to determine tenure, promotion, dismissal or salary, the assessment of their abilities is seen to be for administrative purposes. For an administrator to visit a classroom and watch the teacher and say otherwise about his visit is hard for the teacher to believe. Normally there is some resistance to this kind of procedure. However, when the teacher can be convinced that a classroom visit will be used only for helping the teacher improve a performance, the assessment of the

performance is welcomed. Administrators have much difficulty separating the purposes for their visits, but if student feedback can be presented in a nonthreatening way, the teacher seems more willing to give it a try (43, pp. 570-71).

Support for the work of Zelenak and Snider comes from the investigations of Jane Shaw and her work with a number of public school evaluation programs that incorporated students as part of the process. She found teachers seemed willing to cooperate with the idea of students being a part of evaluation if the feedback from them was seen to have the definite purpose of improving instruction. Teachers wanted nothing to do with the idea, however, if they thought students were going to be used as hatchet men for the administration (37, p. 49).

Shaw reported that at San Mateo, California, when student evaluation of teachers was first proposed at the high school, only about half the teaching staff agreed to participate in the experiment the first semester. However, when the second semester began, about thirty percent more of the faculty decided to become involved in the process. One of the students at the high school at San Mateo explained why she thought there was an increased interest by the teachers in the program. She said, "Teachers are genuinely interested in what students think of them --as long as the student input is well organized and not used as a witch-hunt-type thing. If there's an opportunity to get some student feedback in a responsible manner, teachers really like it and seem to regard it as valuable information" (37, p. 50).

Gage, Runkel and Chatterjee (14) believed that student feedback to

teachers was something more than "valuable information." From their studies they hypothesized that providing teachers information about the relevant feelings and wishes of their pupils would, in fact, change the teaching performance of the instructor.

They reasoned that a sensitive teacher gets feedback every day by watching students react to the teaching methodology being employed. The interpretation that the teacher gives to his/her students' nonverbal feedback helps the instructor restructure teaching behaviors to accommodate the students' needs. By glancing at the students' faces, a good teacher will notice signs of interest or boredom, comprehension or puzzlement, favorable attitudes or resentment. Through such feedback from students, these researchers believe the instructor gauges a performance much like an actor on a stage.

The research team reasoned that if such feedback from students could be more structured and provided on a more regular basis so as to give a more accurate description of what students thought about their teacher's performance, the change in the teacher's behavior would become even more significant.

An experiment with sixth-grade students and teachers proved their assumptions to be correct. Teachers receiving periodic feedback from students about teacher performances in the classroom via an instrument designed by the research team produced a greater change in teaching behaviors than those teachers who were not given any structured feedback (14, pp. 173-181).

Whether the meaningfulness of student evaluations has been

sufficiently documented in the literature still remains an open question. No firm conclusion about the meaningfulness of student feedback can be offered since different instruments contain different items for measurement. However, Doyle believes after studying many examples of student feedback, that in the absence of amply demonstrated empirical fact, that it appears student can provide as good a feedback to teachers about their teaching performances as anyone (9, p. 66).

Perhaps the important thing to remember about the validity of student feedback is that even though it cannot be demonstrated that students are the most adept sources for reporting the effectiveness of a teacher's performance, studies do show that when criteria for educator performance exists from supervisors, peers and students, there is a positive correlation suggesting that students tend to make similar value judgments about teaching performances as do the professionals. Thus, there seems to be an increasing amount of evidence to support the idea that students can make at least as valid assessments about their teachers as those done by administrators if they are given the opportunity to do so.

Importance of Multiple Appraisers

A review of literature regarding any type of teacher evaluation generally points out the conspicuous fact that if teachers' evaluations are to be valid and meaningful, more than one rater must be involved in the evaluation process. Empirical evidence suggest that a more accurate description of teacher behavior results from multiple appraisers making

repeated observations and using carefully worded instruments that allow for each item observed to be ranked.

Thorndike and Allen have experimented with rating instruments used by a single rater and their studies have shown that the reliability of single raters using a conventional rating procedure is low. They reported that the correlation between the ratings of two independent raters on a conventional rating scale would be about 0.55. They felt, however, that if it were possible to pool the ratings of a number of independent raters who know the teacher equally well, the reliability of the appraisal could be substantially increased (40).

Thorndike and Allen's assumption is based on the studies by Remmers who showed that pooled ratings of evaluators function in the same way as lengthening a test in respect to getting improved reliability. Thus, by applying the Spearman-Brown statistical formula to pooled independent ratings, Remmers estimated that if one rater has a reliability of 0.55, then two raters will have reliability of 0.71, and three raters will have a reliability of 0.79. When ten raters are used, the reliability is 0.92 (40, p. 346).

In view of the fact that few classes have fewer than twenty students and that all students have repeated opportunities to view the teaching methodology of the instructor and that most students would have a similar acquaintance with the instructor to be evaluated, it would seem logical to assume that students could give reliable feedback to their instructors. In fact, the research division of the National Education Association has recognized the problem of too few raters being

used to evaluate teachers and has suggested that multiple raters be used in the evaluation process. Using students for this purpose may be the only economical way to accomplish this suggestion.

Experimental Teacher Performance Instruments

The fact that most legislators and universities in many states have left the business of teacher evaluation to the local communities has caused hundreds of evaluation instruments to be developed. A study of these many forms revealed that with few exceptions, the "homegrown" instruments were not very different from those designed by academicians at large universities (37, p. 51).

What administrators who were using such instruments did find important about their acceptance by teachers was that instructors felt better about those instruments in which they had input to their development.

The idea that teachers would be more receptive to an evaluation instrument if it were to include items thought important to the instructor was experimented with by a research team at Purdue University. In 1972 with the aid of a computer and a catalog of rating scale items, they created an evaluation program which allowed a unique rating instrument for each teacher being evaluated. Called a "cafeteria" system, the instrument allowed innovative teachers to select items for the instrument that would assess their unique teaching strategies. Quite literally, each instructor shopped for a match between characteristics of the course--its focus, content, style and goals and the questions by which his/her performance could be evaluated (8, p. 7).

To each group of items selected by the teacher, a standardized and nonoptional core of five items was automatically added to the instrument. These core items served to make comparisons across and between individuals, and they assured that each of the five recurring dimensions of teaching were represented.

The investigators reported that faculty acceptance of the "cafeteria" system had been much greater than they originally expected when the experiment began. In 1974 they reported that this approach to evaluation was voluntarily adopted by the Purdue faculty in two thousand classrooms. The team believed the system's adaptability to most course offerings and the opportunity it afforded faculty to shape and select content in ways the individually preferred was responsible for its success (8, p. 10).

The Purdue study with evaluation of teachers is only one of many being tried in major universities throughout the country. Similar programs using students as part of the teacher evaluation process are being experimented with at the University of Illinois and the University of Michigan.

With so many teacher-observation systems in existence today, an attempt to make mention of them all would be foolhardy and probably impossible. What seems important to this study is that literature does show that students are being used successfully to give feedback to teachers more now than ever before. Perhaps more important, is that this feedback is well-received by teachers if it is understood and felt by them that the purpose of the student assessment is to help the

teachers improve their instructional techniques.

Summary

During the last two decades public schools have been under attack because the education of youngsters is costing more and today's graduates do not appear any better informed than students who were taught in less expensive schools. Efforts by administrators to improve the students' learning skills have recently centered around better assessments of teachers perhaps because increasing teaching salaries have caused a large share of the increase in cost of education. Accountability has been the watchword and basically two approaches have been used to determine the teacher's effectiveness. One approach looks at test scores of children after they have been with the instructor for a given period of time; the other examines the teaching act itself.

The first approach was found to have too many variables for the limits of this investigation, so the second concept was researched. A closer look at how the teaching process is normally evaluated revealed that most evaluative instruments are designed without input from the teachers receiving feedback and that multiple assessment seldom occurs or if attempted, is usually not done on a consistent basis. It was also discovered that when teachers' are evaluated there is little known about whether or not the teaching performances change as a result of the observation.

Current research shows that students have been and are being used to rate teachers and that when such is done, the students look for

the same teaching qualities as those judged by administrators. Research also establishes that, if provided with similar instruments, students do at least as well as administrators and department chairpersons in rating the teachers.

Recently, investigators have been using evaluation instruments which allow teachers to decide some of the items used in the rating. Teachers appear to be more receptive to this type of feedback than that provided by an instrument in which they have had no personal say.

With this evidence about teacher performance evaluation in mind, it was decided to design an instrument which would allow for teacher input that could be used by students on a regular basis. Thus when the principle of multiple assessment could be consistently employed providing the teacher with a more objective picture of his/her performance, an experiment could be designed which would correlate these factors with a positive change in the teacher's behavior in the classroom.

CHAPTER III. METHODS AND PROCEDURES

Introduction

The purpose of this chapter is to explain the procedures used to conduct the study. Having reviewed the literature, three major tasks remained: to prepare an instrument by which teaching behavior could be measured by students, to find teachers and students who would be willing to participate in the experiment, and to develop an experimental design through which statistical data could be manipulated to determine whether or not the experiment had any significance.

This investigation began because it was believed that most teachers want to improve their performances in the classroom, and that the present system for helping instructors do this is not very effective. An examination of current methods for helping teachers revealed that too often feedback about performances is not done with any degree of regularity, the principle of multiple assessment is rarely employed, and too often all the teachers in a system are evaluated with the same instrument--usually one in which they have had no input.

Delimitation

This study was limited to the effects of multiassessor feedback about high school teaching performances and to whether input from teachers about the feedback instrument has any effect on the attention given by the instructor to the feedback. Measurement of teacher performance was limited to student assessments of the performance of high

school teachers of English and Social Studies in Algona, Estherville, Ft. Dodge, and Spencer Iowa high schools. Forty-three teachers participated in the experiment and treatment was limited to four assessments collected every two weeks for a total of eight weeks. Half of the teachers from each school were given the results of their students' ratings. These teachers represented the treatment group. The remaining teachers received no information about their students' opinions and formed the control group.

Preparation of Instrument

The instrument prepared for the experiment was developed from a list of teacher behaviors identified by Madeline Hunter and a team of associates at the University of California at Los Angeles. The instrument desired was one that could be used and understood by high school students and that could meet the following requirements: 1) flexible enough to accommodate a variety of instructional techniques, 2) sufficiently diagnostic to identify instructional strengths and weaknesses, 3) standardized so that comparisons among instructors would be possible.

Hunter's list of teaching behaviors was chosen because many of the items she identified could be placed in categories thought to be important aspects of the productive, successful teacher. Five categories were selected and beneath each of these broad areas, two specific teaching behaviors were listed. Thus, each instrument used to get feedback had ten items upon which a teacher was rated.

In order to compare performances of teachers, the instrument had to contain some items that would be the same for all teachers helping with the experiment. Thus, one teaching behavior under each category was selected by the investigator and appeared on the instrument as numbers 1, 3, 5, 7, and 9. Following is a list of the determined categories and the teaching behaviors chosen by the investigator. Please see Appendix for complete instrument.

- I. THE TEACHER FOCUSES ON PERCEIVABLE OBJECTIVES
 1. Student's effort in class is directed toward a particular learning target of the teacher.
- II. THE TEACHER CHOOSES OBJECTIVES APPROPRIATE FOR THE LEARNER
 3. Student is tested for knowledge about the subject before any study is begun.
- III. THE TEACHER SEEKS EVIDENCE THAT VALIDATES PROGRESSION TOWARD ATTAINMENT OF CHOSEN OBJECTIVES
 5. Student's tests measure what was covered in class and assigned for homework.
- IV. THE TEACHER PUTS LEARNING THEORY TO PRACTICE
 7. Student gets comments from teacher as to how to improve assignments when they are returned.
- V. THE TEACHER ASSESSES TEACHING PERFORMANCE TO DETERMINE WHERE METHODS USED INTERFERED WITH STUDENTS' PROGRESS
 9. Student is given opportunity to critique assignments.

It was also necessary for the instrument to have the flexibility to accommodate a variety of instructional techniques and reflect priorities of the instructors. To do this, several teaching behaviors were listed under each identified category and placed on one sheet that was given to all teachers participating in the experiment. From this list

the teachers were asked to select one behavior under each category he/she thought to be most important to his/her teaching style. Thus, when completed, each teacher had helped design a personalized instrument which asked for judgments about teaching behaviors thought important to the instructor getting feedback. In other words, there were forty-three teachers participating in the experiment and there were forty-three different instruments used for obtaining feedback from students.

Selection of Teachers

After investigating the many teaching behaviors used by teachers, it was thought that differences in teaching methodologies might be reduced if the experimental subjects were instructing in the same area of study. At first it was decided to use only social studies teachers but that presented problems because of the area where the research was being conducted. The high school student population in northwest Iowa is relatively small, thus the number of social studies teachers available for the experiment was limited. For the sake of reducing travel (and expense) both social studies and English teachers were asked to help with the project. It was believed the teaching methodologies in these two disciplines are similar enough to make comparisons. It was possible to get enough subjects in four cities by using instructors from two disciplines.

Once school systems consented to help with the study via mailed responses, the investigator met with the respective teachers in each city and explained the purpose and nature of the experiment. All

teachers were visited before school began in the morning.

After the explanations were given to each teacher group and they had made choices about items wanted on their respective feedback instruments, separate instruments were typed and duplicated for each teacher. So that results for different instructors were not mixed, identification numbers were assigned. Each school was given a number and each teacher within that school was given a subnumber with the same prefix. Thus, Algona High School was one hundred and each teacher from Algona was one hundred and something. A form for each teacher from each school can be found in the Appendix.

To insure that approximately half the subject teachers in each system were getting feedback, every other number used on each school's list was picked to receive feedback. Because the teachers were met only once, there was no way for the investigator to know the individuals receiving feedback.

Collection of Data

Once the feedback instruments were printed for each teacher and arrangements were made for beginning the experiment, two instruction sheets were prepared--one for the teacher and another for the student. A sample of each is found in the Appendix. The teacher instruction sheet explained how the materials were to be distributed to the students, dates when feedback from students was to be asked for and when it would be returned to the teachers selected to receive it.

The process of getting 43 teachers in four school systems to ask

for feedback on the same day was not easy. Several times schedule changes and teacher memory slips caused extra trips or several hours delay so that the instruments could be presented to the students. Even with the extra miles, added hours waiting in school libraries and repeated reminders to school secretaries about dates for feedback, some teachers did not get all the feedback asked for in the project.

The experiment was designed to last for eight weeks with the first feedback from students to occur on Wednesday, February 16. Wednesday was chosen because it allowed for two days to remind teachers to ask for feedback and two days to get the data, process it through the computer at Iowa State University and to return information to teachers in the experimental group.

Every two weeks the same students in each of the teachers' classes were asked to fill out the instrument. Thus, after February 16, data were collected on March 2, March 16, and March 30. After all the data were collected, a thank-you letter was sent to all the teachers participating in the program. This letter also included four questions about the instrument and its value to the instructor. This information was sought to improve subsequent use of such an instrument and also to determine whether or not students understood the questions asked on the instrument. A copy of this letter can be found in the Appendix.

The method for recording the multiassessor responses was to ask each student to blacken the appropriate space on an ISU Form 5 optical-scan answer sheet. Responses were obtained for about ten statements printed on the teacher performance instrument. Each assessor assessed

their teacher on a five-letter, Likert-type scale with A representing the best performance and E being the worst.

Following data collection every two weeks, the raw data were transferred from the optical scan forms to computer cards. Each computer card contained subject identification, school group, time of the data collections, identifications of those in experimental and control groups and the rating for each question on the instrument. On the second collection, the sex of each assessor was indicated to determine whether or not male and female students rated teachers differently.

The student information sheet thanked the pupils for helping with the information and explained the procedures for filling out the answer sheet and directions about the rating scale used to determine the effectiveness of the teachers' methodologies. The Appendix contains a copy of this form.

Treatment

Data reduction and analysis of pre- and postscores were conducted at the computer center at Iowa State University. The data from the first assessment measures were processed so as to yield a mean and standard deviation for each item on the performance inventory created by each instructor. These statistics were interpreted by the investigator and then became the data used in the feedback as treatment to the experimental group.

Within three days after the information was received, it was processed at Iowa State and returned to those teachers selected to

receive feedback. It was assumed that teachers having trouble understanding the computer printouts would comment on the answer sheet or present questions to the investigator. No questions were received, but comments were written about the complexity of the language used in some of the questions on the instrument. When those comments appeared, an answer was provided and put with the next set of evaluation forms. After the first feedback, no more comments from teachers were received about the difficulty of the instrument.

Following data collection which occurred every two weeks for a period of eight weeks, the data were transferred from the optical scan forms to computer cards. The data were examined to be sure that all teachers participating in the experiment asked for and received the same number of feedback loops from their students. Out of the original forty-three teachers asked to help with the program, thirty-seven asked for feedback four times from their students and submitted this data for processing at Iowa State. Each computer card contained teacher identification, the mean score for each item for each data collection, the mean score for the five items chosen by the investigator and the mean score on the total instrument for each time the data was collected. On the second collection, students were asked to identify their sex in order to determine if sex of the students had any bearing on the way a particular question was answered.

Statistical Methods

The statistical methods used in this investigation were: Pearson product-moment correlation (r_{xy}) and pooled and separate variance t-tests. These statistical methods enabled the investigator to analyze, describe and draw inferences from this investigation.

In this experimental investigation, each teacher participating in the study became the unit of statistical analysis for determining the effectiveness of receiving student feedback on teacher performances. For the sake of analysis, each teacher was considered to be a replication of the experiment. Thus, when 19 teachers received feedback and became the treatment group, the experiment was thus replicated 19 times. The same could be said for the 18 teachers who received no feedback after data was collected every two weeks. Individual student scores were not identified other than by sex on the second report.

On questions selected by the teachers, there were small numbers of responses on certain items because few teachers chose them. If a significant difference occurred with one of these questions, it was judged to be less creditable for making assumptions about the entire sample, however the statistic could be helpful to the individual teacher when assessing his/her teaching performance. Thus, with this kind of problem occurring on teacher selected items, it became more plausible to look at differences among questions selected by the investigator when discussing significant changes by all participating instructors.

Generation of Means

Because the mean scores for each of the ten performances variables as perceived by students for each teacher were the units of statistical analysis used for this investigation, the method by which these mean scores were computed should be explained. The mean scores were computed by summing the students' scores for each teacher for each item on the instrument and then dividing these sums by the number of scores obtained for each question.

To get a mean score which could be treated as one score, for either the entire instrument or just the items selected by the investigator, the mean scores of each question were totaled for a given instructor or group and divided by the number of mean scores summed and the number of teachers in the group.

Pearson Product-moment Correlation

The Pearson product-moment correlation was used to express the relationships of the student opinions on each question between the first and last data collections for all the teachers participating in the investigation. Once these correlations were found, the mean of the items chosen by the investigator were correlated with the means of the items chosen by the teachers. The same statistical procedure was used on the second data collection to determine the relationship to the ten questions on the instrument on the basis of sex. The correlation coefficients of these groups are presented in Tables 1-5.

t-tests

The pooled and separate variance t-tests were used to determine the statistical significance for questions relating to the hypothesis in question. When using the t-test, the F test was used for testing the homogeneity of variance. If the F test did not reflect variance difference or, if the calculated F value was smaller than the tabled value, the t-test was computed by pooling the variances. When the calculated F value was larger than the tabled value, the t value was computed by using the separate variance t-test.

The criterion used to determine the significance of the statistics was that if one-half or more of the items related to the hypothesis were significant at the .05 level, the null hypothesis would be rejected. If that kind of condition did not occur, the null hypothesis remained tenable.

Even though sample sizes are not so important when using a t-test, special problems occurred in attempting to find whether the items selected by the teachers had discriminatory power. With only thirty-seven teachers participating in the project, some of the questions offered for selection were picked by only one or two persons in each group. When this occurred, the significance of the test score was meaningless because of the sample size. The most useful questions for determining discriminating power were those chosen by the investigator because there were student assessments for each of the teachers on each of these items and thus the sample was large enough to make the percentage meaningful.

CHAPTER IV. FINDINGS

To determine the usefulness of a multiassessor teacher performance feedback system, this experiment tested four hypotheses based on Madeline Hunter's five premises about teacher performance. The primary interest of the investigation was to ascertain whether students could provide the kind of feedback necessary to cause a positive change in the teaching methodologies of their instructors.

Forty-three teachers from four separate school districts were asked to help with the experiment and approximately 670 students made judgments about teaching performances. Only 37 of the 43 teachers who volunteered completed all four of the data collections. Eighteen of these were selected for the control group and received no student feedback and 19 teachers became the experimental group who were given the results of student assessments every two weeks.

Hypotheses

Ho₁ Teachers receiving feedback from students will not receive a more significant change in student ratings than teachers who were not given feedback.

To test the first hypothesis, three steps were necessary. First, it had to be established whether there was a significant difference in the mean assessments between the first and fourth data collections of the control group (no feedback). Table 1 provides the information for this test. Examination of the table revealed that four items on the feedback instrument solicited significant student response changes.

Table 1. Analysis of student mean responses for control teacher behaviors (correlations and t-tests)

Variable	Mean student responses ^a		r	t-value	Probability
	1st collect	4th collect			
<u>Investigator items</u>					
Learning target identified	2.38	2.14	.384	2.68	.016*
Pretests given	3.41	3.06	.794	3.43	.003*
Tests cover material in class	1.94	2.14	.799	-2.48	.024*
Teacher improvement comments	2.66	2.50	.606	1.21	.224
Student critiques assignments	2.73	2.63	.508	1.05	.308
<u>Teacher selected items</u>					
Question No. 6	2.07	2.09	.318	-0.27	.792
Question No. 7	2.22	2.18	.688	0.58	.560
Question No. 8	2.055	2.30	.873	-3.29	.044*
Question No. 9	2.04	2.09	.518	-0.42	.682
Question No. 10	2.52	2.38	.745	1.13	.273
Mean of 1st 5 items	2.62	2.49	.608	1.84	.084
Mean of last 5 items	2.18	2.21	.701	-0.44	.665

^aLikert scale used for obtaining means was from 1-5 with a rating of 1 (A) best, 2 (B) above average, 3 (C) average, 4 (D) below average, 5 (E) poor.

*Significant at .05 level.

Table 1. Analysis of student mean responses for control teacher behaviors (correlations and t-tests)

Variable	Mean student responses ^a		r	t-value	Probability
	1st collect	4th collect			
<u>Investigator items</u>					
Learning target identified	2.38	2.14	.384	2.68	.016*
Pretests given	3.41	3.06	.794	3.43	.003*
Tests cover material in class	1.94	2.14	.799	-2.48	.024*
Teacher improvement comments	2.66	2.50	.606	1.21	.224
Student critiques assignments	2.73	2.63	.508	1.05	.308
<u>Teacher selected items</u>					
Question No. 6	2.07	2.09	.318	-0.27	.792
Question No. 7	2.22	2.18	.688	0.58	.560
Question No. 8	2.055	2.30	.873	-3.29	.044*
Question No. 9	2.04	2.09	.518	-0.42	.682
Question No. 10	2.52	2.38	.745	1.13	.273
Mean of 1st 5 items	2.62	2.49	.608	1.84	.084
Mean of last 5 items	2.18	2.21	.701	-0.44	.665

^aLikert scale used for obtaining means was from 1-5 with a rating of 1 (A) best, 2 (B) above average, 3 (C) average, 4 (D) below average, 5 (E) poor.

*Significant at .05 level.

Eight weeks after the study began the responses for "identifying learning targets" and "giving pretests" indicated teacher performance had improved. However, a change in the opposite direction occurred with items three and eight. Students reported that teachers' tests were not doing as good a job of measuring work covered in class as had been done at the beginning of the study. Item eight represented a compilation of several teacher items and therefore was only useful in helping to compare items chosen by the teachers with those selected by the investigator.

When applying the criteria that one-half or more of the items needed to be significant ($p < .05$) to determine a significant change in the teachers' behavior, the evidence fell short. Teachers in the control group did not change their behavior significantly during the eight weeks of the experiment.

The second step necessary to test the first hypothesis was to find out if a significant change had occurred over the eight week period in the mean responses of the teachers who received feedback (experimental group). Table 2 revealed that only item three was significantly different ($p < .05$) and that the reported behaviors were less desirable at the end of the eight weeks than at the beginning of the experiment. Students in the experimental group agreed with those in the control group. Both groups judged the teachers' tests were doing a poorer job of measuring class work and homework than had been done earlier.

Once it had been determined that significant differences in mean assessments had been found in both the control and experimental groups,

Table 2. Analysis of student mean responses for experimental teacher behaviors (correlations and t-tests)

Variable	Mean student responses ^a		r	t-value	Prob-ability
	1st collect	4th collect			
<u>Investigator items</u>					
Learning target identified	2.40	2.41	.597	-0.10	.919
Pretests given	3.40	3.24	.589	1.01	.324
Tests cover material in class	2.05	2.27	.548	-2.30	.033*
Teacher improvement comments	2.91	2.88	.444	0.19	.848
Student critiques assignments	2.88	2.79	.342	0.60	.553
<u>Teacher selected items</u>					
Question No. 6	2.44	2.52	.685	-0.96	.352
Question No. 7	2.33	2.31	.375	0.15	.882
Question No. 8	2.29	2.42	.679	-1.11	.280
Question No. 9	2.15	2.30	.564	-1.45	.164
Question No. 10	2.77	2.64	.757	1.21	.243
Mean of 1st 5 items	2.73	2.72	.639	0.09	.929
Mean of last 5 items	2.40	2.44	.589	-0.47	.642

^aLikert scale used for obtaining means was from 1-5 with a rating of 1 (A) best, 2 (B) above average, 3 (C) average, 4 (D) below average, 5 (E) poor.

*Significant at .05 level.

the mean responses from the fourth data collection of the control group (teachers not receiving feedback) were compared with the responses of the experimental group (teachers receiving feedback). Table 3 provided the data to make the third step test of the first hypothesis. Inspection of the probabilities of mean differences at the $< .05$ level showed six of the 28 variables to be significantly correlated. Unfortunately, on items 7, 10, 19, 20 and 22 the significance of the statistic had little credibility because so few teachers had chosen those items for measurement. Of those five items, it should be noted that all but item 20 indicated less desirable final behavior assessments. In fact, 18 of the 28 items carried negative t-values. Thus, the teachers receiving feedback from students were given lower ratings about their terminal teaching performances more often than did those who had not received feedback and the first null hypothesis could not be rejected.

Ho₂ There will not be significant differences in the ratings of feedback items chosen by teachers and those fixed by the research design.

The hypothesis that teachers would be no more responsive to those items they selected than to those selected by the investigator was analyzed with data contained in Table 4. Each time the data were collected, the resulting differences were not significant. Thus, the second null hypothesis could not be rejected.

Ho₃ There will not be significant differences in feedback scores provided by male and female students.

The literature was generally silent as to whether ratings given teachers by students vary according to sex of the rater. Table 5 presents the information to test this hypothesis. The data were taken from

Table 3. Analysis of differences of student mean responses between teachers not receiving feedback (control group) and teachers receiving feedback (experimental group). Data taken from fourth (last) data collection

Variable	Mean student responses ^a			F value	t value	Probability	
	No. of cases	Control	No. of cases				Experimental
1. Learning target identified	(17)	2.135	(19)	2.413	1.64	-2.14	.039*
2. Pretests given	(17)	3.066	(19)	3.247	1.23	-0.73	.470
3. Tests cover material in class	(17)	2.138	(19)	2.279	1.21	-0.81	.425
4. Teacher improvement comments	(17)	2.512	(19)	2.880	1.05	-1.90	.066
5. Student critiques assignment	(17)	2.627	(19)	2.798	1.96	-1.06	.295
6. Student knows purpose of assignment	(8)	2.018	(7)	2.609	2.58	-1.85	.088
7. Student sees practicality of assignments	(1)	0.000	(4)	2.615	0.00	-4.69	.018*
8. Student sees that a.v. aids support class text and discussion	(5)	2.223	(3)	2.348	2.01	-0.34	.742
9. Student is presented with goals and objectives	(3)	1.966	(5)	2.450	3.40	-1.81	.167
10. Student given different ways to learn material	(1)	0.000	(7)	2.303	0.00	-5.71	.001*

11. Students with different abilities can each experience success	(5)	1.916	(3)	2.364	2.36	-1.13	.340
12. Student finds material challenging, but capable of being mastered with effort	(11)	2.237	(9)	2.300	1.82	-0.30	.764
13. Student's daily work provides good feedback about student's progress	(2)	2.169	(5)	2.734	1.31	-1.93	.112
14. Student is encouraged to ask questions	(13)	2.249	(13)	2.282	2.19	-0.19	.853
15. Student's assignments graded quickly and student corrects errors	(2)	2.860	(1)	0.000	0.00	2.05	.289
16. Student's seating arrangement changes to facilitate assigned task	(0)	0.000	(0)	0.000	0.00	0.00	.500
17. Student encouraged to set some goals	(0)	0.000	(1)	0.000	0.00	0.00	.500
18. Student feels comfortable in classroom	(6)	2.129	(6)	2.058	3.35	0.32	.760

^aLikert scale used for obtaining means was from 1-5 with a rating of 1 (A) best, 2 (B) above average, 3 (C) average, 4 (D) below average, 5 (E) poor.

*Significant at .05 level.

Table 3 (Continued)

Variable	Mean student responses				F value	t value	Prob-ability
	No. of cases	Control	No. of cases	Experi-mental			
19. Student treated fairly by teacher	(4)	1.774	(6)	2.577	3.47	-3.05	.016*
20. Student is given time to master an idea before moving to new idea	(2)	2.808	(1)	0.000	0.00	16.13	.039*
21. Student given assignments commensurate with ability	(1)	0.000	(0)	0.000	0.00	0.00	.500
22. Students are not allowed to distract each other when studying	(0)	0.000	(2)	2.356	0.00	-22.23	.029*
23. Student given extra help from teacher when having trouble with assignments	(1)	0.000	(1)	0.000	0.00	0.00	.500
24. Student gets chance to re-view difficult concepts	(0)	0.000	(0)	0.000	0.00	0.00	.500
25. Student gets good comments from teacher when assign-ment is done well	(3)	2.062	(2)	1.825	4.43	0.74	.595
26. Student's progress recorded accurately	(8)	2.062	(9)	2.210	1.46	-0.82	.426

27. Student given opportunity to critique instructional performance	(3)	2.599	(6)	3.457	2.15	-1.81	.113
28. Student who appears bored questioned to find out why	(6)	2.656	(4)	2.420	1.61	0.97	.362
Mean of investigator selected items	(17)	2.496	(19)	2.723	1.61	-1.60	.118
Mean of teacher selected items	(17)	2.203	(19)	2.442	1.32	-1.75	.089

Table 4. Variable (mean of 5 items selected by the investigator and mean of items selected by teachers), means, coefficients of correlation, discriminating value of t-test, and probability for between the two groups. Data taken from last collection

Variable	Means ^a		r	t-values	Prob-ability
	No. of data collection				
Investigator items	1	2.681	.752	9.28	.000
Teacher items	1	2.296			
Investigator items	2	2.635	.853	8.64	.000
Teacher items	2	2.320			
Investigator items	3	2.584	.879	7.89	.000
Teacher items	3	2.301			
Investigator items	4	2.616	.886	8.40	.000
Teacher items	4	2.329			

^aLikert scale used for obtaining means was from 1-5 with a rating of 1 (A) best, 2 (B) above average, 3 (C) average, 4 (D) below average, 5 (E) poor.

Table 5. Analysis of student mean responses for each item on the instrument as reported by males and females on the second data collection

Variable	Mean ^a		F value	t value	Prob-ability
	Male (No. of cases)	Female (No. of cases)			
1. Learning target identified	2.29 (33)	2.26 (33)	1.11	0.20	.842
2. Pretests given	3.29 (33)	3.34 (33)	1.01	-0.24	.812
3. Tests cover material in class	2.09 (33)	1.96 (33)	1.33	1.00	.320
4. Teacher improvement comments	2.64 (33)	2.81 (33)	1.27	-0.92	.362
5. Student critiques assignment	2.81 (33)	2.81 (33)	1.12	-0.01	.995
6. Student knows purpose of assignment	2.43 (14)	2.36 (14)	1.06	0.31	.756
7. Student sees practicality of assignment	2.69 (5)	2.10 (5)	4.10	2.34	.047*
8. Student sees a.v. aids support class text and discussion	1.93 (8)	1.91 (8)	1.96	0.04	.966
9. Student is presented with goals and objectives	2.54 (6)	1.89 (6)	1.44	2.62	.025*
10. Student given different ways to learn material	2.72 (6)	2.55 (6)	9.19	0.72	.490

11. Students with different abilities can each have success	2.18 (9)	2.45 (9)	1.73	-1.01	.329
12. Student finds material challenging but capable of being mastered with effort	2.22 (18)	2.26 (18)	1.62	-0.19	.854
13. Student's daily work provides good feedback about student's progress	2.85 (5)	2.57 (5)	2.88	0.57	.582
14. Student encouraged to ask questions	2.12 (26)	2.09 (26)	1.62	0.25	.807
15. Student's assignment graded quickly and student corrects errors	2.80 (2)	2.45 (2)	3.10	0.38	.743
16. Student's seating arrangements changed to facilitate assigned task	0.00 (0)	0.00 (0)	0.00	0.00	.500
17. Student encouraged to set goals	2.41 (2)	2.05 (2)	5.03	0.42	.714
18. Student feels comfortable in classroom	1.80 (12)	1.97 (12)	1.17	-0.90	.379

^aLikert scale used for obtaining means was from 1-5 with a rating of 1 (A) best, 2 (B) above average, 3 (C) average, 4 (D) below average, 5 (E) poor.

*Significant at .05 level.

Table 5 (Continued)

Variable	Mean		F value	t value	Prob- ability
	Male (No. of cases)	Female (No. of cases)			
19. Student treated fairly by teacher	2.39 (9)	2.26 (9)	1.79	0.48	.640
20. Student given time to master an idea before moving to new idea	3.01 (3)	2.82 (3)	1.57	0.39	.715
21. Student given assignment commensurate with ability	0.01 (1)	0.00 (1)	0.00	0.00	.500
22. Students are not allowed to distract each other when studying	0.00 (0)	0.00 (0)	0.00	0.00	.500
23. Student is given extra help from teacher when having trouble with assignment	1.75 (2)	2.22 (2)	1.38	-7.23	.019*
24. Student gets chance to re- view difficult concepts	0.00 (0)	0.00 (0)	0.00	0.00	.500
25. Student gets good comments from teacher when assign- ment is done well	2.25 (4)	2.48 (4)	14.34	-0.44	.676

26. Student's progress is recorded accurately	2.02 (14)	2.08 (14)	2.43	-0.33	.742
27. Student given opportunity to critique instructor's performance	3.13 (8)	2.91 (8)	1.17	0.46	.652
28. Student who appears bored is questioned to find out why	2.83 (11)	2.96 (11)	1.21	-0.47	.645

the second collection (two weeks after the experiment began). Only three items produced a significant difference by sex. On item seven the girls gave teachers a higher mean rating for "being practical about student assignments" and on item nine, girls rated their teachers higher regarding presentations of goals and objectives. The young men rated teachers higher regarding giving extra help (item 23). These differences were not important, however, because of sample size for the items. Thus, with only three items of 28 showing significant differences and with those three having only scant teacher requests, the third null hypothesis could not be rejected.

Ho₄ Teachers who have participated in a student feedback experiment will be more receptive to using such a tool to help them improve their teaching effectiveness.

Teachers were queried about the value of students' assessments two weeks after the last data collection. A copy of the questionnaire is found in the Appendix. Table 6 lists the questions and presents the response percentages. Even though 68 percent of the teachers believed that students understood the questions on the instrument and a similar percentage of teachers receiving feedback said they understood the computer printout, a majority of teachers did not agree that a student feedback system would be an effective tool for assessing effective teaching behavior. This reluctance from teachers to endorse such a system might be explained by looking at the teachers' response to question three (do you believe students were conscientious in rating your teaching methodologies). Only 60 percent of the teachers responded with a yes to question three. Thus, with only half the teachers

Table 6. Analysis of teacher responses to questions about the validity of student assessments of teaching methodologies

Questions	Totals		Percentage	
	<u>Yes</u>	<u>No</u>	<u>Yes</u>	<u>No</u>
1. Do you think the majority of students in your class who provided feedback understood the questions in the instrument?	28	13	68	32
2. If you were provided with feedback, did you understand the data on the computer printout?	15	7	68	32
3. Do you believe the students were conscientious in rating your teaching methodologies?	27	18	60	40
4. Do you believe a system for getting student feedback similar to the one used in this study would be a useful tool for helping you assess your teaching effectiveness?	20	20	50	50
5. If you did <u>not</u> receive feedback, do you believe a student feedback system would be a useful tool for helping you assess your teaching effectiveness?	7	7	50	50

indicating that the teacher feedback system would be a useful tool and qualifying their answers with restrictive comments, it was not possible to conclude that teachers, who had participated in a feedback experiment were more favorable to using such a tool.

CHAPTER V. SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

The purpose of this investigation was to develop and test a student feedback system that could be used by teachers to help them assess their teaching performances. More specifically, the experiment was designed to find out whether:

1. Teaching behavior could be modified after receiving repetitive feedback about selected teaching objectives.
2. Teachers would react more favorably to items they chose for the instrument as opposed to those chosen by the investigator.
3. Teachers would accept students as reliable assessors of their teaching performance.
4. Teacher ratings would vary by the sex of the student respondent.

Summary

Literature was searched for information relating to the above questions; for information about other student assessment programs; and for evaluation instruments that would adequately measure teaching performances and that could be manipulated by high school students. Professor Hunter's research at U.C.L.A. provided philosophical principles upon which the instrument was designed and the teacher performance studies at Purdue University provided the model which gave teachers choice about evaluation items.

High school English and social studies teachers from Algona,

Estherville, Ft. Dodge and Spencer, Iowa, were assigned to a feedback (experimental), or no feedback (control) group. Each of these teachers chose five items from the twenty-three offered and their selected items along with five items chosen by the investigator became the instruments through which data for testing the hypotheses were obtained. Feedback was obtained every two weeks for eight weeks and at the end of each two weeks, teachers in the experimental group were given the results of the student assessments.

Correlation coefficients of the mean scores for each response period were analyzed by t-tests. Results indicated that repetitive feedback from students did not cause teachers to change their behaviors significantly contrary to Chatterjee, Daw and Gage's equilibrium theory (In Gage et al., 14). Nor was their evidence to support the Purdue University approach which proposes that teachers will respond more favorably to evaluation items which they have chosen as opposed to those chosen by the evaluator.

Teacher opinionnaires about the study indicated that instructors were reluctant to accept students as reliable assessors of their teaching performances; many doubted that students understood the questions.

The literature search produced little evidence that the sex of the student would have little, if any, effect upon the kind of assessments made. Such was the case in this investigation. Male and female students did not differ significantly in the assessment of their teachers.

Conclusions

With so much attention directed at improving teacher effectiveness, the pilot experiment to test whether student feedback would make a positive significant behavioral change appeared warranted. Literature revealed that measurement of teaching performance is difficult at best, but when attempted, the most pragmatic approach is to involve more than one rater per teacher. Considerable research about different kinds of groups who have evaluated teachers indicated that student ratings are at least as valid and reliable as those done by administrators and peer teachers. In the investigation presented, students consistently rated their teachers' effectiveness as average to high average. Few extremes in student responses were found among any of the participating classes. Thus, despite the fact that teachers felt students should have been more serious about the ratings, the young people did report consistent assessments every two weeks about their teachers' behavior.

The major thrust of the experiment was to determine if feedback from students would alter teaching performances.

Even though the study supports the idea that students can make appropriate, meaningful teacher assessments, statistically it was not shown that information given to teachers about their assessments resulted in changed teaching behaviors. Teachers who received feedback did not change significantly from those who were given no information. A close examination of the study indicated that the ratings received may have been influenced as much by the design of the instrument as by the feedback itself. For example, teachers reported that the language used in

the rating instrument included professional terms that were not understood by all the students and that the form for recording ratings was cumbersome and confusing. It was also believed that a longer time span between assessment was necessary. Teachers felt the rapidity of student ratings was a part of the reason they appeared flippant with their responses.

So much of the literature which discussed evaluation instruments stated that teachers were more receptive toward the evaluation process if they were allowed an opportunity to help select the items included on the instrument. This idea was tested in the study by comparing the positive rating changes for feedback items chosen by the teachers with those items fixed by the research design.

Teachers in the experiment reacted no more favorably toward items they chose than toward those selected by the investigator. These statistics however must be viewed in light of the procedures used in accumulating the data. Teachers insisted after the investigation was completed that the computer printout sheet used to provide information about the students' feedback was difficult to interpret, and there was not enough time provided for the investigator to explain the implications of the data. Thus, part of the reason for so little difference between assessments of teacher and investigator items may have been due to the fact that teachers could not properly interpret the computer sheets.

Even though the literature does not indicate that males or females responses differently to various teaching styles, the study included an

investigation of this question. Data from male and female students was compared to find out whether the sex of the students would make a difference as to the kind of assessments made. The finding from the study supports the literature. The male students did not rate their instructors significantly different than did their female counterparts.

Teacher evaluations are commonly done with the idea in mind to improve the teaching performance of the instructor. The literature explains, however, that when administrators are involved in the evaluation process and possible teacher termination becomes part of the reason for such observations, that some teachers are reluctant to endorse the process. This study explored the acceptance teachers would have toward an evaluation design which incorporated the idea of multiple assessment for improvement of instruction, but did not carry with it the threat of termination.

Results of a question given to participant instructors revealed that teachers who had received feedback appeared no more attracted to such an evaluation program than those who had not received the feedback. Again, however, their responses may have been directed as much at the design as to the process itself. For instance, several teachers remarked that the Likert scale used by students to assess the teachers was too similar to the normal grading scale used by teachers in the classroom, thus intimating they were not fond of being "graded" by students. It must also be remembered that teachers felt that students were careless with the assessments and that the computer printouts were difficult for many to interpret. Thus the problems of the design should

be a factor considered when interpreting the acceptance of this student feedback system and the statistical responses given to the hypothesis presented.

Limitations

The field study portion of this research was limited to collecting data from high school English and social studies teachers in four medium-sized school districts in northwest Iowa. These schools were selected because they were geographically close enough so that data could be collected every two weeks and returned in time to meet the time constraints of the experiment. Neither the school districts nor the teachers within districts were randomly selected for the experiment. The schools were selected because they volunteered and were large enough to have both control and experimental groups in each building.

An additional limitation was shrinkage; six of the 43 teachers asked to participate in the experiment did not respond to every data collection and thus information they submitted was not included in the findings of the study.

While this investigation may help in the search for finding ways to improve the quality of teaching performances, there are weaknesses in the design which need to be mentioned:

1. The language used in the questions on the instrument, while pedagogically sound, was confusing to the students.
2. Perhaps the investigation should have centered on only one of the two major hypotheses--to examine the effects of

feedback and flexibility of feedback instrument on teaching performance may have diluted the differences and confused analysis and interpretation.

3. Conducting the research in four separate communities created span of control problems and reduced the amount of time spent with teachers to clarify the feedback data.
4. The Likert-type rating scale used by students (with A, B, C) for rating teachers was disliked by teachers and may have unduly reduced participation.
5. The form used by students to mark assessments was too bulky and complex. Its flexibility for measuring different kinds of items confused the students.
6. The system employed no measure of accountability for teachers. They felt no pressure to take the instrument seriously.
7. The system for allowing teachers to choose feedback questions was not structured enough to get consistent measures on such items.

Recommendations for Practice

Evaluation of teachers is a time consuming process and one which invariably invites problems for both administrators and instructors. Administrators dedicated to improving teaching performances may find the following recommendations gleaned from this study helpful.

1. The study has shown that senior high students can provide reliable data about teaching performances. However, in order

for such assessments to have a positive impact on teaching procedures, the instructors must be convinced that student feedback will be helpful and that administrators will pay attention to the student ratings.

2. Instruments used by students should only measure teaching performance skills and teachers should have input in selecting the items. The criteria selected for the instrument should describe the objectives sought for the course and the vocabulary used on the instrument should be understood by all the students. Research will provide lists of teaching performance skills and discussions between teachers and students about the instrument will reduce vocabulary problems.
3. Rating sheets should incorporate a numerical scale rather than alphabetical marks to reduce the anxiety level of teachers who fear being "graded". The same sheets should be designed for ease of maneuverability and scoring. The standardized computer scoring sheets which are deliberately designed for flexible usage do not work well for such measurement. Again, the use of the scoring sheet should be thoroughly discussed before any assessments are made.
4. Student assessments should be made at least once each six weeks for each class taught and teachers should record the scores. It is imperative that teachers collect the information regularly if growth is to be measured accurately.
5. Data collected and recorded by teachers shall become the basis

for teacher-administrator conferences held at the end of each semester. It is most important that it be stressed that such conferences are held for improvement of instruction and are not to be used for deciding terminations. Those kinds of conferences should be confined to observations and evaluations made by the administrator. It should also be stressed that extreme ratings or singular rating assessments will be ignored and that patterns of ratings will be the basis for discussion.

6. Administrators should let the teachers initiate the areas for discussion and once the conference is completed there should be mutual agreement about the job targets to be written that will improve the deficiencies identified.
7. In the spring and fall of each year, students and teachers should have an opportunity to discuss the reasons for such a system. It would be the goal of such discussions to impress on students that if the rating process is taken seriously, they will have an active part in helping teachers improve performances. The reward to students is obvious; if teachers do a better job, they should have better learning opportunities.

Administrators who would make use of these recommendations would be employing the principle of multiassessments and thus would increase the possibility for reliable measurement. Teachers at the same time would have more regular assessments than are normally provided by administrators and could initiate changes which would enhance teaching performances.

Incorporating students as part of the total evaluation process would help principals improve teacher productivity and thus improve the learning environment of the district.

Recommendations for Further Research

Although the feedback system used in this study did indicate that it is possible to use students to assess teaching performances there were few differences found between teachers who received feedback and those who did not. If a similar study were to be repeated several changes in the design are recommended to tighten the controls for measurement.

1. The scope of the investigation should be reduced. It would be easier to study separately the effect of student feedback on teacher methodology and keep the idea of teachers being more responsive to evaluation items they choose as a different experiment. Studying both together created a problem in determining the effect each had on teaching behavior.
2. The time period between feedbacks should be lengthened - two weeks was too short, and was blamed by teachers for the careless manner in which some students treated the assessments.
3. The vocabulary used on the evaluation instruments should be re-evaluated. Teachers complained that too many students did not understand some of the terms considered "educational jargon". A bigger effort should be made to use words that would be part of students' vocabulary.

4. The answer sheet used for recording student responses was too large and had too many other blanks that were not necessary for the experiment. Even though this was a preprinted form and easily adapted to the computer, it was confusing to the student and thus increased the problems associated with student assessments.
5. The experiment should be confined to one school district; particularly if it is to be conducted by one person. The time it took to collect data from the different cities presented many difficulties in getting the data to the computer center at I.S.U. and back to the teachers in three days when each of the cities were at least 100 miles from the University.
6. More time needs to be given to teachers after each feedback is received. There were not opportunities to visit about the student comments.
7. Computer printouts about student feedback were confusing to teachers. Many did not understand them and thus were not affected by the data. Data from these sheets should be translated to simpler readouts before being presented to instructors.
8. The experiment should be implemented by administrators responsible for evaluating the teachers using student assessment. It appeared that teachers would have paid closer attention to the feedback if the information received had been thought important to improving their teaching careers. As it was, most

saw the experiment as helping another person get a doctorate and some credibility was automatically lost.

Discussion

This investigation was concerned with the effects and relationships of student feedback on teaching performances. The investigation, however, offered no support to the equilibrium theory proposed by Chatterjee, Daw and Gage (In Gage et al., 14). When 18 of the 28 items used for rating teachers were lower at the end of the experiment for teachers who received no feedback than for those who did not, there was little evidence to support the idea that teachers will change to improve their images with their students.

Part of the explanation for such findings may be due to the time frame in which the experiment was conducted. The first reporting period was at the beginning of the second semester when teachers and students were fresh and held great expectations for each other. Unfortunately, the last report came eight weeks later--perhaps too close to the time when student grade reports were due.

Although teachers in the experiment generally thought that students could manipulate a multiple-assessor feedback system, they still resisted the idea of having their teaching assessed by students. Teachers were also quick to point out that poorer students had trouble with the vocabulary of the instrument and that some of the items on the instrument did not measure what was being attempted in the classroom. The following quotes taken from the teacher opinionnaire help illustrate

this point:

When asked if they thought the majority of students in their classes understood the questions on the instrument--

Not the first time--afterwards I explained the questions before they responded.

I tried to explain the questions to them. Some of the questions did not apply to the literature class, including pretesting.

But only because of the class (doing the feedback). I used the top 5% of the senior class types. A "lower" age or group would not have understood them.

The class which did the feedback is comprised of average and below average students. Many of them found it difficult to understand the questions. I found I had to provide much explanation and several examples before they could answer the questions.

Considering the group mean scores for all questions, it was not obvious that students did not understand the instrument, or that they were irresponsible with their observations. As a whole, students were not extreme in their judgments about their instructors. On a five-point scale, the mean of the five items chosen by the students was 2.6 for the teachers who had not received feedback and 2.5 for those who had. The letter attached to such a score, would be a C+ or a B-. In fact, there was only one item on the instrument in which teachers as a group were rated in the low C or high D category. That question pertained to testing for knowledge of the subject before any study was begun. On the last data collection, the mean for that question from the nonfeedback group was 3.41. The feedback group received 3.06. For the most part, students saw their teachers doing an average job in the classroom, but indicated there was room for improvement.

Even though an attempt was made to allow for personalizing the instrument by utilizing some of the ideas from Purdue University's "cafeteria" approach to teaching evaluations, the teachers helping with the study still felt the instrument was too restrictive. If the teachers had taken part in designing the original items from which they were asked to make selections, they might have been more satisfied with what was offered to them.

Perhaps the best explanation of why all teachers in the experiment rated a C+ to B- on each item was that, generally speaking, the social studies and language arts teachers participating in the experiment did not perform in ways congruent with Madeline Hunter's postulates for high teacher productivity. A measure of student gains would have been valuable to test this possibility.

After the study was completed it became obvious that parts of the instrument may have been confusing to both students and teachers. However, the biggest obstacle to gaining an accurate assessment of the effectiveness of the student feedback system appeared to be the time necessary for presenting the ideas of the study to both teachers and students before the observations began and the time needed but not spent with teachers after feedback was provided. Even though visits were made with all of the teachers participating in the experiment before the program began, it now is clear that not enough time was given to clarifying the purpose of the experiment or the nature of the feedback instrument. Sixty-eight percent of the teachers answering the questionnaire reported that they understood the data on the printout

sheet, but the comments on many of the questionnaires indicated they would have liked more clarification. One teacher wrote:

"I studied the first one (computer printout) and I could not understand it. I did not really study the others. It seemed to me the printout was very complicated. Couldn't it have been simplified?"

If more time and money had been available for the experiment, conferences with each of the teachers after each feedback was received would have been desirable. It might have made a difference in the attitude the teachers had about the study. Too often it appeared that the teachers viewed the experiment as helping a stranger complete a doctoral thesis and not as an opportunity to improve or as part of teaching accountability program. If the goal of improved teacher performance is to be reached, it appears that the student feedback system can help meet the desire for objectivity in teacher observations, but unless time is provided for the teacher to discuss the implications of such feedback with either the students or a supervisor, much of the value of the feedback will be lost.

In summary, even though this pilot venture did not indicate that significant differences in teaching methodologies occurred when students gave feedback to teachers, the study does suggest that such a system is workable and, with modifications, could be useful in helping teachers identify strengths and weaknesses of their teaching behavior.

BIBLIOGRAPHY

1. Aleamoni, L. M., and Spencer, R. E. The Illinois course evaluation questionnaire: A description of its development and a report of some of its results. *Educational and Psychological Measurement* 33 (1973): 669-84.
2. Biddle, Bruce J., and Ellena, William J., eds. *Contemporary research on teacher effectiveness*. New York: Holt, Rinehart, and Winston, 1964.
3. Blackburn, R. T. Letter on Research, teaching and faculty fate. *Science* 92 (1971): 172, 1082.
4. Blum, M. L. An investigation of the relation existing between students' grades and their rating of instructors' ability to teach. *Journal of Educational Psychology* 27 (1936): 217-21.
5. Bush, Robert N. *The teacher-pupil relationship*. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1954.
6. Centra, J. A. Effectiveness of student feedback in modifying college instruction. *Journal of Educational Psychology* 65 (1973): 395-401.
7. Costin, F., Greenough, W. T., and Menges, R. T. Student ratings of college teaching: Reliability, validity usefulness. *Review of Educational Research* 41 (December 1971): 511-35.
8. Derry, J. O., and the Staff of the Measurement and Research Center. *The cafeteria system: A new approach to course and instructor evaluation*. Institutional Research Bulletin 74-1 (Measurement and Research Center, Purdue University, Lafayette, Indiana), 1974.
9. Doyle, Kenneth O. *Student evaluation of instruction*. Lexington, Massachusetts: D. C. Heath and Company, 1975.
10. Eastridge, Harry E. Student evaluation and teacher performance. *NASSB Bulletin* 60, No. 401 (September 1976): 48-54.
11. Eickhoff, Robert L. *Perceived effects and relationships of multi-assessor feedback on modifying educator performance behavior*. Unpublished Ph.D. thesis. Iowa State University, Ames, Iowa, 1975.

12. Final Recommendations, subsection devoted to committee's recommendations that deal with professional personnel in education. Iowa Governor's Educational Advisory Board, Des Moines, 1970.
13. Frey, Peter W. The ongoing debate: Student evaluation of teaching. *Change* 6 (February 1974): 47-48, 64.
14. Gage, N. L., Runkel, P. J., and Chatterjee, B. B. Changing teacher behavior through feedback from pupils: An application of equilibrium theory. In W. W. Charters and N. L. Gage, eds. *Readings in the social psychology of education*. Boston: Allyn and Bacon, 1963. pp. 173-181.
15. Herman, Jerry H. Developing a staff evaluation program. *NASSP Bulletin* 60, No. 401 (September 1976): 8-14.
16. Hidlebaugh, Everett J. A model for developing a teacher performance evaluation system: A multiple-appraiser approach. Unpublished Ph.D. thesis. Iowa State University, Ames, Iowa, 1973.
17. Holmes, D. S. Effects of grades and disconfirmed grade expectancies on students' evaluations of their instructor. *Journal of Educational Psychology* 63 (1972): 130-133.
18. Hudelson, E. The validity of student rating of instructors. *School and Society* 72 (1951): 265-266.
19. Hunter, Madeline. Appraising teaching performance: One approach. *The National Elementary Principal* 53 (February 1973): 60-62.
20. Isaacson, R. L., McKeachie, W. J., Milholland, J. E., Lin, Y. G., Hofeller, M., Baerwaldt, J. W., and Zinn, K. L. Dimensions of student evaluations of teaching. *Journal of Educational Psychology* 55, No. 6 (1964): 344-351.
21. Johnson, Henry C. The assessment of teaching in higher education: A critical retrospect and a proposal. *Higher Education* 4 (1975): 173-199.
22. Krathwohl, D. R., Bloom, B. S., and Masia, B. B. *Taxonomy of educational objectives: The classification of educational goals. Handbook 2 Affective Domain*. New York: McKay, 1964.
23. Kratz, H. E. Characteristics of the best teachers are recognized by children. *Pedagogical Seminary* 6 (1896): 413-418.
24. McKeachie, Wilbert J. Student rating of faculty. *American Association of University Professors Bulletin* 35 (Winter 1969): 439-444.

25. Menne, John W. Teacher evaluation: Performance or effectiveness? Paper distributed at teacher evaluation conference at Iowa State University, 26-27 November, 1972. (Mimeographed)
26. Miller, M. T. Instructor attitudes toward, and their use of, student ratings of teacher. *Journal of Educational Psychology* 62 (1971): 235-239.
27. Morsh, J. E., and Wilder, E. W. Identifying the effective instructor. A review of quantitative studies, 1900-1952. *AFPTRC Research Bulletin, Air Force Personnel and Training Research Center, Lackland AFB, San Antonio, Texas, 1954.*
28. Moss, Ronnie L. Teacher evaluation. *School and Community* 57 (January 1971): 16-17.
29. National Institute for Education. Basic Research Grants Announcements. Washington, D.C.: HEW, Spring 1976. pp. 2-7.
30. News release to superintendents and school board presidents. Iowa Association of School Boards, Des Moines, December 3, 1976.
31. Popham, James W. Designing teacher evaluation systems. Los Angeles, California: The Instructional Objectives Exchange, 1971.
32. Remmers, H. H., Shock, N. W., and Kelley, E. L. An empirical study of the validity of the Spearman Brown formula as applied to the Purdue Rating Scale. *Journal of Educational Psychology* 18 (1927): 187-195.
33. Rice, Paul R., and Manatt, Richard P. Module one teacher performance evaluation. Department of Education, Iowa State University, 1976.
34. Root, A. R. Student ratings of teachers. *Journal of Higher Education* 2 (1931): 311-315.
35. School Laws of Iowa, 1976.
36. Shavelson, Richard and Dempsey-Atwood, Nancy. Generalization of measures of teaching behavior. *Review of Educational Research* 46, No. 4 (1976): 553-611.
37. Shaw, Jane. Students evaluate teachers and it works. *Nations Schools* 91, No. 4 (April 1973): 49-53.
38. Stocker, Joseph. Accountability and the classroom teacher. *Today's Education* 59 (March 1971): 41-49.

39. Swanson, Richard A., and Sisson, David J. The development, evaluation and utilization of a department faculty appraisal system. *Journal of Industrial Teacher Education* 9 (January 1971): 64-79.
40. Thorndike, Robert L., and Allen, Elizabeth H. *Measurement and evaluation in psychology and education*. New York: John Wiley and Sons, Inc., 1971.
41. Voeks, Virginia W., and French, Grace M. Are student ratings of teacher affected by grades? *Journal of Higher Education* 31 (June 1960): 330-334.
42. Wise, Charles N. Student ratings of teachers: A perspective for speech communication. *Western Speech* 37 (Summer 1973): 196-203.
43. Zelenak, Mel J., and Snider, Bill C. Teachers don't resent evaluation--if it's for improvement of instruction. *Phi Delta Kappan* 55 (April 1974): 570-571.

ACKNOWLEDGMENTS

My sincere appreciation goes to Dr. Richard P. Manatt, Dr. Robert Thomas and Dr. Anton Netusil for their help in seeing me through this arduous, but necessary task.

I am also especially appreciative of the patience and encouragement given to me by my wife Donna and my three children - Lisa, Bob and Stephanie.

APPENDIX

IOWA STATE
UNIVERSITY

Telephone 515-294-5450

January 10, 1977

Dear

As a superintendent and former principal and teacher I am vitally interested in methods for improving instruction. Thus, I chose as my dissertation topic to determine whether or not teaching performances can be improved when teachers receive an increased amount of feedback about their behavior in the classroom.

There is good evidence to validate that students can accurately assess teaching behavior. At no cost to your school system, I would like to field test an instrument by which students can give teachers feedback about their behavior in classes. The items chosen for students to rate would be selected in part by the teacher observed thus reflecting the teacher's priorities as to what is important for the students to experience.

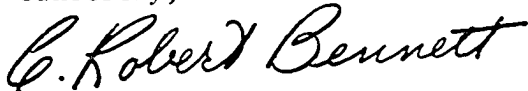
The instrument is not designed to be used as part of the formal evaluation of an instructor. In fact, the feedback from students would only be seen by the teachers participating in the study. Teachers are often reluctant to allow students to assess their performances if the results are to be part of the administrative plan to determine the instructor's worth in the system. However, if the instrument is used to give the teacher an answer to the question about how he or she is doing in a very personal and private way, I believe the hesitancy to participate will be reduced. Naturally, if the experiment does help the teacher improve in the classroom, the whole system benefits too.

My plan is to use high school social studies and English teachers and students. Forty teachers will be necessary to conduct the experiment which should begin the second week of the second semester and last for the following eight weeks. I would be happy to visit with you and your staff about further details of the study if they and you

would consider participating. I have deliberately tried to keep the experimental design simple and short so as not to create much extra work for the participants.

I would certainly appreciate your cooperation in helping me research this question. Enclosed is a self-addressed card for your reply.

Sincerely,

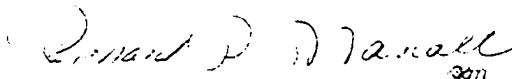


C. Robert Bennett, Superintendent
Ayrshire Community School District

sc

Enclosure

PS We will provide copies of our refined instrument for your use at the end of our research--Please say yes?



Richard P. Manatt
Professor of Education
Section Leader
Educational Administration

STUDENT FEEDBACK ASSESSMENT

FORM A-1

INSTRUCTOR INFORMATION:

Thank you for participating in this experiment to assess the effect of receiving feedback from students on teacher performance in the classroom. There is good evidence to validate that students can accurately assess teaching behavior if the items rated by the students reflect the teacher's priorities as to what is important for the students to experience.

The five categories chosen to reflect a teacher's performance were based on a research effort by the staff of the University Elementary School at the University of California at Los Angeles under the direction of Dr. Madeline Hunter. One item from each of these categories has been selected by the researcher so that all teachers in the experiment can be measured on some common items.

So that the feedback instrument will reflect some personal priorities that you believe important to your students' development, please select and circle one item from the attached list in each of the categories. The item you circle plus the one selected by the researcher will be used to compose the instrument upon which your students will provide feedback about your teaching performance.

The feedback from this instrument is not designed to be used as part of the formal evaluation of a teacher. In fact, the feedback from students is to be seen only by the instructors participating in the study.

STUDENT FEEDBACK ASSESSMENT

FORM A-2

TEACHER: Select one item beneath each category about which you would like to be rated.

-
- I. THE TEACHER FOCUSES ON PERCEIVABLE OBJECTIVES
- A. Student knows the purpose of the assignment
 - B. Student sees the practicality of assignments
 - C. Student understands that audio-visual aids and assignments support what is read in the text and discussed in class
 - D. Student is presented with goals and objectives for the class
- II. THE TEACHER CHOOSES OBJECTIVES APPROPRIATE FOR THE LEARNER
- A. Student is given different ways to learn the material
 - B. Students with different abilities can each have success in the class
 - C. Student finds material challenging but possible to do well if an effort is put forth
- III. THE TEACHER SEEKS EVIDENCE THAT VALIDATES PROGRESSION TOWARD ATTAINMENT OF CHOSEN OBJECTIVES
- A. Student is given daily assignments that provide good feedback to teacher about the student's progress
 - B. Student is encouraged to ask questions when he/she does not understand assignments
 - C. Student's assignments are graded quickly and errors must be corrected by the student
- IV. THE TEACHER PUTS LEARNING THEORY TO PRACTICE
- A. Student's seating arrangement changes to facilitate the task assigned

-
- B. Student is encouraged to set some goals for the class
 - C. Student feels comfortable in the classroom
 - D. Student is treated fairly by the teacher
 - E. Student is given time to master an idea before moving to a new idea
 - F. Student is provided with assignments commensurate with his/her ability
 - G. Students are not allowed to distract each other when studying
 - H. Student is given extra help from teacher when he/she is having trouble with assignments
 - I. Student gets chance to review difficult concepts
 - J. Student gets good comments from teacher when he/she does assignment well
- V. THE TEACHER ASSESSES TEACHING PERFORMANCE TO DETERMINE WHERE METHODS USED INTERFERED WITH STUDENTS' PROGRESS
- A. Student's progress is recorded accurately
 - B. Student is given opportunity to critique instructor's performance
 - C. Student who appears bored is questioned to find out why

STUDENT FEEDBACK ASSESSMENT

FORM A-2 - 1

TEACHER: Select one item beneath each category about which you would like to be rated.

I. THE TEACHER FOCUSES ON PERCEIVABLE OBJECTIVES

- A. Student knows the purpose of the assignment (15)¹
- B. Student sees the practicality of assignments (6)
- C. Student understands that audio-visual aids and assignments support what is read in the text and discussed in class (8)
- D. Student is presented with goals and objectives for the class (8)

II. THE TEACHER CHOOSES OBJECTIVES APPROPRIATE FOR THE LEARNER

- A. Student is given different ways to learn the material (8)
- B. Students with different abilities can each have success in the class (9)
- C. Student finds material challenging but possible to do well if an effort is put forth (20)

III. THE TEACHER SEEKS EVIDENCE THAT VALIDATES PROGRESSION TOWARD ATTAINMENT OF CHOSEN OBJECTIVES

- A. Student is given daily assignments that provide good feedback to teacher about the student's progress (7)
- B. Student is encouraged to ask questions when he/she does not understand assignments (27)
- C. Student's assignments are graded quickly and errors must be corrected by the student (3)

¹Number in parentheses are the number of teachers that selected the items to be part of their rating scale.

IV. THE TEACHER PUTS LEARNING THEORY TO PRACTICE

- A. Student's seating arrangement changes to facilitate the task assigned (3)
- B. Student is encouraged to set some goals for the class (2)
- C. Student feels comfortable in the classroom (12)
- D. Student is treated fairly by the teacher (10)
- E. Student is given time to master an idea before moving to a new idea (3)
- F. Student is provided with assignments commensurate with his/her ability (1)
- G. Students are not allowed to distract each other when studying (2)
- H. Student is given extra help from teacher when he/she is having trouble with assignments (2)
- I. Student gets chance to review difficult concepts (0)
- J. Student gets good comments from teacher when he/she does assignment well (5)

V. THE TEACHER ASSESSES TEACHING PERFORMANCE TO DETERMINE WHERE METHODS USED INTERFERED WITH STUDENTS' PROGRESS

- A. Student's progress is recorded accurately (17)
- B. Student is given opportunity to critique instructor's performance (9)
- C. Student who appears bored is questioned to find out why (11)

STUDENT FEEDBACK ASSESSMENT**FORM A-3**

STUDENT INFORMATION:

Thank you for participating in this experiment to assess the effect of receiving feedback from students on teacher performance in the classroom. There is good evidence to validate that students can accurately assess teaching behavior if the items rated by the students reflect the teacher's priorities as to what is important for the students to experience.

Please use the following five-point scale to rate your instructor on the items listed on the rating sheet. Be as objective as possible in rating the items. Try to base your opinion on your actual experience of learning in the classroom and not on the personality of the teacher. The rating indicates how this teacher compares with all other instructors you have had in school.

A -- Best, B -- Above average, C -- Average, D -- Below average,
E -- Poor

STUDENT INFORMATION TO BE PUT ON THE ANSWER SHEET

- A. Do NOT put your name on the sheet
- B. Put the number of your teacher in block (B)
- C. Put the date in block (C)
- D. Put the name of the course and period in which you are taking the course in blocks (E) and (F)

STUDENT FEEDBACK ASSESSMENT

FORM A-4

FOLLOWING ITEMS WERE THOSE SELECTED BY THE RESEARCHER TO BE USED ON ALL INSTRUMENTS DEVELOPED BY THE TEACHERS SO THERE WOULD BE COMMON AREAS TO MEASURE ALL THOSE INVOLVED IN THE EXPERIMENT

- I. THE TEACHER FOCUSES ON PERCEIVABLE OBJECTIVES
 - A. Student's effort in class is directed toward a particular learning target of the teacher
- II. THE TEACHER CHOOSES OBJECTIVES APPROPRIATE FOR THE LEARNER
 - A. Student is tested for knowledge about the subject before any study is begun
- III. THE TEACHER SEEKS EVIDENCE THAT VALIDATES PROGRESSION TOWARD ATTAINMENT OF CHOSEN OBJECTIVES
 - A. Student's tests measure what was covered in class and assigned for homework
- IV. THE TEACHER PUTS LEARNING THEORY TO PRACTICE
 - A. Student gets comments from teacher as to how to improve assignments when they are returned
- V. THE TEACHER ASSESSES TEACHING PERFORMANCE TO DETERMINE WHERE METHODS USED INTERFERED WITH STUDENT'S LEARNING PROGRESS
 - A. Student is given opportunity to critique assignments

STUDENT FEEDBACK ASSESSMENT

FORM A-5 (101)

STUDENT: Please rate your teacher on the following items. Use a separate answer sheet according to directions on the student information sheet (Form A-3).

1. Student's effort in class is directed toward a particular learning target of the teacher.
2. Student knows the purpose of the assignment.
3. Student is tested for knowledge about the subject before any study is begun.
4. Student is given different ways to learn the material.
5. Student's tests measure what was covered in class and assigned for homework.
6. Student is encouraged to ask questions when he/she does not understand assignments.
7. Student gets comments from teacher as to how to improve assignments when they are returned.
8. Student feels comfortable in the classroom.
9. Student is given opportunity to critique assignments.
10. Student's progress is recorded accurately.

STUDENT FEEDBACK ASSESSMENTFORM A-6

TEACHER INSTRUCTIONS

Please distribute materials to students. Students are to put ratings on separate answer sheets that are provided. The student instruction sheet and evaluation instrument are to remain clipped together and reused each time feedback is asked for. Please collect them with the answer sheets, but only send the answer sheets to the principal's office.

Feedback from students should be done on the following dates:

Wednesday, February 16

Wednesday, March 2

Wednesday, March 16

Wednesday, March 30

Computer printouts showing results of student feedback will be returned to you on the Monday following the Wednesday the feedback was collected. If, for some reason, you know your class will not be meeting on the day scheduled for feedback, please ask for the feedback one day earlier than scheduled. This will insure that I get all the materials for the computer at the same time.

PLEASE NOTE:

Print on some pages is small
and indistinct. Filmed in the
best possible way.

UNIVERSITY MICROFILMS.

Ayrshire, Iowa
April 15, 1977

Dear _____,

Thank you for helping me gather data about what effects student feedback might have on teaching methodology. As soon as I have the material analyzed, I will send you a copy of the results.

I have one more favor to ask of you. Would you take a few minutes to answer the following questions:

1. Do you think the majority of students in your class who provided feedback understood the questions in the instrument?
Yes _____ No _____

Comments: _____

2. If you were provided with feedback, did you understand the data of the computer printout? Yes _____ No _____

Comments: _____

3. Do you believe the students were conscientious in rating your teaching methodologies? Yes _____ No _____

Comments: _____

4. Do you believe a system for getting student feedback similar to the one used in this study would be a useful tool for helping you assess your teaching effectiveness? Yes _____ No _____

Comments: _____

Thanks again for your help with this study.

Sincerely,

C. Robert Bennett

Please return this questionnaire in the enclosed envelope to the school secretary by the end of the school day.

This article will appear in The National Elementary School Principal.

A NEW DIMENSION IN TEACHER APPRAISAL

Madeline Hunter

How to assess the quality of teaching performance is a problem that has plagued educational administration and supervision since the beginning of time. Hundreds of studies have been conducted to determine a "good teacher." Most such studies have lamely concluded with such platitudinous attributes as "warm, accepting personality," "genuine interest in children," "respect for the dignity of the individual," and the like. Those same attributes would be assigned to "good" mothers, social workers, pediatricians or child psychologists, all of whom would be disasters if they were responsible for the daily teaching of a classroom of active, not always appropriately reactive, youngsters.

In spite of the vagueness and ambiguity of such personality attributes, most educators would endorse them as desirable teacher traits. It is when we focus on teacher performance in the classroom that we have as many opinions as there are appraisers. The amount of noise in the classroom, the appearance of the bulletin boards, attitudes of the parents, complaints of the custodian, test scores, involvement with the community, and grooming are some examples of countless criteria in common use, no one of which has ever been demonstrated to yield substantial correlation with the teaching performance that increases the probability of successful learning.

Recently instruments have been developed so one could observe more objectively what was "going on" in the classroom. These instruments never were intended to attribute values to certain behaviors but misunderstanding and misuse have created such value poles. "Interaction is good, teacher talking is bad." "Discussion is good, lecture is bad." "Discovery is good, being told is bad." These are examples of a few of the erroneous extrapolations which dismay the authors of those teacher observation instruments. As a result of such inappropriate extrapolations, many teachers fear that when they give an excellent set of directions, teach spelling effectively, or conduct an orderly fire drill they may be "flunking" inquiry, creativity, and democracy expectations.

For the last five years, the staff of the University Elementary School at the University of California at Los Angeles have been addressing a major portion of their research effort to developing answers to two questions: (1) What does a successful teacher need to be able to do? (2) How do you know when the teacher is "doing it" with a group of learners? The focus has been on the application of science to pedagogy, not on a teacher's personality development or false absolutes in terms of "shoulds" and "shouldn'ts." (This is not to say that personality and curriculum are not important, but maintains that without appropriate pedagogy, learning is not predictable nor is it responsive in the direction desired.)

To hypothesize answers to those two questions, countless teaching episodes were observed and analyzed. Out of this factoring of the

teaching process emerged eleven discrete but interrelated decisions which are carried into action in the teaching process. These identified decisions became the curriculum for a new program of teacher education.

The eleven components identified in successful teaching have been described in detail elsewhere¹ so are merely listed:

1. Deliberate separation of genuine educational constraints from the typical ethnic, financial, intellectual, or emotional excuses which constitute typical (and unfortunately acceptable) excuses for learning failure.
2. Determination of the learning a student has already achieved and what he is read to learn in terms of degree of difficulty (sequence) and complexity in the affective, cognitive, or psychomotor domain.
3. Identification of productive learning behavior for achievement of the learning task and for this particular learner.
4. Determination of an instructional objective with specificity in content and perceivability in terms of learner behavior.
5. Identification of the principles of learning that are relevant to the accomplishment of this instructional objective.
6. Adaptation of those principles to the particular situation and to each learner.
7. Incorporation of professional ethics as the teacher uses his attributes plus his competence in the specific learning area in order to enhance the learner's probability of successful accomplishment. "Use of self as instrument" is the phrase which describes this process. This is the only decision area in the teaching-learning process where, except for "knowing oneself," science has little at present to offer. Here the artist in the profession of teaching often uses the highly operational, but inarticulate, knowledge of intuition. Because much of such knowledge remains, at this point in time, inarticulate, it is

¹"The Teaching Process" and "The Learning Process." Madeline Hunter. Handbook for Teachers. Edited by Eli Seifman and Dwight Allen. Glenview, Ill.: Scott Foresman and Company, 1971. Chapters 3 la, 3 lb.

not systematically transmittable to all members of the profession.

8. Synthesis of the above decisions (one through seven) in the deliberate design or blueprint for teaching-learning episode. It is important to note that excellent teaching is like an iceberg. There is much more to it than appears on the surface. To maximize successful learning, all of decisions one through eight must consciously be made before the teacher-learner interaction takes place.
9. The actual teaching-learning interaction begins. With the first second of encounter, the teacher's observations of the learner augment and/or correct the decision-making process. It is this instantaneous use of current data that characterizes the true professional.
10. Evaluation is an integral and continuous aspect of the teaching-learning process, not merely a terminal function. Constant monitoring of the learner's progress yields essential current information which may modify or validate the teaching-learning design.
11. On the basis of these evaluative data collected during the teaching-learning process, the determination is made to (a) reteach, or (b) practice and extend, or (c) move on, or (d) "abandon ship" because, for some reason, the objective is not appropriate for the learner at this moment in time.

Next came the problem of assessing classroom performance of student teachers as well as teachers in the field to discover which of these decision areas were being implemented productively and which needed teaching or remediation. Out of this assessment practice grew the beginning of a fertile diagnostic instrument of teaching performance which gives the teacher, the teacher educator, and/or the supervisor concrete evidence of what the teacher has learned and can apply in the classroom as well as what still needs to be learned or is being applied incorrectly. Interestingly, the teacher educator must practice what he preaches, for he too must make the same eleven teaching decisions as he instructs teachers. It is the pervasiveness of these eleven decisions in all

teaching, regardless of school or class organization, content, size of instructional group, age or ability of the learner, or his past experience and ethnic derivation, that give this model of teaching its power and universal generalizability. It seems to be applicable to every teaching situation! As such, it holds promise for pre-service and in-service education, teacher evaluation, merit pay, and teacher re-education or dismissal, for it is based on defensible, objective evidence revealed by performance.

Appraisers of teaching performance must cite evidence from their observation supporting answers to the following five questions. All data must come from the current observation. Previous information from or about the teacher and children is not necessary.

1. Is there a perceivable objective? Is the teacher focused on a particular learning target, or is it an "egg on the wall" episode which is the descriptive phrase used when a "little bit of everything and the more the merrier" has been included in the segment of teaching.
2. Is the objective appropriate for this group of learners? This appraisal is based on evidence to validate that the objective has not already been achieved by the learners or is so difficult there is little possibility of its being achieved.
3. Was the objective achieved? Evidence is sought that validates progression toward achievement or attainment of the objective. If for valid reasons the teacher abandons the objective, the appraisers observe the progression of learners toward a new objective.
4. What did the teacher do that facilitated learning? Evidence in this category comes from the appropriate application of principles of learning such as investing content with meaning that is related to these particular learners, massing practice on new material, reinforcing appropriate behavior, giving precise and specific knowledge of results, etc. The Theory into

Practice publications¹ are used to supply the principles of learning to which the appraisers are sensitized.

5. What did the teacher do that interfered with learning? Evidence of violations (albeit unintentional) of the principles of learning cited above constitute such interference. Examples might be unintentional reinforcement of tattling, inappropriate practice, not evaluating attainment of one learning step before proceeding to the next, etc.

Most recently, appraisers have been trained to observe teaching episodes to validate changes in teacher behavior as a result of retraining in this "component model of teaching."² These appraisals are made from video taped teaching episodes. Without prior knowledge of whether they were observing a "before" or "after" tape, in most cases the change in teaching behavior was so dramatic that within the first minute of observation the appraiser recognized classroom performance of a retrained teacher.

Currently UCLA specialists in evaluation and the staff of the University Elementary School are joining their resources to quantify the appraiser's judgements, make more explicit the criteria for evidence being used, establish reliability of the instrument, and train a cadre of appraisers for public schools and for the many other research projects dealing with the improvement of the process of instruction.

While this instrument of teacher appraisal is still in an infant

¹Motivation Theory for Teachers, 1967; Retention Theory for Teachers, 1967; Reinforcement Theory for Teachers, 1967; Teach More--Faster!, 1969; Teach for Transfer, 1971. Madeline Hunter. TIP Publications, P. O. Box 514, El Segundo, California 90245.

²Professional Development Center (P.D.C.) Long Beach Project. Long Beach Unified School District, 701 Locust Avenue, Long Beach, California.

stage, it is a robust infant and holds great promise for growth that will identify and validate those teaching decisions and behaviors that make the difference between success or frustration for the universe of learners.

As such, the production of successful teachers, continued training, the retraining or remediation of those in the field, and the accountability of schools for learning become a possibility in reality rather than an educator's or taxpayer's fantasy.

January 1972