This is to certify that the

thesis entitled

The Relative Impact of Two Forms
of Assessment Data Feedback on a
Teacher's Perceived Strengths
and Needs for Improvement

presented by

Donald George Wilson

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Education

Major professor

Date January 20, 1981

THE RELATIVE IMPACT OF TWO FORMS OF ASSESSMENT DATA FEEDBACK ON A TEACHER'S PERCEIVED STRENGTHS AND NEEDS FOR IMPROVEMENT

By

Donald George Wilson

A DISSERTATION

Submitted to

Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Teacher Education

1981

ABSTRACT

THE RELATIVE IMPACT OF TWO FORMS OF ASSESSMENT DATA FEEDBACK ON A TEACHER'S PERCEIVED STRENGTHS AND NEEDS FOR IMPROVEMENT

By

Donald George Wilson

Purpose

The main purpose of this study was to investigate (1) the extent to which self perceptions fashioned by a training program that focused on self assessment remained stable when teachers received feedback from others, and (2) whether this feedback had a stronger influence on teacher perceptions when the data were processed by the self determining teacher alone, or when the data were analyzed and reported to teachers by an external evaluator. The secondary purpose was to initiate a formative evaluation of the professional development program in which the study was imbedded.

Procedure

Prior to their involvement in the experimental phase of the investigation, all the teachers participated in a graduate

course that focused on systematic self evaluation. During the course which followed, these teachers were assigned to one of three groups, two experimental and one control. The members of one experimental group were asked to make an oral report based on their analysis of feedback they received from three colleagues. The members of the second experimental group had the same sort of data analyzed for them and were given the results in a report prepared by an external evaluator. The control group received no feedback at all.

The main instrument used throughout the study was the Teacher Behavior Survey (TBS). The participants collaborated in the development of this instrument during the initial course. Responses to the TBS served as pretest, posttest, and delayed posttest. The instrument was also used by the three assessors to rate teacher performances. A second instrument, an attitude scale, was given to the members of the experimental groups to find out what they felt were the relative strengths and shortcomings of the two professional courses in which this study was set.

Findings

The three most significant findings with respect to the primary purpose of the study were:

1. Even when teachers have previously rigorously assessed their own classroom performance, their self perceptions are apt to change as a result of receiving feedback from others.

2. Feedback data that are externally analyzed and reported appear to have a somewhat stronger influence on self perceptions than comparable data that are analyzed and reported on by the teachers themselves.
3. Teachers change their self perceptions in accord with the direction and magnitude of the difference between their original self perceptions and the feedback they receive.

Implications

These findings have far reaching implications for teacher education. They clearly indicate that even when teachers have undergone a systematic process of self analysis, feedback from others has a powerful influence on their self perceptions. The results therefore suggest that self analysis by itself is an inadequate basis for decisions about one's professional development. Consequently, a development process based on needs assessment or other forms of self diagnosis must give attention to the provision of feedback from others.

The results also suggest that feedback data, once collected, are most apt to have an impact on self perceptions when these data are analyzed and reported by an external evaluator. It is important to consider that the small number of participants and tightly controlled, almost mechanical, format in which feedback was provided by the external evaluator

severely restricted the power of this experimental treatment. It is therefore likely that receiving feedback data that has been processed by an external evaluator will have an even more powerful influence on one's self perceptions under more normal circumstances. Collectively, these and other findings confirm that, in their traditional role as external evaluators, supervisors and other teacher educators may have a greater impact on teacher perception and behavior than we have previously assumed.

DEDICATION

To Betty

and

Karin, Tim, Steven and Mark

and

to my Mother and Father

ACKNOWLEDGEMENTS

My sincerest thanks to the members of my committee: Dr. Ben Bohnhorst, Dr. Sheila Fitzgerald, and especially to Dr. Robert Hatfield, who inspired me to go in this direction, and to Dr. Donald Freeman, who was a challenging and most generous chairman.

A special word of appreciation to Dr. John Cragun, who was solid in support and encouragement.

Thanks also to Ms. Carol Blumberg who was a patient, kind and resourceful consultant.

I wish finally to acknowledge my great debt to the unselfish and cooperative teachers who participated in this study, and to their colleagues who served as their assessors.

TABLE OF CONTENTS

APPENDICES

LIST OF TABLES

CHAPTER I

Introduction

The contention that teachers should be responsible for their own professional development has gained considerable momentum in recent years. Those who accept this position argue that teachers should decide what their own strengths and deficiencies are, as well as how their developmental needs are to be met. Furthermore, teachers individually should develop and monitor the life-long program for their professional development. In short, teachers should determine and direct their own professional development.

In order to reach these goals, many conditions must be satisfied. Teachers need information about themselves and the availability of developmental resources, and they need self motivation, commitment and opportunity. But their first need is a firm and deep foundation on which to raise up the edifice of their lifelong professional development. The cornerstone of that foundation is the best insights they can possibly get into their own teaching performance--their strengths and their needs for improvement.

If teachers can arrive at sound and firmly rooted insights solely by analyzing their own behavior, then self assessment would be a direct and economical way of

establishing the essential base for professional decision making. If, however, the self perceptions provided by even the most rigorous self analysis techniques are shifting and uncertain, then self assessment by itself would be inadequate. It is possible, for example, that when self perceptions are formed in this way, they can be substantially influenced by the opinion of others. If this is true, the base for professional decision making should include systematic feedback from others in addition to rigorous self assessment.

Purpose of the Investigation

The main purpose of this study, then, is to investigate whether self perceptions fashioned by prolonged and detailed self assessment remain firm and relatively unchanged or are substantially altered by external feedback based on classroom observation. A second major question is whether feedback has a stronger influence on teacher self perceptions when it is analyzed and organized by the self determining teacher alone, or when it is analyzed and reported to the teachers by an external evaluator.

Importance of the Investigation

Because these questions are of fundamental importance in the professional education of teachers, they need to be

rigorously tested in a field-based experiment. The critical nature of the questions can be traced to several sources. First of all, it is important to recognize the magnitude of the professional development or in-service teacher education enterprise in the United States of America. Joyce, Howey and Yarger (1976) point out that although there is great discontent with it, in-service teacher education (ISTE) exists on an enormous scale. They estimate that there may be "as many as a quarter of a million persons in the United States who engage as instructors in some form of ISTE activity--this is about one instructor for every eight persons". This is indeed a huge investment, involving in all over two and a quarter million teachers.

Within this vast enterprise, the principle of teachers as self determiners of their own professional development has recently gained widespread popularity. Its popularity may be traced in large measure to the feeling of discontent about ISTE that Joyce et al. (1976) mentioned. As a result, teachers, individually, in groups, and through their professional associations have been demanding and getting more control over their own in-service training. The growth and spread of teachers centers have motivated and been motivated by the increasing desire of teachers to have more responsibility for their own professional training. One result of this movement for greater control has been greater emphasis on making ISTE responsive to teachers' on-the-job needs and to their emergent roles.

Such a teacher centered orientation has lead to the development of the teacher directed approach as one of the major modes of ISTE. In this approach:

> The teacher is seen as a self motivated craftsman or professional who is interested in maintaining the currency of his skills and knowledge. In this context, motivation and direction for learning come from the teacher, but certain enabling factors--time, money, educational resources--must be provided either by schools or higher education institutions. (Rubin, 1978).

Rubin (1978) also concedes that the continuing professional education of teachers was once held in disfavor, but maintains that it is now regarded as "a major force in school improvement". As a consequence, "a good many state departments of education have sought to organize a variety of exploratory programs," most of which place "heavy emphasis on individual need and are aimed at general professional enhancement..."

Rubin and the other contributors to his book strongly support the view that professional development should be growth oriented and determined by the needs of the individual teacher. As Rubin states:

> In authentic growth, each man directs his own evolution because when the incidental trappings of circumstance are shorn away, each man is responsible only to himself.

This sort of dedication to teacher self determination may appear a bit excessive, but shows the extent to which influential writers are committed to it.

Once we agree that teachers themselves are the decision makers, the ones to determine the course of their own self development, then their self perceptions become centrally important. For, the way they can best decide what their needs are, is to have accurate and firm perceptions about their own professional behaviors.

There is a closely related reason why accurate and firmly rooted self perceptions are important. Self perceptions generate self motivation and commitment, without which the best laid plans for professional development will fail. Combs (1965, 1976) presents a convincing case for self perceptions as the basic motivator for self actualization. Argyris (1970) is substantially in agreement with Combs' position. In Argyris' view, there are three fundamental requirements for professional education: (a) valid and useful information (b) informed and free choice and (c) internal commitment to that choice. He states that it is "accepted as axiomatic that valid and useful information is the foundation" of effective professional development (through a process which he describes as "intervention"). Thus, valid data or information about oneself is necessary not only so that one can have the best self perceptions on which to base decisions, but also because it generates self motivation and internal commitment to these freely chosen and firmly rooted decisions.

If one accepts the argument that decisions about professional development should be determined by teachers'

perceptions of their needs, critical questions follow: What constitutes an adequate source of these self perceptions? Can a rigorous process of self appraisal be designed and implemented that will result in relatively realistic and stable self perceptions? Or would self perceptions formed in this way prove to be malleable to the influence of the opinions of others?

Rubin (1978) would be likely to answer the second question in the affirmative. He believes that in the quest for quality performance, teacher self appraisal is perhaps "more useful than external self ratings." In-service education "must begin with perception, kindle the freedom and lust to change, then provide a method and support, and end in the confirmation of new born habits". Thus, for Rubin, the process of professional self determination and development begins with perceptions which arise from self analysis rather than from external evaluation.

In this tradition, Curwin and Fuhrmann (1975) give an example of a comprehensive model of do-it-yourself self assessment. Their book is addressed in the second person singular to the individual teacher. They encourage, but do not demand, participation in "support groups". The characteristic focus is on the individual teacher becoming more self aware and more self actualizing as a result of various self assessment activities.

Combs (1965) is directly opposed to this approach. He declares that "objective analysis of self...has been vastly

overrated as a device for personality change," and warns that it can "even be highly destructive." For Combs and other interactionist psychologists, one's self perceptions are formed and changed by social interaction with significant others. To base one's decision bearing self perceptions on self analysis and not on the careful evaluation of significant external opinion, is to build one's edifice on sand. On the other hand, if Combs is right, professional development models in teacher education ought to include the assessment of data from others. This approach suggested by Combs' theoretical position is more in keeping with traditional models of supervision and professional development.

The two sets of theoretical considerations discussed above underscore the major questions being asked in this investigation. First will teachers change their self perceptions, after rigorous self assessment, in response to feedback from others? Second, does externally analyzed and reported feedback cause greater changes in teacher self perceptions than self assessed feedback? Because of the clear relationship between these questions and the two theoretical positions, attempts to answer the questions should provide an incidental test of the theoretical positions.

Limitations of the Study

As the preceding analysis suggests, the fundamental questions that have been presented here have critical

implications for the professional education of teachers. They, therefore, demand precise consideration. This investigation sets out to test them in an exacting experiment. There are two salient features in the design of the investigation which were crafted to provide a severe test of the major research questions. First of all, in order to test the extent to which self perceptions, formed as a result of self assessment, can be changed by external feedback, subjects had to be selected who had undergone or could undergo a prolonged and detailed process of self assessment as a preliminary stage to the experiment. In the second place, in order to compare in a one-on-one situation the strength of the influence of self assessed feedback as against externally assessed feedback, a very exacting experimental procedure was followed that effectively resulted in a conservative test of the external feedback condition. To isolate and test the effect of external analysis and reporting, the role of the external evaluator was restricted to summarizing and reporting the data according to a prescribed schedule. Thus, whereas self analysis was allowed full play, a careful effort was made so that only the defined function of external evaluation was operational in this experiment.

If, under these stringent conditions, external feedback is shown to have a strong influence on teachers' self perceptions, or externally evaluated feedback proves more potent than self analyzed feedback, then these findings would be persausive indeed.

Another limitation of this study should be noted. Research suggests three possible levels of response to feedback which can be represented in the following model:

Feedback -> changes in self perceptions -> intention to change behavior -> actual changes in behavior. This study is limited to an investigation of the relationship between feedback and changes in self perceptions. It should be kept in mind that the broad concern underlying this study is the adequacy of self perceptions as a basis for making decisions about one's professional development.

Research Questions

A. Experimental Questions

The basic purpose of this study is to provide definitive answers to the four specific experimental questions which follow. These questions are directly derived from the general questions which have been the focus of the discussion to this point.

The four experimental questions are:

1. Does either of two forms of feedback--self-analyzed and reported (SAG), or externally analyzed and reported (EAG) cause any change at all in a teacher's self perception of the relative levels of his or her abilities?
2. Which of two forms of feedback (SAG or EAG) will yield the greatest changes in a teacher's self perceptions?

3. Is the magnitude of change in self perceptions resulting from assessment feedback a function of empirically established levels of agreement (high, low) between feedback received and original self perceptions?
4. Will the changes in self perceptions revealed in the posttest immediately following the experimental intervention (seven to 10 day interval) differ from corresponding changes in perceptions revealed in a delayed posttest administered approximately six weeks after the intervention?

B. Questions Concerned With Formative Evaluation

A secondary aim of this investigation is to gather formative data on the program of evaluation which the teacher will be involved in during this experiment. Three questions concerned with instructional outcomes related to this program serve as the focus of this secondary purpose. Answers to these questions should in turn provide useful information for teachers and teacher educators involved in similar programs of professional development.

The three questions are as follows:

1. What elements in the two term evaluation program do the participants find most valuable or least valuable?
2. How do they feel about participating in the program?

3. What are their perceptions of the feedback they receive?

Overview of the Dissertation

The remaining four chapters of this dissertation deal with (1) a review of the literature (2) the design of the investigation (3) the analysis of the data, and (4) conclusions and implications.

The literature review is concerned with a theoretical investigation of the three central topics of this study. First, what is the importance of feedback in teacher education? Second, what is the importance of self perceptions in teacher education? Third, what does previous research say in relation to the four experimental questions? With regard to the first topic, a number of field related teacher training models that make use of self analyzed and externally analyzed feedback will be described and discussed. A number of issues will be explored, including the need for external support in evaluative feedback situations. The second section explores the history, nature and structure of self perceptions, how they are formed and changed, and how consistent and stable they are. These conjectures will be described from the perspective of Combs' perceptual psychology and symbolic interactionist psychology. The concepts of the "looking-glass self" and "significant others" will be highlighted in the discussion. The third section looks at empirical research in relation to the four experimental questions that are raised in this study.

Chapter Three begins by outlining the self analysis process which participants underwent in preparation for the experimental stage of the investigation. The two treatments, involving feedback that was self analyzed and reported by one group of participants (SAG), and externally analyzed and reported for the other group (EAG) are then described. Other sections of this Chapter deal with the selection and assignment of participants to experimental groups, and the development of the three measurement instruments used in the study--including the Teacher Behavior Survey (TBS) used for gathering feedback and measuring participants' self perceptions. The nature of the dependent variables--including how they were determined, evidence of their internal and external validity, and the analyses used to determine if critical findings were statistically significant are also discussed.

In Chapter Four a summary and analysis of the data are presented. The statistical analyses that are reported include _t_-tests, analyses of variance tests, and correlational analyses. The data collected in response to the formative evaluation questions required only the computation of frequency scores and mean scores.

Chapter Five briefly summarizes the purpose and design of the investigation, the significant findings which resulted, and, finally, the implications of these results for teachers and teacher educators. Some far-reaching implications requiring further investigation are highlighted in this discussion.

CHAPTER TWO: REVIEW OF LITERATURE

INTRODUCTION

The present investigation is directly concerned with two sets of variables. The treatment variables are two different ways of giving feedback to teachers. The dependent variables, which are being measured to determine the effects of the treatment variables, are derived from differences between self perceptions on three separate occasions. These variables are described operationally in Chapter Three. In this chapter, the literature will be reviewed in relation to these variables. First, the importance of feedback in teacher education will be explored. Second, the literature in relation to the importance of self perceptions in teacher education will be reviewed. Third, answers to the four research questions regarding feedback and change in teacher self perceptions will be sought from the literature.

IMPORTANCE OF FEEDBACK IN TEACHER EDUCATION

Feedback is very important in teacher education, especially in the area of field experiences. Where change and development of teacher behavior is concerned, many educators would agree with Tuckman (1976) who writes that in

teacher education "the sine qua non of change is feedback."

The importance of feedback is attested to by the many ways in which feedback is utilized in various models of teacher field experience. A number of these uses will now be discussed.

A. Clinical Supervision

It would perhaps be difficult to think of any form of educational supervision in which some form of feedback is not given to the teacher. Whether in Cogan's eight stage model (Cogan, 1976), or in Goldhammer's five stage model (Goldhammer, 1969) or in any of the humanistic variations proposed by Sergiovanni (1975) or in any of Harris' models (1963), clinical supervision makes use of feeback, usually resulting from classroom observation, in a systematic cycle of supervision. Usually teacher and supervisor agree on goals and the specific focus of the observation, and the supervisor aims to give descriptive feedback within the limits set, after which plans for improvement and another cycle of supervision are made.

An important aspect of clinical supervision is the democratic, supportive, optimistic ambience in which it operates. This is to be deliberately nurtured by the supervisor, who should operate as a colleague or professional helper, not as a superior. More generally, Sergiovanni (1976)

describes the practial philosophy of clinical supervision in these words:

> In clinical supervision it is assumed, for example, that operationally the curriculum of the school is manifested in what teachers do day by day; that changes in curriculum and in teaching formats require changes in how teachers behave in classrooms; that supervisors are not teachers of teachers; that supervision is a process for which teachers and supervisors are both responsible; that the focus of supervision in on teacher strengths; that given the right climate teachers are willing to improve; that teachers have large reservoirs of talent--often unused; and that teachers want to increase competencies and to be successful for they seek and derive satisfaction for accomplishing challenging and important work.

This supportive element together with non-threatening, descriptive feedback are critical features in clinical supervision.

B. Peer Feedback

Peer feedback based on classroom observation is recommended by a number of writers. Blumberg (1974) is often cited in support of peer evaluation. In answer to the question, "Can teachers supervise teachers?" he reproduces two articles in support of the affirmative. The first is a research report by Amidon, Kies, and Palisi (1966) on the use of the Flanders System of interaction analysis in a peer group situation for giving feedback to teachers. He lists six useful rules for giving feedback that were devised by the faculty group in the study:

1. The person giving feedback describes rather than evaluates
2. Feedback is offered only in areas that are perceived as susceptible to change by the recipient
3. Feedback is given only upon request of the person whose teaching is being discussed.
4. Feedback is concerned with those aspects of teacher behavior that are characteristic of the teacher at the time that the discussion is taking place.
5. Feedback does not require a teacher to defend his personal opinion or feelings about the way in which he is teaching
6. Feedback is concerned with specific teaching acts, not with generalized interpretations.

Amidon concludes that group supervision offers advantages not found in the teacher-principal conference. Also, teachers do become more sensitized to verbal interaction, and the group activity seems to influence faculty interpersonal relationships, communications, goal setting and behavioral norms in a positive way.

The second article cited by Blumberg is by Abramson (1970) who provides a number of examples of the use of peer evaluation schools to point to its potential in staff improvement. He does caution that evaluators need to be trained.

Brophy (1979) recommends that teachers work in a group to observe one another's classrooms, and give feedback

and assistance to one another. Glassberg (1978) advocates peer supervision for student teachers, because it helps students to analyze their own and their colleagues' behavior, and in this way broadens their perspective from self to others with respect to the learning experience. Her experimental study gave evidence of signifcant gains in ego development. She stresses the need for a supportive environment to encourage reflection, and the integration of experiences in order to promote the higher levels of ego development.

Support for the efficacy of peer evaluation, or collegial evaluation, as they call it, also comes from Roper et al. (1976). They field tested a six step model involving goal setting, setting of criteria or standards, observing, appraising performance, communicating appraisals (providing feedback) and planning a program for improvement. They conclude that it is a most useful and flexible approach.

Bryant and Haack (1977) point to growing popularity and success of peer-centered systems of evaluation. They stress the need for training programs to define criteria for categorizing goals and behavior, to develop skills in data gathering and in giving feedback, and to gain expertise in planning improvement programs. Finally, Blumenthal (1977) reports favorably on the use of peer non evaluation feedback together with the use of video taped lessons.

A highly organized feedback system that relies centrally on peer feedback is discussed separately in the next section.

C. Tuckman Feedback System

The Teacher Feedback System developed by Tuckman (1976) is based on the Tuckman Teacher Feedback Form (TTFF). The form is a list of 28 adjectives, each describing a human element in teaching and paired with its opposite, for example, "original" is paired with "conventional," and "patient" with "impatient". Four factors were derived from the data collected by this list: Creativity, Dynamism, Organized Demeanor, and Warmth and Acceptance. The TTFF and the four factors are the basis for describing and giving feedback on a teacher's performance.

There are seven stages to the Feedback System:

1. Collect a team of volunteer teachers
2. Each teacher fills out the TTFF describing the "good" teacher
3. Teachers observe one another
4. Each teacher is given a consensus summary of his or her ratings. Teachers meet as a group to discuss feedback.
5. Teachers engage in strength training. They learn their deficiencies and find out what they can do to improve by giving one another specific ideas. Role playing also takes place.
6. Leadership training is essential for group leaders.
7. They observe one another a second time to determine whether there has been change, especially with reference to the four general factors.

Variations of this system have been tested in different environments. Spencer (1973) experimented with trainee teachers of vocational subjects and found that "warmth" and "acceptance" of the ideal teacher was rated much higher after the workshop, and that improvement in TTFF ratings was greater for the treatment group. Walencik (1973), substituting supervisors for peers, and high school students as the source of feedback using the TTFF, found results that supported the model. Student teachers who received TTFF feedback changed more than others. Kotula (1975) found that the group approach led to greater increase in creativity in the experimental group, but that inexplicably, the control group had the greater gains in warmth and acceptance.

Finally, Tuckman gives 12 specific rules for effective feedback. These include the following, as summarized by Brophy (1979). Feedback should:

1. Involve specific, concrete behaviors or characteristics
2. Be credible and presented with good intentions and in understandable terms
3. Include specific guidelines for changes, and
4. Lead to a commitment to initiate specific changes

D. Microteaching

Cooper and Allen (1971) report that since its inception in 1963, microteaching has become an established teaching training procedure in many colleges, universities,

and school districts, to the extent that a national survey showed 44 percent of all teacher education programs to be using some form of microteaching. They define microteaching in the following way:

> Defined most succinctly, microteaching is a teaching situation which is scaled down in terms of time and numbers of students. Usually, this has meant a 4-20 minute lesson involving from three to ten students. The lesson is scaled down to reduce some of the complexities of the teaching act, thus allowing the teacher to focus on selected aspects of teaching. Frequently, one microteaching episode includes teaching a lesson and immediate feedback on the teacher's effectiveness. This feedback may come from videotape or audiotape recordings, supervisors, pupils, colleagues, or from the teacher's self-perceptions. Some of the variable aspects of microteaching include lesson length, number of reteaches, the amount and kind of supervision, the use of videotape or audiotape recordings, and number and types of pupils.

Two points to note are the limited focus of microteaching, usually on one specific teaching skill, and the immediacy of feedback from a variety of sources. With respect to the feedback, microteaching does not require the supportive presence of a supervisor or colleague.

Cooper and Allen describe the classic process as having eight steps: (1) trainees receive instruction in particular skill, (2) trainees see a videotaped or filmed model of a teacher demonstrating the skill, (3) the model is discussed until trainees are clear as to the skill they will be practicing, (4) trainees teach a short lesson to a small number of students; this is usually videotaped or audiotaped, (5) usually a supervisor helps trainees analyze

lesson and discuss improvements; frequently colleagues assist; sometimes student feedback is utilized, (6) trainee replans lesson, (7) trainee reteaches lesson to different group of students, (8) repeat feedback process as in step 5. They describe the basic model as having a teach-critique/reteach-critique format.

Cooper and Allen give examples of the wide scale use of microteaching in a number of settings, naming the most comprehensive development of microteaching for inservice training as that conducted by the Far West Laboratory for Educational Research and Development. In addition to these descriptive studies, they also cite 10 experimental studies as evidence in support of the effectiveness of microteaching in changing and improving teaching behavior. In summarizing the research, they state: "The feedback dimension of microteaching is probably the crucial one in terms of changing the trainee's behavior." Their endorsement of the efficacy of feedback echoes the remark by Tuckman quoted at the beginning of this chapter: "The sine qua non of change is feedback."

Many other studies have been done which support and many which question various aspects of the microteaching model. But these are not the concern of this investigation. The point to be made is the central importance of feedback to the process-and how widely used the process is. A secondary point is that, whereas clinical supervision and peer evaluation emphasized the importance of a support

system--whether supervisory or peer, microteaching found this to be optional. The mechanical device (audio- or video-tape) removed one need for human interaction, and presented the possibility of the individual confronting the objective data of his or her performance. The teacher could be assisted in the analysis of the performance by a period of training or the provision of a guidebook. The effectiveness of this last aspect of microteaching has been tested, but as discussed later in this chapter, the evidence is inconclusive.

Following from this discussion of microteaching, two other forms of giving feedback to teachers will be discussed: first, two "computer assisted" models, and then Flanders Interaction Analysis. Both of these involve ways of analyzing teachers' behaviors and presenting feedback to teachers.

E. Computer Assisted Models

Two recent studies of the use of computer printouts as the source of feedback to teachers did not yield significant results (Froman and Owen, 1980; Trank, 1978). The first made use of student ratings, and the second made use of the Student Perception of Instruction (SPOT) survey.

But other ventures have been much more fruitful. As a result of computer program analysis and feedback using data collected by the Flanders Interaction Analysis teachers become more able to alter their behavior and more conscientious about planning (Hail, 1978).

A comprehensive, complex program, the Computer Assisted Teacher Training Program (CATTS) has been found effective with pre-service teachers of special education, where specific teaching skills were the aim (Semmel et al., 1976). The program is described in the System Document Manual (Semmel and Olson, 1977) which also presents extensive documentation for this system. CATTS is described "as a system capable of providing continuous, instantaneous, and/or delayed feedback of relevant teacher student interaction data in order to modify behavior through regulatory moves." The component systems include CATTS stations, data flow, data collection, data analysis, feedback, storage and retrieval, and an observation system trainng subsystem.

Chissom and Morris (1976) describe a system for the evaluation of student teachers employing automated data processing as an integral part of the system. It employed data gathered from four sources: public school pupils, student teachers, supervising teachers, and college supervisors. Feedback provided from the evaluation of the four sources was used to identify strengths and weaknesses of individual student teachers and cumulatively to evaluate the total student teaching program.

Based on a Faculty-Course Questionnaire (FCQ) evaluation instrument, a computerized feedback system was developed at the University of Colorado (Whetsone, 1974). Its chief advantage is that each instructor receives detailed and

comprehensive information that is individually specific and at the same time allows comparisons with various norm groups within the university.

Another computer assisted program makes use of a questionnaire (Pohlman, 1976). The Instructional Improvement Questionnaire (IIQ) has four parts designed to collect evaluative feedback from students: (1) the Class Characteristics Section; (2) the Instructor Evaluation Section; (3) the Course Evaluation Section; and (4) the Optional Item Section, which consists of 60 "response positions" that the instructor may use to record student responses to locally supplied items. The answer sheets are optically scanned and responses are coded and written on magnetic tape. A computer program analyzes the data and produces a printed report of the results.

Again, it is to be noted that these computer assisted programs make it possible to bypass a human evaluator and reporter. The feedback can be transmitted without supportive human interaction.

F. Flanders Interaction Analysis

Flanders Interaction Analysis is a system for analyzing teacher behavior originally designed by Flanders as a research tool, and later utilized by Amidon and Flanders (Amidon and Giamatteo, 1965; Amidon and Hunter, 1966; Amidon and Hough, 1967; Flanders, 1965, 1968, 1970) to provide teachers and student teachers feedback about their verbal

behavior in the classroom. The analysis can be done on data collected by audiotape, videotape or peer observation, and can involve either self analysis or external evaluation by peer or supervisor.

In itself, Flanders Verbal Interaction Analysis is not a process for giving feedback to a teacher; it is a system for analyzing teacher behavior. Both the procuring of the data and the reporting of results are not essentially determined by the Flanders analytical system. As reported in the section above, this system has even been used in combination with a computer program (Hail, 1978).

There is a great deal of research evidence to support the validity of the Flanders system (Amidon and Flanders, 1961; Rosenhine, 1971). But the reason for including it in this discussion is the experiential one, that it is widely used as a device for giving feedback to teachers, with or without the involvement of an external human agent.

To continue this section and to broaden its focus a little, two programs which make use of feedback and point to future directions in teacher education will be discussed. The first of these is concerned with the role of feedback in professional intervention; and the second is concerned with the development of specialist teachers in New York City and their use of feedback.

G. Professional Intervention

The work of Argyris (1970) and Argyris and Schon (1974) are concerned with the role and nature of intervention in

professional development and the need to respect and develop the client's professional autonomy. Many of their ideas are relevant to teacher education, and the place of feedback in the professional development of teachers.

In Argyris' view, an "intervenor":

> ...assists a (client) system to become more effective in problem solving, decision making and decision implementation in such a way that the system can continue to be increasingly effective in these activities and have decreasing need for the intervenor.

The successful intervention depends on the three primary tasks of the intervenor. He must ensure that the client is supplied with valid and useful information:

> First, it has been accepted as axiomatic that valid and useful information is the foundation for effective intervention. Valid information is that which describes the factors, plus their interrelationships, that create the problem for the client system.

Next the intervenor must ensure that the client is able to make a free choice. But in order to have a free choice, the client must have a cognitive map of what he wishes to do:

> Free and informed choice entails what Simon has called "satisficing," that is, selecting the alternative with the highest probability of succeeding, given some specified cost restraints. Free choice places the locus of decision making in the client system. Free choice makes it possible for the clients to remain responsible for their destiny. Through free choice the clients can maintain the autonomy of their system.

But Argyris has a practial reason for insisting on free choice for the client:

> Free choice is important because there are so many unknowns, and the interventionist wants the client to have as much willingness and motivation as possible to work on the problem. With high client motivation and commitment, several different methods for change can succeed.

The final stage of the intervention leads to internal commitment on the part of the client:

> Internal commitment means the course of action or choice has been internalized by each member so that he experiences a high degree of ownership and has a feeling of responsibility about the choice and its implications. Internal commitment means that the individual has reached the point where he is acting on the choice because it fulfills his own needs and sense of responsibility, as well as those of the system.

These three tasks defined by Argyris were translated into practice in the Master of Arts in Classroom Teaching (MACT) program described by Cragun and Wilson (1980) and Bradley et al. (1980)., first course in the MACT program called Classroom Analysis, incorporated the following three aims:

(1) to help candidates collect and assess valid and useful information about their classroom behavior related to students' needs and desired outcomes

(2) to help candidates who have constructed this "cognitive map" of their professional behavior make as free as possible decisions about courses, projects and learning experiences for the rest of their MACT program.

(3) as a result of (1)and (2), to facilitate the development of a high degree of internal commitment

to their own professional development in the short term of the 2-year MACT program and in the long term thereafter as a lifelong goal.

H. Individualized Professional Development Program

Many professional development programs claim to be individualized and are based on a combination of self assessment and the evaluation of external feedback, frequently referred to as a needs assessment. This is the basis, for example, of the needs assessment described by Cragun and Wilson, (1980) and Bradley et al. (1980), as part of the Master of Arts in Teaching Program (MACT).

Another sort of individualized professional development program was developed by the New York City Teachers Center (McDonald, 1980). The program has the following features:

1. It is an in-depth program which the teacher undertakes
2. It is developmental in character, that is, the activities of the program lead the teacher through a variety of progressive stages
3. It is built around the teacher's stated needs, so that it is not imposed by the Teacher Specialist.
4. It is revised on the basis of evaluations and analyses by both teacher and Teacher Specialist.

The program is built around the stated needs of the teacher. But once these needs are stated, the teacher specialist sets about collecting data from the teacher's classroom. Feedback is given in the context of detailed discussions of the data, with the goal that the teacher will make the diagnosis. Thus, one of the important functions of the Teacher Specialist is the skillful gathering and evaluation of data.

The Teacher Specialist is an innovation in this program. Each Teacher Specialist, defined as a teacher who specializes in teaching teachers, undergoes a carefully designed training program. The aims of the program are:

1. To provide Teacher Specialists with opportunities to gain further knowledge in curriculum and instruction in the elementary school
2. To provide instruction for the Specialists in developing programs for individual teachers and groups of teachers at the Teacher Center sites
3. To provide training experience for the Specialists to learn to:
   (a) diagnose teachers' needs and problems, plan interventions for and evaluate instructional development
   (b) help teachers to diagnose their own needs and problems, plan their own strategies and make their own evaluations of instructional development

(c) plan workshops based on the stated needs of teachers and needs perceived by the Specialists

(d) select, organize and use appropriate instructional materials with teachers in their Teacher Center programs.

Two features stand out in this program. Specialists have to be able to help teachers to disgnose their needs, but the Specialists must also be able to diagnose these needs and give feedback to the teachers that will help them in their diagnoses. They also have to give evaluative feedback. The second feature is the supportive role of the Specialist which is fundamental to the program. Specialist intervention, based on feedback and developmental programs, is seen in this model as an essential ingredient.

I. Change Agents in Education

To go a little further afield, it would be useful to look at feedback more broadly, in terms of the diffusion of knowledge and its use in schools in order to improve educational practice. Most schools apparently fail to utilize the fruits of much of the research that has been done, by not applying the knowledge or making use of the products that have been developed from it.

From a synthesis of findings across five recent studies in educational dissemination and change (Emrick and Peterson,

1978), five major generalizations were derived. They stress once again that in order for this kind of feedback to bring about change, "directed personal intervention" and "continuous personal participation" are essential "to root and sustain utilization." In short, in this context, effective dissemination requires a support system.

The five major generalizations are:

1. Meaningful change occurs as a process not an event
2. Directed personal intervention is by far the most potent technical support resource and may be a necessary condition for many forms of utilization
3. Continuous personal participation of the implementing staff is needed to firmly root and sustain utilization
4. Administrators occupy a crucial role in supporting the utilization process
5. Descriptive, instructional, and support materials are needed, particularly for utilization including organizational or instructional changes.

J. Summary

Feedback is used in a wide variety of teacher education models from clinical supervision to peer evaluation and from humanistically oriented to the mechanically or computer assisted forms of feedback from microteaching and classroom observation.

There can be no doubt about the rich contribution feedback has made in teacher education.

An interesting question is raised by the different levels of human support that characterize the models. On the one hand, clinical supervision in some forms cherishes an ideal of a helping relationship and peer supervision depends on the dynamic of group support and interaction. At the other extreme, some computer assisted systems have displaced the human support system, and microteaching does not require it. Instead, the dynamic is that of the self determining individual confronting the objective data fed back from his or her own performance.

IMPORTANCE OF SELF PERCEPTIONS IN TEACHER EDUCATION

A. Introduction

Ryans and Teacher Characteristics

Ryans (1960), in his monumental study of the characteristics of teachers, reveals that effective teachers differ in their perceptions of themselves, others, and their overall classroom behavior. Effective teachers, who were "high" with respect to overall classroom behavior, saw themselves as ambitious and having initiative and were more satisfied with regard to their emotional adjustment. They liked other people and were willing to participate in school and college social groups. They were extremely generous in their perceptions of the behavior and motives of other people. The

"low" group of teachers were less satisfied with their emotional adjustment, and were restricted and critical in their appraisals of others.

Ryans was interested only in describing these perceptions and correlating them with different kinds of teachers. But his work draws attention to the importance of teachers' perceptions and self-perceptions as features that distinguish between teachers.

Combs' Perceptual Psychology

Combs (1965) states that the basic concept of perceptual psychology is that the behavior of a person is the direct result of his field of perceptions at the moment of behaving. His behavior at any instant is the result of (1) how he sees himself, (2) how he sees the situation in which he is involved, and (3) the interrelations between the two. Thus, teacher behavior is a function of these three aspects of perception, the first of which is self perception. As Combs writes:

> of all the perceptions existing for an individual none are so important as those he has about himself.... It is the organization of seeing self that the modern psychologist calls the self concept. It represents the most important single influence affecting an individual's behavior.

Just as Ryans did, Combs notes that there are certain kinds of self perceptions associated with effective teaching. Basically these are rooted in the fundamental need all people share to be adequate, and in a willingness to see others as

always motivated to be and become as adequate as they can be in all situations.

Convinced of the importance of self and the self concept, Combs developed the idea of self as instrument. Thus, the effective teacher is "a unique human being who has learned to use himself effectively and efficiently to carry out his own and society's purposes in the education of others."

B. History and Nature of Self Perceptions/Self Concept

Webster (1974) gives a good account of the history of the self and self concept in modern times. It should be noted that "self concept" and "self perception" are being used interchangeably. There are two different ways in which the origin and development of the self have been viewed in sociological and psychological literature. On the one hand there is the view of the developmental self, which is roughly equivalent to a set of innate personal characteristics, or "personality," which develops by interacting with the possibilities inherent in one's environment. The crucial element in this school of thought is the "primacy of inborn needs and traits."

On the other hand, there is a view that focuses on the development of the social self. Those who share this view stress the importance of contact with others for the development of the personality, and are referred to as "environmentalists." There are at least two groups of environmentalists: behaviorists, who stress the assumption that

"human behavior may be studied or analyzed as if it were determined only by specifiable external influences." In this view, the individual is acted upon more than he acts. The second group, which this study will be most concerned with, employs the interactionist approach. The members of this group reject the idea of innate sources of behavior, and stress instead the effects of social interaction in shaping one's personality. The idea of "self concept" plays a unique role in their understanding of human behavior.

The interactionist approach owes its origins to the writings of Cooley (1964) and Mead (1934). Cooley put forward the idea of the "looking-glass self," which means as Webster (1974) says, that:

> a person's self concept is considered to be dependent on observing the reactions and opinions of others toward the individual. In other words, the personality is formed, not partially, but wholly through the experiences the individual has interacting with others.

But the individual is not wholly at the mercy of these interactions. He is self aware and aware of others, and can control his choices of action to some degree.

Cooley also contributed the notion of the "internalized other." This is a mental image of others which an individual develops as a result of social interaction. Mead modified this idea into the notion of the "generalized other". This concept described the organized community or social group which gives the individual his unified view of himself, and is central to Mead's idea of the development of

the self. Mead emphasizes the importance of social context as a source of self concept.

> The individual thinks about himself in categories determined by his social groups, and probably he also applies to himself standards of comparison derived from the range of variation he sees in others. (Webster, 1974).

Mead made another important contribution to the idea of the looking-glass self. He saw the self as having a semi-permanent structure, partially resistant to change--not needing to change with every change in an individual's environment.

Sullivan (1947) made a very important contribution to the development of the idea of the looking-glass self. He contributed the idea of the "significant other," first applied by him to parents, but since generalized to others who are instrumental in forming the individual's self concept. Important here is the notion that not all people equally influence the formation of one's self-concept. The individual judges some to be significant, and is much more influenced by them.

In summary, the idea of the self that influences this investigation belongs to the tradition of the looking-glass self, developed by the school of interactionists, headed by Cooley, Mead and Sullivan. These ideas include the central perspective that one's self concept is directly dependent on the opinions and actions of others. But the individual interprets the significance of others in the light of what he knows of them, and is not blindly influenced. Also, the individual develops in a social context--within which he

cperates a concept of the generalized other. His concept of himself is a semi-permanent structure, depending for change on his assessment of the significance of others and his own self image. This notion of the assessment of others is highlighted in Sullivan's concept of the significant other.

C. Structure of Self Perceptions/Self Concept

To this point, self concept and self perception have been used interchangeably. Combs, Richards and Richards (1976), however, do make a distinction. For them, the perceived self is the phenomenal self, and the core of this is the self concept:

> ...each person develops a large number of more or less discrete perceptions of self which he regards as characteristic of his being...these perceptions do not exist in the perceptual field as a simple enumeration of ways of seeing the self. Rather the concepts of self constitute an organization representing a person's own conception of himself in all his complexity.... This organization of all the ways a person has of seeing himself we have called the phenomenal or perceived self.

Combs et al. (1976) distinguish between the phenomenal self (all perceptions of self irrespective of their significance) and the self concept (those perceptions about self which seem most vital to the person himself). Crucial to both of these conceptions is the notion of organization.

Wylie (1961), in her survey of the research literature concerned with self concept, pointed out that phenomenological theorists like Combs are so called because they stress

the role of the conscious self concept in determining a person's behavior.

A more elaborate structure of the self concept is described by Kash and Borich, (1978). Kash and Borich bring together both poles of self psychology, and include the "developmental self," and the "performing self" and the idea of the "significant" and "salient" other in their theoretical framework. Their description of the self concept is derived from Allport's categories of the self (1961). Self concept consists of five senses of the self: the senses of (1) bodily self (2) self identity (3) self extension (self-as-doer) (4) self esteem, and (5) self image. All component senses are interactive, interrelated and interdependent.

Webster (1974) is not so interested in static pictures of how the self is constructed, but in the dynamic of how the structure of self is determined--how it is formed and maintained.

D. How Self Perceptions/Self Concepts Are Formed And Maintained

In discussing the history and nature of the self concept, the active role of the individual in reacting with significant others has been discussed in relation to the genesis of the self concept. Kash et al. (1978) also supports this point of view. Purkey (1978) also is of the opinion that no one is born with a self concept. For Purkey, the development and structure of self awareness is a "lifelong research project.":

> By experiencing the world through interactions with significant others, the developing person develops a theory of personal existence...we learn to identify ourselves both with categories (female, black, southerner, American) and with attributes (good, bad, valuable, able, unable, etc)... Gradually each person forges a self concept, complete with a complex hierarchy of attributes and categories.

Perhaps the main thing left in this discussion of the formation and maintenance of the self concept, is to indicate some empirical basis for these theoretical ideas. Webster (1974) cites three studies on the structure of the self concept by Miyamoto and Dornbusch (1956), Moore (1964) and Reeder, Donohue and Biblarz (1960), which agree that "self evaluation is a direct function of the (perceived) opinions of others, that such others may not be equally important to the individual, and that individuals may misperceive the others' opinions."

Other studies that deal with change and stability will be discussed later in this chapter. Some further theoretical ideas about change and self perception will be discussed in the next section.

E. How Self Perceptions Are Changed

Combs (1965) has very interesting ideas about the ways in which changes in self perception are and are not brought about. An individual changes his self perceptions not by being told to do so; nor by the "objective analysis of self" which has been "vastly overrated as a device for

personality change," and can even by highly destructive. His position on perceptual change is this:

> Changes in behavior, including changes in one's personality, are most effectively brought about, not by introspection and analysis, but through slow changes in perceptions about outside events and their relation to the self. To produce a change in a person's self requires some new experience which helps him to perceive himself in a new way.

Combs suggests that this sort of change can be brought about in three ways--(1) through some direct provision of experience (2) as a consequence of perceiving an event in a new perspective, and (3) through interaction following changed perception of others that is, a change in the perception of others causes them to behave in ways that change the self. For, as Combs concludes, "the self is learned from the looking glass held up for us by others."

The third way for changing self perceptions enunciated by Combs above, puts him firmly in the vanguard of those who see the need for a supportive presence, a significant other, perhaps, in the change process when feedback is given. Combs, as indicated at the beginning of this section, is skeptical of the virtues of solitary self analysis for the purpose of changing one's self perceptions.

F. How Consistent Are Self Perceptions?

Hamachek (1978) points out that perceptions need to be consistent for two reasons. First, there is less strain and anxiety if the social environment is not in a continual state of change. Second, consistency serves as the

foundation for stable human relationships. In fact, the world of social perceptions is very stable; once conceptual judgements are made, they tend to remain intact and unchanging. Hamachek then relates behavioral consistency to self concept theory, and reports that the values of "inner sameness" has been stressed by many psychologists. For Hamachek, self concept theory strongly suggests that we will "act like" the sort of person we perceive ourselves to be, and that as we encounter new experiences in everyday life, we will tend to accept or reject them in terms of their compatability with our present concepts of ourselves. By behaving in this way, we reduce conflicts and maintain our individuality as persons. He then discusses Festinger's model of cognitive dissonance and the need for consistency. Festinger's model will be discussed later in this paper.

Combs et al. (1976) support Lecky's position (1961) that the basic need of the organism is the maintenance of a unified organization. To achieve adequacy, one must develop a high degree of consistency within his phenomenal self. As did Hamachek, Combs then goes to Festinger's theory in order to explain the individual's need for consistency and the effect of dissonance on his self concept.

G. Festinger's Theory of Cognitive Dissonance

Festinger (1957) sees a person as continually striving for cognitive consistency. His basic hypotheses are as follows:

1. The existence of dissonance, being psychologically uncomfortable, will motivate a person to try to reduce dissonance and achieve consonance.

2. When dissonance is present, in addition to trying to reduce it, the person will actively avoid situations and information which are likely to increase the dissonance.

The need for consistency or consonance drives the person to try to reduce dissonance or to avoid situations that increase it.

Another of Festinger's maxims is that dissonance gives rise to pressures to eliminate the dissonance, and that the strength of these pressures to reduce the dissonance is a function of the magnitude of the dissonance. There is a limit to the amount of dissonance that can exist between any two elements, and this limit is set by the total resistance to change of the less resistant element. At the point of maximum possible dissonance, the less resistant element would change and the dissonance would be eliminated.

Festinger describes three methods for reducing or eliminating dissonance stemming from social disagreement:

1. One person may change his opinion so that it corresponds more closely with one's knowledge of what the others believe.
2. One may try to get the others who disagree with him to change their opinions to conform more closely with his opinion.
3. One may attempt in some way to make the others not comparable to himself, either by attributing different characteristics, experiences or motives

to them, or by derogating or rejecting them outright.

One should mention that Festinger's theory rests on an abundance of research evidence. Hamachek (1978) reports that while not all the evidence is unequivocal, over 500 experimental investigations do suggest that some such tendency as postulated by Festinger, does exist. In more recent times, Festinger's theory has continued to be fruitful ground for empirical research, although it has been modified and adapted to suit particular contexts. For example, Rosenberg (1979) defines a particular adaptation of the theory in terms of "contextual dissonance" and describes a number of empirical studies in this area. Contextual dissonance is based on the social similarity and dissimilarity of the individual to those around him, which affects his experience, and consequently his self concept.

One of the questions being asked by this study is directly related to Festinger's theory. The question asks whether the magnitude of change in self perceptions resulting from assessment feedback is a function of levels of agreement between the feedback and one's original self perceptions. Festinger's theory would lead one to expect that this is so.

H. How Stable Are Self Perceptions/Self Concepts

Already the discussion of the literature concerning consistency suggests that self perceptions will tend to be

stable. Combs et al (1976) are quite emphatic about this. Stability, or resistance to change, is one of the characteristics of an organization. Once established, self perceptions have a high degree of stability. Combs et al. describe the perceived self as our "fundamental frame of reference, our anchor to reality," and claim that even an unsatisfactory self organization is likely to be highly stable and resistant to change. This stability has been demonstrated by a number of researches which they list (Balester, 1956; Bloom, 1964; Engel, 1959; Gollin, 1954; Kagan and Moss, 1962; Roth, 1959).

I. Summary

This section began with a discussion of the importance of self perceptions in the work of Ryans and in the perceptual psychology of Combs. Combs is in the tradition of interactionist psychologists, who see the self concept, or one's self perceptions, as largely the result of social interaction especially with significant others. This notion of the self as a "looking-glass self" was traced from Cooley through Mead to Sullivan. Two views of the structure of the self concept were reviewed, and then a number of its characteristics were discussed, including how it is formed and changed, and how consistent and stable it is thought to be.

RESEARCH AND THE FOUR RESEARCH QUESTIONS

Throughout this literature review, research has been cited in support of a number of the theoretical positions discussed. In this section, the review of research studies will be focused on those studies that relate directly to the research questions.

> Question 1: Does either of two forms of feedback--self analyzed and reported (SAG), or externally analyzed and reported (EAG)--cause any change at all in a teacher's self perception of the relative levels of his or her abilities?

This first question, precisely defined for the purpose of experimental investigation, can be broadly restated for the purpose of this literature review: Is there any evidence that feedback of any sort changes self perception? Theoretical arguments have already been presented in the earlier sections of this chapter.

Webster (1974) in his review of self evaluation research, describes a number of experimental investigations that bear on the question. The prototype of change studies is the experiment by Israel (1956). Participants in the experiment were 107 students from four classes at a Swedish physical education college. Each class was divided into two experimental groups of 16 to 18 participants, who knew one another well. Each participant was asked (1) to rank all

members including himself on leadership ability (2) to estimate the level of leadership others thought he possessed, and (3) to rank himself as he wished the others would. An attempt was then made to get individuals to change their self-ranking by reporting fictitious rankings from others. The significant findings are first, that the evidence supports the basic idea of the looking-glass self, and consequently, the idea that the perception of others' opinions influences one's self perception. Second, the likelihood of change in one's self perceptions is directly related to the attractiveness of the group. Third, the effect of others in changing the self concept is inversely related to the individual's accuracy of perception; so that, if the individual did not accurately perceive the others' perceptions, he was more likely to change. This last finding is also relavant to the third question in this study to be discussed later.

In a study by Backman, Secord and Peirce (1963) college students rated themselves on personality items and told how close friends and relatives would rate them on the same items. Then they filled out personality measurement scales which were to be scored by an expert psychologist and this "objective information" was fed back to them. But this "objective information" included attempts to alter their self perceptions on certain personality items. The significant findings are that participants do change their self ranking as a result of the manipulated reports, and that they change most

in those areas where they feel friends and relatives do not agree with them. Incidentally, this study also demonstrates the potency of expert evaluators in changing self perceptions.

Videbeck (1960) experimented with 30 students from introductory speech classes, rated as superior by their instructors. They rated themselves on a nine point scale on each of 24 items to do with adequacy of oral presentation. After this, they read six poems for a "visiting speech expert." The major findings were that this procedure is effective in changing the individual's self rating. A second major finding has to do with the direction of the changes. These data show no greater tendency on the part of individuals to change in a positive than in a negative direction. In fact, the data suggests that those who received positive ratings changed less than those who received negative ratings. This finding is contrary to a finding by Moore (1964) who reported that individuals tend to raise their self-perceptions more in response to feedback than to lower them. Maehr, Mensing and Nafzger (1962) replicated the Videbeck study using 31 boys enrolled from a high school physical education class, and confirmed Videbeck's finding except that regarding greater changes in the disapproval condition.

These studies strongly suggest that feedback does change self perceptions. Whether positive or negative feedback is more potent is uncertain. But it does seem that the attraction of the assessing group is a strong influence,

as is the "expertness" of the evaluator. This last point is more relevant to the second research question discussed below.

Question 2: Which of two forms of feedback (SAG or EAG) will yield the greatest changes in a teacher's self perceptions?

The evidence presented earlier, in the section dealing with the importance of feedback in teacher education, suggests a split verdict on this question. Practice and research findings from supervisory and peer evaluation seem to favor an external analysis of feedback, whereas experience and data from microteaching and computer assisted programs seem to promote self analysis of data. Two of the experiments reported in response to question one above, support the efficacy of the "expert" evaluator (Backman et al., 1963, and Videbeck, 1960).

Hartman (1978) reports a number of very relevant findings from the Teacher Self Appraisal Research Project (Brooks, 1967). The major components of this project were (1) voluntary participation (2) leaders with the ability to provide an accepting, non-threatening climate (3) frequent videotaped feedback viewed only by the participants who were taught encoding and analytical skills, and (4) a year long inservice program of weekly meetings in which the principles of perceptual psychology were taught, in addition to research findings about teaching and the analytical skills already mentioned. The assumption was that teachers

provided with these tools would be intrinsically motivated to improve. The findings clearly indicate that looking at videotape feedback is not enough; teachers need professional competencies in order to describe and evaluate their behavior. Also there was evidence that group support was vital for the success of this sort of feedback, and that a longitudinal rather than short range approach is necessary.

This study is mentioned here to offer a caveat where the self assessment approach is concerned. Careful tutoring is necessary for it to be effective. This study also gives some support to the efficacy of self evaluation under these conditions, especially with the support of a peer group as part of the system.

On the other hand, a study by Litwack (1974) which compared three types of feedback treatments: authority feedback, peer feedback and self feedback, using data from Flanders Interaction Analyses, suggests that participants feel more secure when they receive feedback from an authority figure, or peer group, than when they receive it from the videotape alone.

The evidence cited in answer to this question is not unequivocal, but does seem to weigh in favor of external analysis as a more potent force for change.

Question 3: Is the magnitude of change in self perceptions resulting from assessment feedback a function of empirically established levels of agreement (high, low) between feedback received and original self perceptions?

This question has also been anticipated by some of the research studies discussed in relation to question one. Israel's study (1956) indicated as one of its findings that if an individual had not accurately perceived others' perceptions of him, he was more likely to change, than if he had accurately perceived them. The degree of difference between one's original perception and the feedback received created dissonance which exerted pressure on the individual to change his self perceptions.

Studies by Videbeck (1960), Moore (1964) and Maehr et.al (1962), though they disagree about relative potency, do agree that both positive and negative differences between self and other ratings exert pressure on the individual to change one's self perceptions.

A few studies report on the effect of discrepancies between one's espoused platform of behavior and the platform one puts into use (Simon 1976), or discrepancies between one's ideal and actual behavior (Fages, 1978), or discrepancies between students' perceptions of their ideal and actual teacher fed back to the teacher (Gage, Runkel and Chatterjee, 1960).

Two studies are concerned with the specific question of the effect of magnitude of discrepancy between self perceived and observed teaching behavior. Tuckman, McCall and

Hyman (1969) hypothesized that changes in teacher behavior, self perception, or the discrepancy between the two would be an increasing function of (a) the magnitude of the initial discrepancy between self perception and observed behavior, and (b) the nature and specificity of the feedback. This hypothesis is based on Festinger's writing and research (1957) which show that the greater the dissonance between cognitive elements, the greater the pressure to change one or other of them. Twenty-four high school teachers were assigned to a High or Low Discrepancy group depending on discrepancies between their scores on the Self Perception Inventory and the scores given them by trained coders of audiotapes of their behavior. Treatments were based on three different types of feedback (Flanders, verbal feedback, and audiotape), and there was a control group. The findings were that the treatments had no differential effects on changes in self perception, but that initial discrepancy level did have a differential effect on changes in perception. Teachers in the High Discrepancy level group changed their total perceptions across all treatments to a significantly higher level than did Low Discrepancy teachers. Thus teachers with initially high discrepancies changed their view of their own teaching significantly to a greater extent than did teachers with low discrepancies. It was also found that initial discrepancy levels seemed to effect self perception change but not behavior change.

Doyle and McNally (1974) reported on the effect of intent-action discrepancy and student feedback on teacher behavior change. They surmised from the study of the Tuckman et al. (1969) that in the absence of an external model, reduction of discrepancy induced by feedback results in a modification of personal intentions rather than teaching behavior. Their findings tended to confirm this surmise, for they concluded that in the absence of an externally validated and supported model of approved behavior, verbal feedback appears to affect perceptions rather than behavior.

The participants in their experiment were 36 junior high school teachers who volunteered to participate in an experiment in microteaching. Teachers were first introduced to the teaching task and asked to fill out an intent inventory based on the amount of time they expected to devote to various classroom behaviors during the teaching session. Students were tested on the material taught after the lab teaching session. Participants then received feedback on intent-action discrepancy and student test performance, and were given a similar teaching task to perform.

The results showed that feedback concerning student learning outcomes did not have a significant impact on either the teachers' perceptions of how they would teach or how they actually taught afterwards. However, the data indicated that discrepancy conditions did have a significant effect on the amount of intent change in the areas of direct and indirect influence. These results indicate that when teachers are

asked to reteach the same content, they react to intent-action discrepancy by revising their intentions of how they will teach rather than by changing their actual classroom behavior. Thus, knowledge of student outcomes seems to have little impact on teaching behavior, and feedback regarding differences between teacher intention and teacher performance seems to lead teachers to change their perceptions but not their performance.

As to the precise question of the differential effects of magnitude of discrepancy, a comparison between high-discrepancy and low discrepancy participants showed that high discrepancy participants changed more on the two measures of teacher intent (direct and indirect teaching behavior).

In summary, the evidence seems to indicate that magnitude of change in self perceptions as a result of assessment feedback is a function of high or low levels of agreement between feedback and original self perceptions. All the literature reviewed showed that dissonance is a potent motivator of change, and two studies suggested that the greater the level of dissonance the greater the change in self perceptions. Perhaps we should note with caution a limitation reported by Glassberg (1978) in reference to Maves (1972). Glassberg writes, in the context of ego development and student teachers:

> Developmentalists emphasize the role of the environment in creating disequilibrium and point out at the same time that too much disequilibrium can become overwhelming resulting in fixation at a stage rather than progression to the next stages.

The studies reviewed here also suggest that there are three levels of response to feedback. Feedback can affect (a) self perceptions (b) intention to change, and (c) actual behavior. The relationship might be represented in the following way:

Feedback -> changes in self perceptions -> intention to change behavior -> actual changes in behavior.

The present investigation is concerned with the relationship between feedback and self perceptions. As indicated by the studies reviewed, further research needs to be done on the relationship between feedback and intention to change behavior, and between feedback and actual changes in behavior.

Question 4: Will the changes in self perceptions revealed in the posttest immediately following the experimental intervention (seven to ten day interval) differ from corresponding changes in perceptions revealed in a delayed posttest administered approximately six weeks after the intervention?

Psychologists seem to agree on the stability of self perceptions. Combs' arguments (1976) and the research he cited in favour of the stability of self perceptions have already been discussed in a previous section dealing with the importance of self perceptions in teacher education.

Webster (1974) describes two studies that tested the stability of self perceptions over the exact time period (six weeks) used in this investigation. A series of studies reported by Haas and Maehr (1965) were designed to change various aspects of self evaluation. Participants were eighth grade boys in physical education classes. The findings from the first experiment indicate that changes induced by the treatment

were great immediately following the evaluation, and that they persisted at the same level for the duration of the six week study. In a second experiment involving differences in the number of treatments, participants were found after the second "dosage" to have made a greater change in the predicted direction, and this level of change persisted for six weeks.

The conclusion is obviously that in the absence of other treatments, time alone, at least such a period of time as six weeks, will not affect self perceptions that have been changed by strong experimental treatments.

Conclusion: Critical Issues Still Outstanding From Review of Research

Although the research literature supports the view that feedback does influence self perceptions, a number of critical issues are left outstanding. The main question not addressed is whether self perceptions are altered by external feedback after they have been established by self assessment. This question is the main focus of the empirical section of this investigation.

In the context of professional development, we need to know more about the specific attributes of "significant others". Questions about the validity of their feedback need also to be pursued. What are the comparable effects and validity of feedback from "experts"?

With regard to the nature of feedback, levels of specificity, and alternative instruments and categories of

teacher behavior should be studied. The effects of different modes of analysis and transmission also need further comparison.

Questions about directionality are unresolved. Do positive or negative differences between feedback and self perceptions cause the greater changes in self perceptions?

A number of demographic variables could influence judgement about the effect of feedback on self perceptions. These include sex, geographical location, level of school taught, age and years in teaching. They need to be investigated for their possible influence. In this investigation, an attempt will be made to control them.

The question of whether an external support system is more effective than a self reliant one needs careful examination. This question is a second major empirical focus of this investigation. An attempt will be made to distinguish between the function of the external agent as the analyzer and transmitter of feedback and other support functions sometimes described as "interaction".

Magnitude of dissonance as a motivator of change in self perception and behavior has not been sufficiently researched in teacher education. Is magnitude of self change always the result of magnitude of dissonance if one does not reduce dissonance by changing the opinion of others or one's opinion of them, as Festinger suggests. Perhaps it is important to test also whether the expected result is the only result, and whether the "maximum possible dissonance", even

if it does cause the expected change, has harmful side effects.

Another important question is whether dissonance or intentions, changes only self perceptions or intentions, or also influences a teacher's classroom behavior, and if so, under what conditions. This is not a concern of the present empirical investigation, but it is a question that needs further experimental study in the context of teacher development.

Finally, theory strongly indicates that self perceptions are very stable, and the evidence cited here confirms that opinion, at least for periods up to six weeks. Whether stability is maintained beyond this time could be determined in other experiments. In this experiment, stability of any changes resulting from the treatment will be tested over the normative six week period.

CHAPTER III

DESIGN OF THE STUDY

Introduction

This investigation had two distinct stages: a preparatory stage that focused on instrumentation and instructional methodology, and an experimental stage that focused on relations between feedback and self-perceptions.

The preparatory stage represented an attempt to develop a process and a set of instruments for guiding teachers in their professional self assessment. These products were developed in the first of two consecutive graduate courses in a Master's Degree Program for Classroom Teachers (MACT), and were incorporated into the curriculum for that course.

The main, or experimental, stage took place in the second course of the two term sequence. It consisted of two experiments which investigated the general question of whether externally mediated feedback is more likely to change a teacher's self perceptions than is self mediated feedback, or no feedback at all.

The events which occurred in the first term had a major influence on the conduct of the two experiments. It is, therefore, important to review these events prior to describing the experimental design.

Before undertaking this review, two reasons for the importance of the preparatory stage should be brought to mind. First, the basic question under investigation was not whether assessment feedback in general makes a difference to a teacher's self perceptions, but specifically whether it makes a difference when the teacher has already undergone a careful process of professional self assessment. The kind of preparatory self assessment process the teachers in this investigation underwent is what is to be described here.

In the second place, the Teacher Behavior Survey (TBS)--the instrument used for collecting feedback data and for measuring changes in participants' self perceptions during the experimental investigation--was developed with the teachers as an integral part of their experience in the first term. Thus, the development of the TBS instrument is important both as a product which was used in the experiments, and as a part of the instructional process that guided the formation of the teachers' self perceptions in the first term.

Description of Critical Events in the Preparatory Stage

Throughout this description, then, two considerations should be kept in mind. The events were intended to contribute (a) to the participants' sophistication in self analysis, and (b) to the development of the Teacher Behavior Survey (TBS).

As already mentioned, the preparatory stage of this investigation occurred during the first course in a two course

graduate level sequence for classroom teachers. Throughout this initial course, the teachers underwent an individual needs assessment of their professional strengths and deficiencies. These personal evaluations became the basis for consultation with the program adviser so that each teacher could frame his or her own individual plan for professional development within the overall resources of the program of the Master of Arts in Classroom Teaching (MACT).

The investigator had a shared involvement in the design and conduct of the course, drawing on a set of materials and procedures for a needs assessment developed by the professor, Dr. John Cragun, with whom he worked. The investigator's role was to assist in reorganizing the course and developing new materials so that a more systematized and replicable model for conducting the needs assessment might emerge. The investigator had not previously met any of the teachers on the course.

The needs assessment process which the participants experienced had six steps. These might be briefly described as follows:

Step 1. Small Group Brainstorming on Teacher Abilities

In the very first class session, the participants were divided into small groups and asked to discuss and list the characteristic strengths of the skilled teacher and the deficiencies of the unskilled teacher within each of the following categories of teacher performance:

Planning

Organization and Management

Working with People

Classroom Climate and Control

Command of Subject Matter

Teaching Methods

Use of Audio-Visual and Other Materials

Understanding Human Growth and Development of Children

Understanding Curriculum and Curriculum Development

Evaluation Procedure

Personal and Professional Characteristics

Other

The rough notes that described strengths and deficiencies were collected from the groups. They were read, and then sifted and elaborated by the investigator into a set of statements under each of the category headings. In this way two comprehensive lists were prepared. One, entitled "List of Teacher Abilities," listed the strengths of the skilled teacher and the other, entitled "List of Teacher Deficiencies," listed the deficiencies of the unskilled teacher, within each of the general categories.

Step 2. Rating of Teacher Strengths and Deficiencies

In the second session, the participants were asked to do two things. First, they were asked to rank from one to three the most important strengths in each category on the first list, and then to perform the same ranking of the most

characteristic deficiencies of the unskilled teacher in each category on the second list.

Having in this way internalized some specific indicators of teacher behavior within a set of explicit categories, the participants then began initially to examine their own performance.

Step 3. Presentations on Areas of Teacher Behavior

In the next four weeks, while the teachers were engaged in systematic self analysis of their professional behavior, lecture discussions and seminars were held on a number of these categories of teacher behavior. The intention was to help the teachers to a deeper understanding of the particular category and to provide the opportunity for them to evaluate their current practices--their strengths and deficiencies--in the light of this understanding. The presentations were descriptive of good practice and reviewed research on teaching in each area. The categories on which presentations were made were the following:

Teaching Strategies

Questioning Skills

Verbal Interaction Analysis

Helping Relationship Skills

Classroom Management

Planning Strategies

The other categories were informally discussed from time to time during class sessions.

Step 4. Individual Needs Assessment

Next, after they had been assisted by the presentation to evaluate their own behavior in a number of specific categories, the teachers were now asked to prepare a written outline of their needs assessment, using the lists of teacher abilities and deficiencies as a guide. They discussed these, first of all, in a small group session, and the following week they presented a full written account of their needs assessment to the instructor. These accounts were discussed in individual conferences.

Step 5. Initial Program Planning

Based on the results of the personal needs assessment, each teacher began to construct an individualized plan of courses, experiences and projects for the MACT program, with the help of the course adviser.

Step 6. Individualized Mini-Projects

Candidates planned and carried out a mini project of not more than three weeks duration to develop some aspects of their own classroom behavior suggested by their needs assessment. For example, a teacher might have discovered a limitation in questioning technique. The teacher would follow a set of project planning guidelines in mapping and executing a strategy for improving skills or a single skill, in this area. These guidelines required the teacher to describe his or her problem exactly, to state limited objectives as concretely as possible, to set out specific activities to be done

in the few weeks available and, finally to specify how progress was to be evaluated.[1]

The intent of this summary of the six step needs assessment process has been to describe how the preparatory stage contributed to:

a. the participants' sophistication in self analysis and the firm foundation of their perceptions at this point, and

b. the development of the Teacher Behavior Survey (TBS).

With respect to (a), it should be stressed that throughout the term, and particularly as a direct result of undergoing the first four steps of the process, the participants had been concentrating on a specific set of categories of teacher behavior and had been examining their own behavior in the light of their understanding of these categories. By the end of the term, the teachers should have had quite specific insights into their own behavior in each of these specified areas.

With respect to (b), it should be noted that the first of the six steps resulted in the development of a "List of Teacher Abilities" and a "List of Teacher Deficiencies." The two lists were similar in overall structure: both contained about 12 statements in each of the 11 categories of teacher behavior (for example: Planning, Organization and Management,

---

[1]This brief summary of the six step needs assessment process may prove inadequate for those who might wish to replicate the investigation. A more detailed account is provided in Cragun and Wilson (1980).

or Working with People). In step two, participants were asked to rate the dozen or so statements in each category on a scale of one to three in order of importance. From these weighted lists, the investigator selected the items with the highest ratings, and used them in constructing the two sections of the Teacher Behavior Survey (TBS).

General Overview of the Design of the Experimental Investigation

A. Introduction

As the preceeding discussion suggests, the initial course in a two term graduate sequence is based on the assumption that teachers can gain detailed insights into their professional behavior quite economically and effectively through a process of self assessment. Evidence collected by others suggests that this assumption may be warranted (De Marte,Kelly, and Freeman, 1980). However, this evidence does not speak to the stability of self perceptions that are formed. This investigation therefore, focused on the general question of whether or not self perceptions formed as a result of a systematic self assessment process will remain relatively unchanged when teachers receive feedback from others regarding their classroom performance. There are two distinct ways in which this feedback might be received after the data has been collected. Feedback data might be given directly to the individual to be self-analyzed, or it might be externally analyzed and then reported to the individual. Thus, the two experiments that represent the main focus of this investigation

also addressed the more specific question of whether feedback that is mediated by an external evaluator will alter a teacher's preformed self perceptions to a greater extent than is true for feedback in which the data is processed by the teacher himself or herself.

B. Summary of Experimental Design

The two experiments may be schematically summarized as follows:

(R) 0_1 X_1 0_2 0_3

Experiment I

(R) 0_1 X_2 0_2 0_3

0_1 0_2 0_3 Experiment II

0_1 = Pretest (Teacher Behavior Survey, April 16-25, 1979)

0_2 = Posttest (Teacher Behavior Survey, May 28-June 6, 1979)

0_3 = Delayed Posttest (Teacher Behavior Survey, July 10-17, 1979)

X_1 = Data self analyzed and reported

X_2 = Data externally analyzed and reported

This outline makes use of the Campbell and Stanley notation (1963), and represents the experimental design over time. The steps in conducting this experimental investigation are summarized in the "Calendar of Events" presented in Appendix A. In Experiment I there is random assignment (R)

of pairs of participants selected from a total sub-population, two experimental treatments (X_1 and X_2), and the repeated use of the same measure on three separate occasions as pretest, posttest and delayed posttest (0_1, 0_2 and 0_3). The design of Experiment II is related to that used in Experiment I in two ways. Both experiments took place over the identical time intervals, and used the same test measure repeatedly. The participants in Experiment II, however, were not randomly assigned, as in Experiment I. They also received no experimental treatment, as they were being considered in the role of a quasi control group.

C. Narrative Outline of the Investigation

Experiment I

This experiment took place in the Spring term of 1979, and involved the MACT candidates who had taken the course already described as the preparatory stage to this experiment. They were now doing the second course in the professional development sequence of the MACT program. Both these courses were offered in two middle sized cities in mid-Michigan, hereafter referred to as City A and City B.

The teachers in these two classes were sorted into fourteen matched pairs according to geographical location (City A or City B), sex, and level of school (elementary or post elementary). Members of each pair were then randomly assigned to one or other of the two experimental groups. The first experimental group, the self assessed group (SAG), was

told that its members would be given evaluation data concerned with their teaching performance, which they would have to process and report on. The second experimental group, the externally assessed group (EAG), was told that its members would have the data processed for and reported to them. The members of both classes who could not be paired were dropped from the sample.

The instrument used for pretest, posttest and delayed posttest, and for collecting the data, was a questionnaire called "Teacher Behavior Survey" (TBS) with two sections:

I. Questionnaire on Teacher's Strengths

II. Questionnaire on Teacher's Possible Areas for Improvement.

A copy of the TBS questionnaire is to be found in Appendix B and will be discussed later. It should, however, be noted that this questionnaire was developed from the two lists ("List of Teacher Abilities" and "List of Teacher Deficiencies") which had come out of the small group brainstorming session in the previous term. The items on these lists had also been rated for each category of teacher behavior by these same teachers.

The pretest was given to all the teachers at both locations. Approximately six weeks later the posttest was given, and six weeks after this the delayed posttest was given. An attitude scale developed for this purpose was also distributed for the teachers to complete after the delayed posttest (see Appendix C).

During the period between pretest and posttest, the experimental treatments were done. Members of both experimental groups selected two peers and an administrator from their schools to observe them and provide assessment data on their performance. This the assessors were to do by filling out and returning the questionnaire to the investigator. A letter containing instructions for the assessors was attached to each questionnaire (see Appendix D). The assessors were also asked to complete and return a brief questionnaire giving information on how they performed as assessors (See Appendix E).

The next stage involved two separate operations. First, the three assessment questionnaires for the teachers in the first experimental group (SAG) were copied out by hand (to avoid identification of evaluators by their handwriting), and these transcribed copies were given to the teachers to be analyzed and reported on according to a set of instructions read by the investigator (See Appendix G). For the second experimental group (EAG), the quesionnaires were processed and a report prepared by the investigator for each member of the group. (Appendix H provides a copy of the structured outline the investigator followed in preparing reports for the members of the EAG group).

Next, the members of both groups had individual interviews with the investigators. Those in the SAG group brought to the interview the written report on the feedback they received, and made an oral presentation. The

investigator's role was that of a carefully interested listener who responded only by requests for clarification. The members of this group then wrote reports on the feedback they had received, following the same guidelines that were given to the SAG group.

Approximately one week later, all teachers were required to take the posttest, and six weeks later, they completed the delayed posttest and the attitude scale.

Experiment II

Experiment II set out to investigate whether teachers who had previously experienced a self assessment process like that undergone by the teachers in Experiment I, but who did not receive any further treatment by way of external assessment feedback, would differ in the pattern of change or lack of change in professional self perceptions over the same period of time as in Experiment I.

Teachers in the second year of the MACT program who attended the City A center were the participants (EIIG) in this experiment. There was no second year program in City B. The City A second year candidates had done their needs assessment approximately a year previously under the direction of the professor with whom the investigator had been associated in the preparatory stage of the investigation, and they had continued to take courses in the MACT program.

First, the nature of the experiment being done with the first year MACT candidates was explained to these second year candidates. They were then told the purpose of their

involvement in Experiment II. Subsequently, they were treated as a control group, in that they received only the tests which the participants in Experiment I had received. They were given the pretest, posttest and delayed posttest during the same periods that were established for the first experiment.

Description of Experimental Treatments

A. Introduction

The experimental treatments involved three operations:

1. Data Collection, which was identical for both of the groups in Experiment I (SAG and EAG).
2. Data Analysis, which was done by the members of the SAG, but was done for the members of the EAG by the investigator.
3. Reporting, which, for the SAG involved writing a report before the Interview, and making an oral presentation at the Interview. For the EAG, an oral presentation was made by the investigator to each member of the group, who then wrote a report after the Interview.

A major concern where the data analysis and reporting were concerned was for the reliability of the experimental procedures. So, great care was taken to ensure that identical procedures were carried out within and between groups, where these were required for the integrity of the experiment.

B. Data Collection

At the first class meeting of the Spring term, the members of the SAG and EAG groups were each given three copies of the Teacher Behavior Survey (TBS), and three large stamped, addressed envelopes, and asked to give these to two of their peers and one administrator at their schools. Attached to each questionnaire was a letter to the assessor (See Appendix D), and a single sheet, headed "Description of Evaluator" (See Appendix E). Participants were to choose the people who they felt confident would do the job conscientiously and give useful feedback on their teaching performance. The members of these experimental groups were themselves familiar with the TBS since they had contributed to its development in the preparatory stage of the experiment.

The letter attached to the questionnaire gave instructions to the assessors as to the minimum required for doing the job properly. They were asked to observe the teacher for about one hour or one full class period. They were to look at the teacher's lesson plan book, teacher records and reports, teacher prepared materials, samples of students' work, etc. and they were to have informal talks with the teacher. They were also told to draw on their overall knowledge of the teacher over the period they had known him or her. Finally, they were given instructions about preserving their anonymity.

The single sheet, entitled "Description of Evaluator," was intended to gather descriptive data on the assessors, but also to provide a check on how they carried out their functions.

C. Data Analysis

Self Assessed Group (SAG)

When the three completed TBS forms were returned by the assessors for this group, the investigator copied them out by hand onto blank TBS forms. The summaries at the end of each section were also typed by him. These copies were given to each participant in the SAG group, and members of the group in City A and City B were given identical oral instructions by the investigator. To ensure this, the instructions had been written out beforehand, and virtually memorized by the investigator (See Appendix G).

The instructions stressed that there was no unique way to analyze the data; so, SAG members should make the most careful and meaningful analysis of the data, keeping in mind that what they wanted from the data was help in making decisions about their future development as teachers. When the analysis was completed, they were to write a summary/outline to be used as the basis of their oral report at the Interview with the investigator. Their summary should adequately cover the following points:

1. Show how they went about processing the data.
2. Give a summary of their findings--what the main messages were.
3. Give an account of what they found that should be most helpful to them from the point of view of their professional development.

4. Explain what there was about the evaluation and feedback process that they found helpful, not helpful, or needing improvement.

Externally Assessed Group (EAG)

The investigator was responsible for analyzing the data for this group. The results of the three TBS surveys for each teacher were transcribed onto one blank TBS form, using three different colored pens--partly so as to be able to distinguish between the assessors, but more importantly, so that the researcher could quickly scan and read the results to each participant in the later Interview. Then, the sectional summaries were retyped.

The next step was to calculate the mean ratings of the subscales for each teacher, so that the top three strengths (or four, in the case of a tie), the middle strengths, and the bottom three (or four) strengths could be discerned and listed. The subscales were analyzed in terms of relative strengths, because a study of the questionnaires indicated that this was how the assessors rated the teachers. It was a relatively rare thing for an assessor to rate even a single item on the first section of the questionnaire as "Below Average". On the second section of the TBS, which dealt with possible areas for improvement, assessors suggested very few areas in which the teachers seemed to them to need improvement. Where the assessors did indicate such areas, the investigator reported the areas in the Interview.

Further analysis of the data had to be done in preparation for the presentation required in the Interview. The investigator had to be able to give a comprehensive review of the data to each participant on the spot in a live interview situation, and he had to do it in a way that was identical, or at least uniform, for all subjects. Reliability was a very important consideration here. How the investigator prepared himself and the data for this presentation will be discussed in the next section.

D. Reporting

Self Assessed Group I (SAG)

The teachers in this group had been told to prepare a written report which they should use as the basis of their oral presentation. When they came to the Interview, they were told that they should talk through their written reports feeling free to expand, explain, raise new points, and otherwise alter the written version in any way they wished.

The problem for the investigator was to respond in such a way that he was perceived to be interested and encouraging, and yet behave in such a manner that could be uniformly replicated for each teacher. Simplicity seemed the best solution. With respect to verbal interventions, he restricted these to requests for clarification. But overall, the main device was to explain to each teacher at the beginning of the Interview that in order to keep his attention

focused and also more easily remember afterwards the points the teacher raised, the investigator wished to take full notes. No one seemed put off by this request, nor did in seem to bother anyone in practice. The investigator attempted, then, to present to each teacher a uniform picture of absorbed attention as he sedulously took notes and occasionally interrupted to ask for clarification of a particular point.

At the end of the session, the teachers handed in their reports, were given a copy of the posttest to be filled out and handed in by the end of term, and were warmly thanked.

Externally Assessed Group (EAG)

(a) Investigator's Preparation

It was of the greatest consequence in terms of reliability, that the reporting to the individuals in this group should be carefully uniform. The first step was to make decisions about what form the data was to be presented. The next step was to develop an almost ritual format which the investigator could thoroughly practice before the Interview. The following decisions were made:

1. The teachers should be given the re-typed summaries of their strengths and needs for improvement.
2. They should know the order in which their abilities were rated, that is, those categories that were rated top, middle, or bottom on a scale of mean ratings.

3. Then they should be given an account of the ratings for each item in each of the categories in descending order. Categories and now, in turn, items within the categories were to be presented in descending order.
4. Any areas needing improvement suggested by the assessors should be reported.
5. They should be given a preliminary example of the procedure as part of the format.
6. Summaries should be given at the end of each section of the presentation.

Following on these decisions, a detailed scenario was put together, which incorporated the decision points. There folflowed a process of trying out the format in simulated Interview situations and modifying the script until the investigator felt that it was most likely to enable him to do as completely as possible what he wanted to do in the same manner for all teachers in this group.

A copy of the final version of the script, which was used in the Interview, entitled "Researcher's Schedule for Reporting to Group II (EAG)," is to be found in Appendix H. After a rehersal, in which he went completely through a few of the questionnaires, the investigator was ready.

(b) Interviews

After the candidates had been greeted, they were given a copy of the re-typed summaries which had appeared at the end of both sections of their assessors' questionnaires.

After a few minutes, when they seemed to have finished reading, they were given a blank TBS form on which to make notes. Then the interview scrupulously followed the prepared script, as described above. The investigator departed from the script only to answer requests for clarification.

At the end of the reports, the investigator reminded the teachers that they had to write a summary/outline on the feedback they had received, which was to be turned in by the end of term. The report should cover the same four areas prescribed in the "Instructions to Group I" already cited under the description of data analysis for the SAG (see also Appendix G).

They were then given a copy of the TBS for the posttest which was to be completed after they had written the report, and were asked to hand it in at the last class for the term. Then they were sincerely thanked.

Participants in the Investigation

A. Population

The target population which this study has in mind is made up of teachers in any inservice training program that features a process of professional assessment for the purpose of professional development. The defined sub-populations which the investigation dealt with were, for Experiment I, the first year MACT candidates in two middle sized mid-Michigan cities (City A and City B), and, for Experiment II, second year MACT candidates in City A.

The fact that the defined sub-populations of both experiments consisted of teachers who had participated in a course that focused on self-analysis prior to the initiation of the treatments may severely restrict the generalizability of the findings. This issue is, therefore, discussed in a later section of this chapter that describes the external validity of the study.

b. Sample - Selection and Assignment

For Experiment I, the sample was selected from 36 teachers enrolled in the first year of the MACT program in City A and City B in the Winter and Spring terms of 1979. The selection of participants and their assignment to the two experimental treatments went as follows.

The names of the 16 teachers from City B and the 20 from City A were written on slips of paper and sorted into eight groups:

	City A	City B
Male post elementary	Group 1	Group 2
Male elementary	Group 3	Group 4
Female post elementary	Group 5	Group 6
Female elementary	Group 7	Group 8

Two containers were prepared, one labelled City A, the other City B. The slips from Group 1 were crumpled into balls and put in the City A container, and the slips from

Group 2 crumpled and put in the City B container. A flip of a coin decided which group would be selected first. Then one slip was blindly selected for the first experimental group, the Self Assessed Group (SAG), from the City A container, as the flip of the coin directed, and placed in a box marked SAG. A second slip was taken from the same container and put in the box marked EAG (Externally Assessed Group). These were matched by selecting two names from the City B container in the same manner. When the matched pairs in these two categories were exhausted, slips from Groups 3 and 4 were placed in the City A and City B containers. The pairing and matching process continued in the same way until all the possible matched pairs were assigned. Fourteen pairs were assigned--seven pairs from City A matched by seven pairs from City B. Eight teachers, who could not be paired on these variables, were not included in the sample.

It should be noted that in an attempt to control for certain potentially important confounding variables, the participants were blocked on sex, geographical location and level of school taught before the random assignment of pairs began.

Late in the experiment, at the very end of the Spring term, one of the female elementary school teachers from the EAG group became tragically ill. During the Summer another female elementary teacher from the SAG group, from the same geographical location, had a prolonged illness and was unable to complete the delayed posttest. Both these teachers had to be dropped from the experiment. Nevertheless, the matching

of the groups was not seriously affected, as both subjects were female, elementary teachers, from the same geographical area, belonging to different experimental groups. Where it did affect balance, however, was that now each group had seven teachers from one geographical area and only six from the other. The experimental groups, therefore, continued to be identical in terms of the demographic criteria being used, but one variable (location) was not equally distributed in both.

Finally, the total sample contained 26 participants, 13 in each of the two experimental groups. The demographic composition of each group was as follows:

Two male post elementary teachers, one from each of the two locations.

Two female post elementary teachers, one from each of the two locations.

Five female elementary school teachers from one location

Four female elementary school teaches from the other location.

For Experiment II, the total group of 17 teachers in the second year MACT program in City A was initially included in the study. Eventually, two teachers who did not hand in one of the tests, were excluded from the sample. The final group of participants was made up as follows:

Two male post elementary school teachers

one male elementary school teacher

12 female elementary school teachers

A summary of the demographic characteristics of the two groups in Experiment I (SAG and EAG) and the group in Experiment II (EIIG) is presented in Table 1. A description of the individual participants is included in Appendix I.

As mentioned earlier, the defined sub-population from which the SAG and EAG were selected had been blocked on location, sex, and level of school. It had not been blocked on age and years taught for two reasons. First, the relatively small number of participants limited the amount of blocking that could be done. Second, it was felt that since age, and to a lesser extent, years taught, were more freely distributed among participants than were the blocked variables, these would be less vulnerable in a process of random assignment. However, as a study of Table 1 shows, leaving the assignment of these two variables to chance resulted in their uneven distribution among the SAG and EAG groups. The consequences in terms of internal validity will be discussed later in this chapter.

Instrumentation

A. Measures Taken

In all, eight measures were taken. Six of these--the pretest, posttest, delayed posttest and feedback data collection by three assessors--made use of the same instrument, the questionnaire or Teacher Behavior Survey (TBS). The seventh measure was based on response to a one page

TABLE 1: DEMOGRAPHIC CHARACTERISTICS OF GROUPS IN EXPERIMENTS I AND II

Age	Years Taught
1 = 20 - 24	1 = 0 - 4
2 = 25 - 29	2 = 5 - 9
3 = 30 - 34	3 = 10-14
4 = 35 - 39	4 = 15-19
5 = 40 and above	5 = 20 and above

			Location		Sex		Level of School		Age					Years Taught				
Experiment	Group	n	A	B	M	F	Elem.	Post Elem	1	2	3	4	5	1	2	3	4	5
I	SAG	13	7	6	2	11	9	4	4	5	1	2	1	8	2	2	1	0
II	EAG	13	7	6	2	11	9	4	3	1	2	2	5	4	4	1	2	2
II	EIIG	15	15	-	3	12	13	2	2	3	1	4	5	6	2	3	1	3

questionnaire entitled "Description of Evaluator." The eighth was based on an attitude scale entitled "Debriefing for Participants in the Two-Term MACT Professional Evaluation," and measured the teachers' attitudes to various elements in the two stage evaluation process. The administration and timing of the eight measures have been described earlier and are summarized in Appendix A.

The number of respondents and rate of return was very high for all measures. Only two of the participants in Experiment II failed to return one of the tests and were dropped from the sample before the analysis of the data was undertaken. In Experiment I, only two female elementary school teachers from City B (one from the SAG and one from the EAG) did not participate in the posttest, due to serious illness in both cases. They were also omitted from the analysis.

B. Instrumentation Characteristics

Teacher Behavior Survey

(a) Description

The TBS is an original questionnaire developed by the investigator. It was derived from the two lists of statements describing general categories of teacher behavior ("List of Teacher Abilities," and "List of Teacher Deficiencies") that were developed with the teachers in the preparatory stage of this experimental investigation. After the statements on the two lists had been rated by the teachers, the investigator took the five or so statements with the highest ratings

in each category and converted them into the items for the questionnaire.

It should be noted that the two lists of statements describing teacher behavior were produced by the teachers in the first sessions of the initial course in the MACT professional sequence. Under these conditions the lists, and the TBS instrument based on them, were expressions of the teachers' entry level value systems rather than those shaped by participation in the MACT program. Thus, the value system implicit in the instrument should also be representative of the school systems from which the teachers came, rather than those of the instructors in the MACT program. When the assessors filled out the TBS instrument, they should then have found its implicit value system fairly congruent with their beliefs.

The final version of the TBS contained 114 items, divided into two sections:

Section I: Questionnaire on Teachers Strengths (58 items).

Section II: Questionnaire on Teacher's Possible Areas for Improvement (56 items).

Items in each section were organized in 11 categories of teacher behavior:

Planning

Organization and Management

Classroom Climate and Control

Command of Subject Matter

Teaching Method

Use of Audio-Visual and Other Materials

Understanding Human Growth and Development

Understanding Curriculum and Curriculum Development

Evaluation Procedure

Working with People

Personal and Professional Characteristics

Responses to each item were on a five point scale:

For Section I:

5 = Exceptional (This teacher is in the top 5% of teachers at this level or in this subject.)

4 = Strong (This teacher is in the top 15% of teachers at this level or in this subject).

3 = Above average

2 = Below average

1 = I have had no opportunity to observe or know this

For Section II:

5 = Strongly agree

4 = Agree

3 = Disagree

2 = Strongly Disagree

1 = I have had no opportunity to observe or know this

At the end of Section I, respondents were asked to list in rank order the five most significant strengths of the teacher being assessed. At the end of Section II, they were asked to list, again in rank order, five specific areas

in which they thought the teacher might most need to improve. In both cases, they were requested to be as specific as possible and not to restrict themselves to the categories or items on the TBS. A complete copy of the TBS questionnaire is provided in Appendix B.

(b) Purpose

The TBS was used for two purposes:

(i) Self Assessment: On three different occasions, about six weeks apart, teachers filled out the TBS. The administrations represent the pretest, posttest and delayed posttest of the teachers' self perceptions.

(ii) Feedback Data: Each teacher asked two peers and one administrator in his or her school to complete the TBS. Differences in how these data were interpreted and communicated to the teacher represent the two treatments in the experimental study.

(c) Steps in Refining the TBS

Two significant changes are made in the TBS instrument as a result of statistical analyses that focused on its

measurement properties. The most radical change prompted by this analysis was the deletion of the second section, "Questionnaire on Teachers' Possible Areas for Improvement." This section was derived from the "List of Teacher Deficiencies" developed with participants in the preparatory stage of the experiment. Whereas the first section of the TBS required that the teacher be rated in positive terms, the second section required a negative rating of performance based on levels of deficiencies or "possible areas for improvement." Furthermore, most of the items in the second section were negative versions of corresponding items in the first section. Thus, the participants themselves complained that the TBS was too long and repititious, especially since they had to complete it on three occasions as pretest, posttest, and delayed posttest.

But more fundamental than this complaint about the tediousness of the task of completing the TBS was the lack of variance of response to the second section among the assessors. A few ignored the section altogether; but most simply rated all the items either (3) or (2)--indicating that they either disagreed or strongly disagreed that the teachers needed improvement in any of the suggested areas. A few wrote comments as to the confusion or resentment they felt as a result of the negative tenor of this section. It was decided, therefore, to drop this section from further consideration. Unfortunately, time had not allowed for a pretest of the TBS survey, and the author was genuinely

surprised by the assessors' response to the second section. Their reaction was not typical of the author's experience in his native country.

The second major change in the TBS instrument was the decision to combine the "Organization and Management" and "Classroom Climate and Control" subscales into a single subscale labeled "Management". This decision was based on the following considerations:

(i) The interscale correlation for "organization" and "climate" was higher than that for any other subscale pairing.

(ii) The internal consistency (alpha coefficients) of these two subscales was relatively low. Creating a single, longer subscale improved this reliability (See Table 2).

(iii) The more general category, "Management," suggested by the combination of the two subscales was conceptually meaningful.

(d) Reliability

Based on the responses from the 41 participants in both experiments, reliability analyses were done for the total scale and for each of the 10 subscales. Table 2 provides a summary of the internal consistency of the ten subscales. As these figures suggest, the subscale reliabilities were comparatively high. As might be expected, the reliability of the total scale was even higher (coefficient alpha = .97).

TABLE 2: RELIABILITY ANALYSIS FOR INDIVIDUAL SCALES

Subscale	Coefficient Alpha
Planning	.77
Organization	.74
Climate	.76
Organization/Climate Management } Combined	.88
Subject Matter	.81
Teaching Method	.82
Use of A.V. and Other Materials	.95
Understanding Human Growth	.79
Understanding Curriculum Development	.77
Evaluation Procedure	.80
Working With People	.92
Personal and Professional Characteristics	.92

Item level analyses suggested that this figure would not be altered by the omission of any one item on the questionnaire (See Table in Appendix J). The item level analysis also revealed that item - total score correlations were consistently high across all items on the questionnaire. These correlations were greater than .60 for 37 of the 58 items and fell below .35 for only three items on the scale (6, 9 and 34).

As a result of these analyses, item 16, which had an unusually low item - total correlation and a strong negative influence on the alpha level for the subscale to which it was assigned, was dropped from the scale.

(e) Inter-Subscale Correlations

As a final step in the statistical analysis of the TBS instrument, inter - subscale correlations were determined. These Pearson Product Moment correlations are summaries in the correlation matrix portrayed in Table 3.

As the figures in Table 3 suggest, the inter-subscale correlations were typically quite high. There are at least two ways in which these data might be interpreted. Some would argue that the TBS instrument should be interpreted as a single global measure of teaching performance rather than as 10 independent sub-scales that measure distinct aspects of performance.

Others would argue that high inter-subscale correlations do not necessarily suggest that all subscales are measuring essentially the same general phenomenon. In other words,

TABLE 3: INTER-SUBSCALE CORRELATION MATRIX (PEARSON PRODUCT MOMENT CORRELATION COEFFICIENTS)

Subscales

1 = Planning
2 = Management
3 = Subject Matter
4 = Teaching Method
5 = Instructional Materials
6 = Human Growth
7 = Curriculum Development
8 = Evaluation
9 = Working with People
10 = Personal/Professional Characteristics

	Subscales									
Subscales	1	2	3	4	5	6	7	8	9	10
1	1.00									
2	.82	1.00								
3	.81	.78	1.00							
4	.73	.66	.75	1.00						
5	.55	.41	.53	.71	1.00					
6	.65	.65	.74	.78	.54	1.00				
7	.52	.46	.67	.48	.38	.56	1.00			
8	.56	.58	.67	.66	.55	.82	.69	1.00		
9	.47	.50	.59	.61	.47	.75	.55	.74	1.00	
10	.59	.56	.58	.65	.58	.73	.48	.75	.84	1.00
Total Scale	.83	.81	.88	.87	.72	.88	.71	.85	.78	.80

it is appropriate to treat the subscales as conceptually independent measures. Those who support this position would argue that teacher behavior is commonly discriminated into categories such as these, that is, teaching is a multi-faceted phenomenon. They would also argue that these categories are recognized as being conceptually different. For example, indicants of "planning" are different from the indicants of "working with people" in that sets of actions which teachers perform, grouped together as "planning," are conceptually different from sets of actions grouped together as "working with people."

It seems, therefore, most useful to look at the categories as representing different aspects of teacher behavior, which are nevertheless highly correlated. For, it might well be that good teachers who are good in management in the classroom are in practice very likely to be also good in planning or in their command of subject matter.

The most striking statistics in Table 3 are the high correlations between each of the subscales and the total scale. These figures range from .71 to .88. The lowest of these were for subscale 7, Curriculum Development (r = .71), and for subscale 5, Instructional Materials (r = .72). It is also interesting to note that the interscale correlations were lowest for this pair (r = .38).

The three subscales which correlated most highly with the others and with the total scale were subscale 3, Subject

Matter (r = .88), subscale 6, Human Development (r = .88), and subscale 4, Teaching Method (r = .87).

It would seem, then, that the three strongest categories from this scale for describing teacher performance in global terms are Subject Matter, Human Development, and Teaching Method. In general, these might well be the dominant characteristics of the able teacher. On the other hand, the aspects of teaching performance that would be least closely related to global measures were Curriculum Development and Instructional Material.

(f) Validity

As the TBS is an instrument developed by the investigator for the purpose of this study, there is of course no prior evidence of its validity. Nevertheless, evidence of "face validity" does exist. First, the categories of teacher behavior on which the TBS is based, have served on numerous occasions as a framework for self analysis activities in a graduate course taught by the professor, Dr. Cragun, with whom the investigator worked for the duration of the project. According to Dr. Cragun, teachers have always responded favorably to his classification of teaching performance. Second, the process by which the items on the TBS were derived should insure high levels of "face validity" when viewed by teachers (See Cragun and Wilson, 1980). Finally, the participants in both experiments found the first section of the TBS most helpful, and the assessors who used it commented favorably. A

couple of assessors, who were principals, have expressed interest in its further use in their schools.

A vote of confidence in its value was also expressed by the rate of response to its use in the investigation. The participants in Experiment I were to some extent captive within the project, but all 78 assessors also completed and returned the TBS on time, and 15 second year MACT candidates filled them out and returned them on three separate occasions, after only two short meetings with the investigator. This was a very high rate of response, in spite of the fact that many found the second section of the TBS repititious.

But the best evidence of the validity of the TBS instrument will be the extent to which results based on this instrument support the conjectures underlying this investigation. The presentation and analysis of the data in the next chapter should, therefore, be viewed as a test of the validity of the TBS instrument as well as a test of the conjectures on which the study is based.

Description of Evaluator

This form was also developed for the purpose of this investigation. It is a short questionnaires with 12 items, designed to get a description of the relevant personal and professional characteristics of the assessor, and also to check on the manner in which each operated. A copy of this form is in Appendix E, and a summary of the data collected is in Appendix F.

Debriefing for Participants in the Two-Term MACT Professional Evaluation

This form, also referred to as the "Debriefing Form" or "Attitude Scale", was the third to be developed by the investigator for this study. Its purpose was to provide data regarding the teachers'

a. appraisal of the effectiveness of the various instructional processes, which they experienced during the investigation, and

b. attitudes towards the various parts of the program of evaluation, and the program as a whole.

There are four sections to this form. The first set of items asked teachers to indicate the extent to which they felt specific activities in the two term evaluation process influenced their perceptions of relative strengths and shortcomings as a teacher. The next set of 17 items asked respondents to indicate their level of agreement with statements that described favorable or unfavorable attitudes toward specific or general features of the evaluative process. The third set of five items asked teachers to describe their response to the assessors and the feedback they received. The final set of five items invited respondents to complete a set of open-ended statements that provided an opportunity to elaborate on their feelings about the program.

Participants were asked to type or print their comments so as to preserve anonymity. This procedure was designed to encourage honest answers even when these were likely to be unfavorable.

A copy of this debriefing form is in Appendix C.

Dependent Variables

Analyses of responses to the TBS questionnaire provide measures of each of the following dependent variables: (a) Feedback Differential, (b) Change Scores I, (c) Change Scores II, and (d) Change Scores Differnetial.

A. Feedback Differential

Feedback differential is the difference between a teacher's original self perceptions and the perceptions of his or her three assessors. The teacher's original self perceptions were measured by the pretest. The perceptions of the assessors were taken as the average of their ratings. Feedback differential was, therefore, the difference between pretest and the mean feedback scores.

In this investigation, feedback differential is analyzed at two levels--at the subscale and at the total scale levels. For each subscale, feedback differential is the difference between the means of the pretest and the feedback scores. For the total scale, it is the difference between the weighted means of pretest and feedback scores.

If feedback differential is large and positive, it suggests that assessors' ratings were considerably higher than the teacher's self ratings. Negative feedback differential scores occurred when assessors' ratings were lower than the teacher's self rating.

B. Change Scores I

A change score in this investigation is a measure of the degree of change in a teacher's self perceptions. Change Scores I refers to the difference between the teacher's original self perceptions as measured by the pretest and his or her self perceptions six weeks later, as measured by the posttest. The means of subscales and the total scale for the pretest were subtracted from the corresponding subscale and total scale means for the pretest. These constituted Change Scores I. They represented an attempt to measure any changes that might have occurred in the teacher's self perceptions over the specified period. These changes could have been caused by the treatment received by the participants in Experiment I, or by potential confounding variables, which were the only conditions affecting the participants in Experiment II. Large change scores suggest that a teacher may have altered his or her self perceptions to a considerable extent. Small change scores suggest that the teacher's self perceptions may have remained unchanged or been only slightly altered.

C. Change Scores II

The second set of change scores attempted to measure the changes in teachers' self perceptions which might have occurred over the longer period of time from pretest to the delayed posttest, which was administered six weeks after the

posttest. These scores were derived in exactly the same way as for Change Score I, except that mean scores on the delayed posttest were substituted for mean posttest scores.

D. Change Scores Differential

One of the purposes of the investigation was to test the stability of any changes in self perceptions that took place during the time of the investigation as measured at six week intervals. Change Scores I measured such changes as might have taken place six weeks after the teachers' self perceptions were originally measured by the pretest. Change Scores II measured the changes six weeks later. Change Score Differential was the difference in corresponding mean scores at the subscale and total scale levels between Change Scores I and II. It should represent those changes in self perceptions that took place between posttest and delayed posttest, and so, should give a measure of the relative stability or instability of any changes in self perceptions that occurred as a result of treatment or non treatment conditions. A large change score differential suggests a great deal of instability in a teacher's post-treatment self-perceptions, whereas a small change score differential suggests a greater degree of stability.

E. Computational Formulas

The computational formulas that guided the derivation of the dependent variables may be briefly summarized as follows:

(1) Feedback Differential

(a) For each scale: Subtract subscale for pretest from corresponding subscale mean for feedback.

(b) For total scale: Subtract weighted total scale mean for pretest from weighted total scale mean for feedback. To get the total scale mean, multiply the mean score for each category by the number of responses in the category, add them, and divide by the total number of responses.

(2) Change Scores I

(a) For each scae: Subtract subscale mean for pretest from corresponding subscale mean for posttest.

(b) For total scale: Subtract weighted total scale mean for pretest from weighted total scale mean for posttest.

(3) Change Scores II

Same as for Change Scores I, substituting delayed posttest means for posttest means.

(4) Change Score Differential

(a) For each scale: Subtract subscale score for Change Scores I from corresponding subscale score for Change Scores II.

(b) For total scale: Subtract total scale score for Change Scores I from total scale score for Change Scores II.

Internal and External Validity of the Study

The following discussion of the validity of the design of the study is guided by the arguments of Porter (1980), and Campbell and Stanley (1963). According to these authors, the quality of an experimental design should be judged in terms of internal validity, precision and external validity.

A. Internal Validity and Precision

Porter defines internal validity in the following way:

> If the independent variable is the only reasonable explanation of the differences in the dependent variable, the study is said to have internal validity.

Confounding variables, which are alternative explanations of the differences in dependent variables, are the main threats to internal validity. The best way to control confounding variables at the beginning of an experiment is to randomly assign the subjects from the initial pool to the experimental treatments. Porter cautions, however, that:

> The utility of random assignment for controlling variables must, of course, be tempered by the realization that the process is based on chance, and by chance alone, experimental groups will differ at least to some extent.

He goes on to point out that the smaller the pool of subjects initially identified for the experiment, the greater the chance that there might be "worrisome differences between experimental groups at the outset of the experiment," when random assignment only is used. To lessen this possibility, the largest pool of subjects should be sought.

About precision, Porter writes:

> One way to think about precision is that the more precise an experiment, the less likely it is that the experiment will yield large chance differences between experimental groups.

One way to improve precision is, again, "to increase the number of subjects assigned to each experimental condition." And, as pointed out above

> ...The larger the number of subjects to be randomly assigned, the less likely it is that random assignment will result in unusual groups of subjects.

With respect to these two criteria of internal validity and precision, this study would seem to be caught in double jeopardy. The small number of subjects in the experimental groups in Experiment I (13 each) and in Experiment II (15) would seem to severely threaten the internal validity and precision of the investigation. Porter, however, does suggest a way out of this double predicament. Using the example of student aptitude as a variable in a particular experiment, he writes:

> If you want to make sure that not all of the most able students are assigned to a single experimental condition, the most straightforward method is to first group students according to aptitude. Then for each aptitude group of students, randomly assign equal numbers to each experimental condition. This insures that each experimental condition has subjects with similar aptitude levels. The procedures is called blocking and is one way to improve precision.

So, in designing the first experiment in this study, recognizing the problems in having a small number of subjects, the investigator decided to block the subjects on three variables: sex, location and school level taught. Then having done this, from each block he randomly assigned matched pairs of subjects to the experimental groups, as described earlier in this chapter. By this device, he hoped to avoid the "worrisome differences" between experimental groups on these three variables. It was also hoped in this way to increase precision in the experiment.

There are at least two other variables that may confound the results: age and years of teaching. However, practical limitations in the ability to form matched pairs on several variables as well as anomalies in the distribution of participants across these variables prevented the investigator from blocking on age and teaching experience.

Thus, worrisome differences in age and teaching experience did exist at the outset of Experiment I. These differences are summarized in Table 4. As the data in this table suggest, members of the SAG group were younger and had fewer years of teaching experience than was true for the EAG group.

TABLE 4: IMPORTANT DEMOGRAPHIC DIFFERENCES ACROSS EXPERIMENTAL GROUPS

Group	n	Mean Age	Mean Years Taught
SAG	13	31.5	8.5
EAG	13	37	12.5
SAG & EAG	26	34	10.5
EIIG	15	37.5	12.5
Total	41	35.3	11.2

These differences must therefore be considered when interpreting findings of this investigation. In short, age and teaching experience may serve as confounding variables in this study.

As defined by Campbell and Stanley (1963), between the pretest and posttest many types of confounding variables can also develop during the course of a study even though random assignment may have ensured an unblemished beginning. The experimental designs in both experiments, where three tests were taken over a period of 12 weeks, would seem to make this study vulnerable in four respects: history, maturation, testing effects and mortality.

With regard to history, even within the context of the course they were doing as part of the MACT program, the groups in both experiments completed major action research projects and interacted with one another and with the instructor in ways that could have altered their professional self perceptions. Whether these effects were uniform or not, they still could have caused changes in the dependent variable not due to the influence of the experimental treatments. Maturation also could have been a factor affecting the dependent variable in both experiments. So, too, could testing effects, where the subjects took the same test on three occasions. Responses on the second and third administration may have been influenced by the teachers' recall of how they had marked the items in the past. A negative attitude could also have developed as a result of too frequent exposure to the same

test. The effect of experimental mortality will be discussed a little later in this section.

One of the main reasons for undertaking Experiment II was to "control" for the effects of three of the four potential confounding factors mentioned above. The participants in Experiment II (EIIG) shared the same sort of history, maturation and testing effects as experimental groups, but did not receive either of the experimental treatments. Thus it does not seem that differences between experimental and control groups could be attributed to systematic differences on these three variables. It might, however, be argued that the members of the control group could have been negatively affected by a sort of Hawthorne effect. For, they did not get the same level of attention which the members of the treatment groups experienced, particularly in the process of receiving or reporting on their feedback. The investigator could have controlled for this effect but felt that a half an hour visit with each of the participants would not have had an impact on their self perceptions in any of the areas under study.

But there were two other confounding variables which Experiment II could not "control". One was an aspect of history peculiar to the experimental groups, and the other was experimental mortality. The history of Experiment I in itself could have nurtured elements of the Hawthorne effect, or it may have had the opposite effect, namely, growing resentment at being asked to satisfy the demands of the experiment during a busy term.

Experimental mortality also had an effect in Experiment I. Two participants, because of serious illness, had to be dropped from the experiment. One withdrew in the last two weeks of the Spring Term, the other got ill during the summer holidays when the delayed posttest was due. Both were female elementary school teachers from City B; so, on these two variables they constituted a pair. However, they left their counterparts from City A without a match, with the result that the variable, geographical location, was now unevenly distributed between the groups.

Over the duration of Experiment II, the EIIG group also lost two members because they did not hand in one of the tests. But this is not seen to have had a serious effect on that experiment.

The professional background of the 15 participants in Experiment II were very similar to those of the participants in Experiment I (See Appendix I). On the other hand, certain differences between these two groups suggest that participants in Experiment II should not be viewed as a "true" control group. The main differences were time spent in the MACT program, and to a lesser extent, restricted geographical location. Nevertheless, results from the second year MACT teachers in the EIIG group should give some picture of what changes teachers in an inservice program of this sort, who have not been given the experimental treatments, are likely to experience in terms of their professional self

perceptions over a period of 12 weeks. To this extent, qualified comparisons with the experimental groups will be attempted.

In summary, there is at least some reason to question the internal validity and precision of the experiments. Attempts were made, however, to identify and neutralize the major effects of a number of confounding variables. Because blocking was used, the virtues of random assignment and precision, though beleagured by the fact of a small pool of participants, should have been preserved. Nevertheless, age and years in teaching remain as potential confounding variables.

B. External Validity

External validity is concerned with the extent to which valid generalizations can be made from a study. In dealing with external validity, Porter quotes Bracht and Glass (1965) as follows:

> ...threats to external validity appear to fall into two broad classes: (1) those dealing with generalizations to populations of persons (what population of subjects can be expected to behave in the same way as did the ...experimental subjects?), and (2) those dealing with the environment of the experiment (under what conditions ... can the same results be expected?)

Both of these threats to external validity will be considered.

What Populations?

To make generalizations to a population, ideally two conditions are necessary: a well defined population, and

random selection from that population. The problem is that random selection from a large well defined population is rarely the case in experiments. Porter pursues the consequent argument in this way:

> ...since random selection is rarely used in experiments, are the results of experiments typically limited to the subjects used? Most people would answer these questions with a decided "no". People who are willing to generalize from an experiment, however, require a careful description of subjects. The description defines a hypothetical population to be used for purposes of interpreting results.

It is possible, then, for the results of an experiment to be generalized beyond its subjects to a hypothetical population based on a description of the subjects. In order to provide the reader with such a basis for the projection of an acceptable hypothetical population, a full description of subjects is provided in Appendix I.

However, one critical characteristic of the sample must be considered. As described earlier, all teachers in the sample had engaged in an intensive self assessment exercise prior to their participation in the experiment. This characteristic tends to set them apart from other comparable groups of teachers and may therefore severely restrict the generalizability (external validity) of the study. This characteristic, however, also provides a unique opportunity to examine the impact of feedback on teacher self perceptions. If participants with this background do change their self perceptions in accord with feedback they receive, it is logical to infer that other teachers will also alter their self

perceptions under these conditions. In short, these unique background experiences of the participants may limit external validity, but they also provide a powerful test of the degree to which feedback may alter a teacher's perceptions of relative strengths and shortcomings.

Under What Conditions?

Certain conditional features of this investigation should be noted. With regard to the experimental treatments, the need to limit the degree of spontaneous interaction between investigator/instructor and participants/students during the reporting sessions for both experimental groups (SAG and EAG) was a distinguishing feature. It is less typical of evaluators, especially in the role of instructor or supervisor, not to interact more flexibly with their students or clients. This controlled interaction was a particularly troublesome feature of the EAG treatment.

A positive feature of this study, where external validity is concerned, is the nature of the dependent variables. As Porter states:

> External validity can also be threatened to the extent that the experimenter limits his consideration of dependent variables.

In this study, professional behavior, which is the object of the teachers' self perceptions, is widely represented in its various aspects in the subscales of the Teacher Behavior Survey.

Porter also warns that:

> Differences on a variable observed immediately following the experiments may not perservere to later points in time.

This study seeks to obviate that threat by a repeated measures design that includes a delayed posttest as well as an immediate posttest.

The concluding judgement might well be that the external validity of this investigation, in so far as a general population of inservice teachers is concerned, is severely restricted by the self analysis backgrounds of the participants. The conditions of the experiment, especially the need for controlled interactions, also caution against making blind generalizations. But the value of this investigation is in its attempt to explore a complex, practical question under conditions in which feedback might be least apt to influence self perceptions. In brief, the sharp focus of the experiment might discourage replication, but it should increase confidence in those inferences that are formed.

Research Questions

Two sets of questions are addressed by this investigation. The first set is concerned with changes in teacher self perception. The second set is concerned with the attitudes of the participants in the experiment.

A. Questions Concerned With Changes in Self Perception

Question 1: Does either of two forms of feedback--self analyzed and reported (SAG), or externally analyzed and reported (EAG)--cause any change at all in a teacher's self perception of the relative levels of his or her abilities?

Question 2: Which of two forms of feedback (SAG or EAG) will yield the greatest changes in a teacher's self perceptions?

Question 3: Is the magnitude of change in self perceptions resulting from assessment feedback a function of empirically established levels of agreement (high, low) between feedback received and original self perceptions?

Question 4: Will the changes in self perceptions revealed in the posttest immediately following the experimental intervention (seven to 10 day interval) differ from corresponding changes in perceptions revealed in a delayed posttest administered approximately six weeks after the intervention?

B. Questions Concerned With Attitudinal Outcomes

Question 1: What elements in the two term evaluation program did the participants find most valuable or least valuable?

Question 2: How did they feel about participating in the program?

Question 3: What were their perceptions about the feedback they perceived?

Analytical Procedures Used

A. Summary of Statistics Used

1. Questions Concerned With Changes in Self Perception.

Data for the four research questions concerned with changes in self perceptions were analyzed using one of three statistical tests: *t*-tests, correlation coefficients, and analyses of variance. For all analyses, alpha has been fixed at .05. A simple summary of these tests is as follows:

Question 1: Are there changes in self perceptions following feedback?

(a) *t*-tests of means for Change Scores I and II at subscale and total scale levels for experimental groups (SAG and EAG) and for EIIG.

(b) Tests to determine if correlations between Feedback Differential and Change Scores I or II at subscale and total scale levels are significantly different from zero.

Question 2: Are there differences between SAG and EAG groups?

(a) ANOVA tests of means for Change Scores I and II at subscale and total scale levels for SAG, EAG and for EIIG.

(b) Tests to determine the correlations between Feedback Differential and Change Scores I or II at

subscale and total scale levels and to compare these coefficients by groups.

Question 3: Is the level of change in self perceptions a function of the level of feedback differential? Tests to determine the correlations between Feedback Differential and Change Scores I or II at subscale and total scale levels for the combined experimental groups (SAG and EAG).

Question 4: How stable are the changes in self perceptions?

(a) ANOVA tests of means of Change Score Differentials at the total scale level by groups (SAG and EAG combined and EIIG).

(b) Tests to determine if correlations between Change Scores I and II at subscale and total scale levels for SAG, EAG and EIIG are significantly different from zero.

2. Questions Concerned With Attitudinal Outcomes

The questions concerned with attitudinal outcomes focus on three aspects of the participants' reactions to the program: first, what they found most or least valuable in the program; second, how they felt about participating in it; and third, what their perceptions were with respect to the feedback they received.

For each of the three research questions in this section, mean scores of all items concerned with the question will be calculated and revised. For the open ended items a

frequency count will be made of the times a given activity or program feature is mentioned.

B. Why Data Is To Be Analyzed In This Way

The four research questions on self perceptions are motivated by the more general question of whether feedback which is different from one's initial self perceptions changes these self perceptions or not. The concern is with differences between feedback and original self perceptions, and with changes that may have occurred later as a result of these differences. Thus, the nature of the investigation requires that difference scores, or change scores, be the focus of analysis.

Where it is necessary to see whether any change at all has taken place for a single or combined group, simple *t*-tests are appropriate. These test whether an apparent change is significantly different from zero, or should be attributed to change.

Where it is necessary to distinguish between group differences, again *t*-tests (for 2 groups) or one way analyses of variance (for more than two groups) are appropriate for two reasons. First of all, only single dependent variables are being tested; and, second, these tests discriminate between statistically significant and chance differences between groups.

A number of tests for significant correlations are to be performed. These are useful to test whether relationships

between feedback and original self perceptions are significantly related to later changes in self perceptions. The relationship between immediate and delayed change scores are also to be tested for correlational significance. A consistent pattern of significant correlations would suggest a high degree of stability in those changes that occur.

The alpha level, or level of significance, has been set at .05, which indicates a five percent likelihood of rejecting the null hypothesis when it is true. This is the most commonly chosen level, and seem suitable for this investigation.

The simple reporting of mean scores and frequency counts should provide adequate answers to the questions concerning attitudinal responses.

CHAPTER IV

PRESENTATION AND ANALYSIS OF DATA

Part One: Changes in Self Perceptions

Introduction

The central concern of this investigation was to determine to what extent the self perceptions of teachers who have undergone a thorough process of self analysis may nevertheless be altered by either of two types of external data feedback. If, after this training, teachers do not significantly alter their perceptions as a result of either form of external data feedback, then there is good reason to believe that they are in a good position to make decisions about their needs for professional development following an initial process of self analysis. If, on the other hand, teachers do respond to additional feedback data from significant others, the provision of feedback may represent an essential step in needs assessment.

Four specific research questions were addressed by this investigation. The data analysis that focus on each of these questions will be summarized in the sections that follow. While data relating to these questions were being analyzed, an interesting question arose which could not be

answered within the context of the original four questions. The new question concerned individual responsiveness to external data feedback. This was a more exacting question than any of the original four, and will therefore be considered in the final section of this chapter.

> Question 1: Does either of two forms of feedback--self analyzed and reported (SAG), or externally analyzed and reported (EAG)--cause any change at all in a teacher's self perception of the relative levels of his or her abilities?

This question asks whether feedback received by the teachers will cause them to change the perceptions of their professional behavior which they held at the conclusion of the self analysis course. The most direct answer to this question was to determine whether the changes in self perceptions that occurred for the combined treatment groups between pretest and posttest were significantly different from zero. These results had then to be compared with similar results for the control group (EIIG) in an attempt to determine whether differences in pre and post test performance should be attributed to treatment effects or to some other source.

Means for the two sets of change scores (Posttest-Pretest and Delayed Posttest - Pretest), for the ten subscales and the total scale are presented in Tables 5 and 6. Results of *t* - tests which compare these means with a value of zero are also described in these tables. Table 5 summarizes the

TABLE 5: SUBSCALE AND TOTAL CHANGE SCORES I (POSTTEST-PRETEST) BY GROUPS: MEANS, STANDARD DEVIATION AND RESULTS OF *T* - TESTS

Subscales

1 = Planning
2 = Management
3 = Subject Matter
4 = Teaching Method
5 = Instructional Materials
6 = Human Growth
7 = Curriculum Development
8 = Evaluation
9 = Working with People
10 = Personal/Prof. Character.

Group	n		Subscales 1	2	3	4	5	6	7	8	9	10	Total
SAG	13[a]	$\bar{x}$	.15	.08	-.03	-.08	-.10	-.11	-.11	-.23	.05	.19	.01
		SD	.49	.59	.58	.81	.83	.71	.59	.88	.87	.61	.60
EAG	13	$\bar{X}$	.19	.33	.37	.42	.18	.57	.23	.46	.55	.43	.37
		SD	.60	.46	.58	.67	.64	.65	.60	.47	.50	.43	.43
SAG & EAG Combined	26[a]	$\bar{X}$	.17	.21	.17	.17	.05	.23	.06	.12	.30	.31	.19
		SD	.54	.53	.60	.77	.73	.75	.61	.77	.75	.53	.55
		t	1.57[1]	1.95*	1.41[1]	1.10	.30	1.54[1]	.51	.74	2.01*	2.92*	1.74*
EIIG	15	$\bar{X}$	.12	.10	-.01	.07	-.21	.09	.03	-.10	-.01	-.17	-.01
		SD	.43	.44	.42	.60	.61	.47	.71	.90	.67	.50	.43
		t	1.04	.85	-.06	.42	-1.30	.72	.14	-.40	-.07	-1.29	-.04

[a] n = 12 on subscale 5

[1] $p < .10$

* $p < .05$

TABLE 6: SUBSCALE AND TOTAL CHANGE SCORES II (DELAYED POSTTEST-PRETEST) BY GROUPS: MEANS, STANDARD DEVIATIONS AND RESULTS OF *T* - TESTS

Subscales

1 = Planning
2 = Management
3 = Subject Matter
4 = Teaching Method
5 = Instructional Materials
6 = Human Growth
7 = Curriculum Development
8 = Evaluation
9 = Working With People
10 = Personal/Professional Characteristics

Group	n		1	2	3	4	5	6	7	8	9	10	Total
SAG	13[a]	$\bar{X}$	.06	.11	.03	.02	.26	-.03	-.09	-.04	.11	.08	.07
		SD	.65	.60	.63	.71	.66	.80	.64	.92	.60	.74	.60
EAG	13	$\bar{X}$	.25	.34	.21	.51	.49	.52	.15	.42	.39	.42	.36
		SD	.54	.40	.59	.77	.69	.55	.57	.52	.56	.53	.45
SAG & EAG Combined	26[2]	$\bar{X}$	.15	.22	.12	.26	.38	.25	.03	.19	.25	.25	.22
		SD	.59	.51	.60	.77	.67	.73	.61	.77	.59	.66	.54
		t	1.30	2.18*	1.01	1.70[1]	2.73*	1.68[1]	.27	1.22	2.10*	1.89*	1.99*
EIIG	15	$\bar{X}$	.04	.06	-.02	.05	-.15	.16	.21	.00	.11	.03	.05
		SD	.47	.55	.71	.72	.70	.66	.57	.89	.62	.54	.50
		t	.32	.41	-.11	.27	-.79	.90	1.37[1]	.00	.65	.21	.40

[a] n = 11 on subscale 5

[1] $p < .10$

* $p < .05$

results for Change Scores I (Posttest - Pretest). This table indicates, first of all, that the means for the combined treatment groups were in all cases higher than those for the control group. The corresponding *t*-tests suggest that the means for the treatment groups were significantly different from zero for the total scale and for three subscales when the likelihood of Type I error was set at .05. Also, three other subscale means for the treatment groups approached statistical significance. On the other hand, none of the means for the control group were significantly different from zero under the same test conditions. These results suggest that there were greater changes in self perceptions for the treatment group than for the control group.

A study of the data in Table 6 reinforces the insights already gained. Means for Change Scores II (Delayed Posttest - Pretest) were always higher for the combined treatment groups than for the control group (EIIG). Furthermore, *t* - tests of the means for the combined treatment groups were significantly different from zero for the total scale and for four subscales, when the likelihood of Type I error was set at .05. The means of two other subscales approached statistical significance. None of the means for the control groups were significantly different from zero, and only one approached statistical significance under these test conditions.

The strong impression left from a study of Tables 5 and 6 is that the members of the treatment groups changed

their self perceptions to a greater extent than did the members of the control group. And these changes from the original pretest levels were greater for the treatment group members both as measured in the short and long term by the posttest and the delayed posttest. In general, there is a strong suggestion that feedback influenced the considerable changes in self perceptions that took place for the members of the treatment groups.

A further study of the data in Table 5 reveals that the means and *t* - tests results for the combined treatment groups were all positive, while six of the eleven means for the control group were small and negative. (The latter was also the case for teachers who were members of the SAG treatment group, but this negative tendency was overcome by the strong positive scores for the EAG group). In Table 6, all of the combined treatment group means were positive, whereas two of the control group means, and three of the SAG means were negative. The significance of these relative positive or negative tendencies must be examined in the light of the data summarized in Table 7.

The data summarized in Table 7 show that for every subscale and for the total scale, the means for feedback differential (Feedback-Pretest) were high and positive for the treatment groups, with most scores over .50. Feedback scores were, therefore, typically higher than the teachers' original self perceptions as reflected in pretest scores.

To return to the study of the data in Tables 5 and 6, the fact that the mean difference between posttest and

TABLE 7: SUBSCALE FEEDBACK DIFFERENTIAL SCORES (FEEDBACK- PRETEST): MEANS, STANDARD DEVIATIONS AND *F* - RATIOS

Subscales

1 = Planning
2 = Management
3 = Subject Matter
4 = Teaching Method
5 = Instructional Materials
6 = Human Growth
7 = Curriculum Development
8 = Evaluation
9 = Working With People
10 = Personal/Professional Characteristics

Group	n		Subscales 1	2	3	4	5	6	7	8	9	10	Total
SAG	13	$\overline{X}$	.63	.61	.57	.54	.73	.24	.39	.59	.47	.50	.55
		SD	.66	.56	.61	.93	.75	.62	.58	.58	.78	.67	.52
EAG	13	$\overline{X}$	.49	.48	.64	.70	.73	.64	.56	.54	.53	.51	.59
		SD	.71	.70	.81	.90	1.06	.85	.71	.72	.76	.71	.70
		F	.26	.27	.06	.30	.00	1.55	.44	.06	.05	.001	.050

pretest scores were positive seems to suggest that these teachers had been positively influenced by feedback, that is, they changed their self perceptions in the direction of the positive feedback they received. The control group (EIIG), by definition received no feedback, and the majority of the means for this group were small and negative.

Up to this point, the data analyses have been based on means and tests of their significance. Another kind of analysis--one based on a comparison of correlations--also gives support to the findings that have been described so far. The process used was to correlate feedback differential scores with both sets of change scores to determine whether there is a link between feedback and later changes in the teachers' self perceptions.

The first step in deriving these correlations was to arrange each participant's scores on the three critical variables in rows as in Table 8. Each of the 11 columns (one per subscale and total scale) of 26 feedback differential scores was then correlated in turn with the corresponding columns of Change Scores I and II. The first column of 26 feedback differential scores for subscale one, for example, was correlated with the first column (subscale 1) of Change Scores I, and then with the first column (subscale 1) of Change Scores II. In this way, it was possible to determine that the correlation between feedback differential scores and Change Scores I for the total sample was .67 for subscale 1 (See Tables 8 and 9). The corresponding correlation coefficient for Change Scores II was .71.

TABLE 8 AN ILLUSTRATION OF HOW FEEDBACK RESPONSIVENESS CORRELATIONS WERE COMPUTED

Participants	Feedback Differential Scores - (Subscale 1) Feedback - Pretest	Change Scores I (Subscale 1) Posttest-Pretest	Change Scores II (Subscale 1) Delayed Posttest-Pretest
1	1.19	.40	.60
2	.67	-.80	-1.60
3	.57	.40	- .20
4	.60	.20	.00
5	-.20	.00	- .40
6	.20	-.40	.20
. . .			
26	1.07	1.00	1.00

$r_1 = .67$

$r_2 = .71$

Table 9 describes the complete set of correlations between feedback differential scores and the two change scores. The correlations were computed for SAG and EAG separately (n=13), and then for both groups combined (N=26). In view of their relevance to question one, only the results for the combined groups will be discussed in this section.

Part A of Table 9 shows that the 11 correlation coefficients representing total and subscale relations between feedback differential and Change Scores I for the total group (N = 26) were consistently very high and were significantly different from zero, when the likelihood of Type I error was set at .01. Part B shows exactly the same pattern of relationship between feedback differential and Change Scores II. In fact, these correlations were even somewhat higher than those for Change Scores I. Eight of the eleven correlations were above .80.

In brief, the data in Table 9 provide convincing evidence that there was a strong positive correlation between feedback differential scores and both measures of change in self perception. In other words, those who received higher feedback scores than they had awarded themselves on the pretest were likely to change their self perceptions in this direction; those who received lower feedback scores were likely to lower their self perceptions. When these results are combined with those summarized in Tables 5 and 6, there can be little doubt that the 26 teachers in the sample were strongly influenced by the feedback they

TABLE 9: CORRELATIONS BETWEEN CHANGE SCORES I OR II (POSTTEST - PRETEST or DELAYED POSTTEST - PRETEST) AND FEEDBACK DIFFERENTIAL (FEEDBACK - PRETEST)

Subscales

1 = Planning
2 = Management
3 = Subject Matter
4 = Teaching Method
5 = Instructional Materials
6 = Human Growth
7 = Curriculum Development
8 = Evaluation
9 = Working with People
10 = Personal/Professional Characteristics

A. Correlation Coefficients Feedback Differential/Change Score I

Group	n	1	2	3	4	5	6	7	8	9	10	Total
SAG	13[a]	.61**	.28	.34	.11	.61*	.35	.39[1]	.34	.73**	.57*	.42[1]
EAG	13	.74**	.87**	.76**	.77**	.82**	.82**	.75**	.77**	.90**	.88**	.84**
SAG & EAG Combined	26	.67**	.51**	.56**	.46**	.69**	.61**	.59**	.39**	.73**	.67**	.57**

B. Correlation Coefficients: Feedback Differential/Change Scores II

Group	n	1	2	3	4	5	6	7	8	9	10	Total
SAG	13[b]	.41[1]	.38[1]	.52*	.48*	.43[1]	.41[1]	.55*	.44[1]	.63**	.47*	.49*
EAG	13	.69**	.70**	.56*	.83**	.76**	.77**	.53*	.82**	.76**	.87**	.76**	
SAG & EAG Combined	26	.71**	.91**	.84**	.89**	.86**	.88**	.69**	.88**	.85**	.89**	.95**

[a] n = 12 on subscale 5 [b] n = 11 on subscale 5

[1] p < .10 *p < .05 **p < .01

received. Despite the training in self analysis these individuals had received, there were significant and predictable changes in self perceptions that resulted from their receiving feedback from others.

Question 2: Which of the two forms of feedback (SAG or EAG) will yield the greatest changes in a teacher's self perceptions?

This question seeks to determine whether those participants in the SAG who analyzed and reported on the data themselves were more affected by feedback than those in the EAG for whom the data were analyzed and reported by an external evaluator.

In order to provide a convincing answer to this question, it was important to demonstrate that the experimental conditions were such that:

a. the three groups (SAG, EAG and control) were similar in pretest performance (self perceptions did not differ prior to the start of the experiment), and

b. the two treatment groups received the same level of feedback differential (one group did not receive more favorable feedback than the other).

To test the equality of the groups in these two fundamental areas, tests were carried out in response to two specific questions:(1) How similar were the groups' initial self perceptions as reflected by mean scores on the pretest?

(2) To what extent were the groups alike with respect to the feedback they received as reflected by mean feedback differential (feedback-pretest) scores? The second question was particularly important. If the participants in one experimental group had received feedback that was much higher or lower than their initial self perceptions when compared to the other group, then this contaminating variable could account for any difference in self perception that occurred. The data summarized in Part A of Table 10 attempt to answer the first question; those in Part B address the second.

In Part A, the group means for the total pretest scores are seen to be minimally different, only .13 separating the most widely different groups (EAG $\bar{X}$ = 3.73, EIIG $\bar{X}$ = 3.86). The analysis of variance yielded an F - ratio of less than 1.00. Thus, the groups do not appear to be significantly different in terms of their initial self perceptions.

The data in Part B of Table 10 suggest that the three means for total feedback differential were even closer to each other. A mean difference of only .04 separated the two treatment groups. The analysis of variance produced a correspondingly miniscule F - ratio. The groups, therefore, appear to be substantially equal where total feedback differential is concerned. This finding was supported by more exacting analyses done at the subscale level. As the results summarized in Table 7 suggest, mean feedback differential

TABLE 10 TESTS OF THE EQUALITY OF GROUPS: MEANS, STANDARD DEVIATIONS AND RESULTS OF ONE-WAY ANOVA TESTS FOR PRETEST AND FEEDBACK DIFFERENTIAL SCORES

A. Pretest

Group	n	$\overline{X}$	SD	Analysis of Variance Source	df	MS	F
SAG	13	3.81	.64	Main Effect	2	.06	.18
				Error	38	.34	
EAG	13	3.73	.55				
EIIG	15	3.86	.46				

B. Feedback Differential: Feedback - Pretest

Group	n	$\overline{X}$	SD	Analysis of Variance Source	df	MS	F
SAG	13	.55	.52	Main Effect	1	.019	.050
				Error	24	.383	
EAG	13	.59	.70				
SAG & EAG Combined	26	.57	.61				

scores for the SAG and EAG groups were very close across all the subscales; none of the differences approached statistical significance.

It is a strong feature of this study that the two groups did not differ significantly in initial self perceptions or feedback differential. As a consequence, results of tests of subsequent changes in self perceptions can be that much more persuasive.

Perhaps the simplest way to address the question as to which of two forms of feedback (SAG or EAG) will yield the greatest changes in a teachers' self perceptions is to determine whether or not changes in total scores vary across these two groups. Means for the two change scores (posttest - pretest and delayed posttest - pretest) are summarized in Table 11. As an examination of these data suggest, the mean for Change Scores I was higher for the EAG than for the SAG or the control group EIIG. The corresponding analysis of variance tests suggests that these differences approached but did not reach statistical significance when alpha was set at .05. Means of Change Scores II do not vary to the same extent. Once again, the mean of the change score for the EAG was higher than for either the SAG or "control" group. However, as the results of the corresponding analysis of variance test suggest, these differences did not approach statistical significance. In brief, the data summarized in Table 11 suggest that the changes in self perception were slighly higher for the members of the EAG than for

TABLE 11: GROUP COMPARISONS OF CHANGE SCORES I AND II: MEANS DEVIATIONS AND RESULTS OF ONE-WAY ANOVA TESTS

Change Score I: Posttest - Pretest

Group	n	$\bar{X}$	SD	Analysis of Variance Source	df	MS	F
SAG	13	.01	.60	Main Effect	2	.62	2.56[1]
				Error	38	.24	
EAG	13	.37	.43				
EIIG	15	-.01	.43				

Change Score II: Delayed Posttest - Pretest

Group	n	$\bar{X}$	SD	Analysis of Variance Source	df	MS	F
SAG	13	.07	.60	Main Effedt	2	.41	1.52
				Error	38	.27	
EAG	13	.36	.45				
EIIG	15	.05	.50				

[1] $p < .10$

members of the SAG or control groups. Differences across the three groups were most pronounced when posttest -pretest scores served as the dependent variable. However, none of these differences were found to be statistically significant when alpha was fixed at .05.

At the subscale level, the results are summarized in Tables 12 and 13 and support the conclusions suggested by the analyses of the total scores. In these tables, the 10 subscale means for Change Scores I and II are summarized. For all 10 subscales of Change Scores I and for the nine of the ten subscales of Change Scores II, means were highest for the EAG group. Results of the analysis of variance tests suggest that two of the Change Scores I differences were statistically significant when the likelihood of a Type I error was set at .05. Furthermore, subscale differences for three Change Scores I and one Change Scores II approached statistical significance under these test conditions. Thus, the analysis of these data again suggest that changes in self perception were slighly higher for members of the EAG than for members of the SAG or EIIG groups. These differences were somewhat more pronounced in favor of the EAG group when Change Scores I (posttest-pretest) served as the dependent variables. These findings were particularly striking since they were suggested not only by the analysis of total scores but by the analyses of 19 of the 20 subscale scores as well.

A final point should be made about the data presented in Tables 12 and 13. As already mentioned in the previous

TABLE 12: GROUP COMPARISONS OF SUBSCALE CHANGE SCORES I, MEANS, STANDARD DEVIATIONS AND *F* - RATIOS

Subscales

1 = Planning
2 = Management
3 = Subject Matter
4 = Teaching Method
5 = Instructional Material
6 = Human Growth
7 = Curriculum Development
8 = Evaluation
9 = Working With People
10 = Personal/Professional Characteristics

Group	n		Subscales 1	2	3	4	5	6	7	8	9	10
SAG	13[2]	$\bar{X}$	.15	.08	-.03	-.08	-.10	-.11	-.11	-.23	.05	.19
		SD	.49	.59	.58	.81	.82	.71	.59	.88	.87	.61
EAG	13	$\bar{X}$	.18	.33	.37	.42	.18	.57	.23	.46	.55	.43
		SD	.60	.46	.58	.66	.63	.65	.60	.46	.50	.43
EIIG	15	$\bar{X}$	.12	.10	-.01	.07	-.21	.09	.03	-.10	-.01	-.17
		SD	.43	.44	.42	.60	.61	.47	.71	.90	.67	.50
		F	.06	1.03	2.43[1]	1.76	1.17	4.25*	.92	2.89[1]	2.71[1]	4.86*

[a] n = 12 for subscale 5

[1] $p < .10$ *$p < .05$

TABLE 13: GROUP COMPARISONS OF SUBSCALE CHANGE SCORES II: MEANS, STANDARD DEVIATIONS AND *F* - RATIOS

Subscales

1 = Planning
2 = Management
3 = Subject Matter
4 = Teaching Method
5 = Instructional Materials
6 = Human Growth
7 = Curriculum Development
8 = Evaluation
9 = Working with People
10 = Personal/Professional Characteristics

Group	*n*		Subscales 1	2	3	4	5	6	7	9	9	10
SAG	13[a]	$\bar{X}$	.06	.11	.03	.02	.25	-.03	-.09	-.04	.11	.08
		SD	.65	.60	.63	.71	.66	.80	.64	.92	.60	.74
EAG	13	$\bar{X}$	.25	.34	.21	.51	.49	.52	.15	.42	.38	.42
		SD	.54	.40	.59	.77	.69	.55	.57	.52	.56	.53
EIIG	15	$\bar{X}$	.04	.06	-.02	.05	-.15	.16	.21	.00	.11	.03
		SD	.47	.55	.71	.72	.70	.66	.57	.89	.62	.54
		F	.57	1.09	.48	1.84	3.12[1]	2.23	.96	1.33	.97	1.63

[a] $n = 11$ for subscale 5

[1] $p < .10$

section with reference to Table 5, all means for the subscale scores were positive for SAG and EAG combined as a result of the strong positive influence of the EAG scores. Thus, whereas all the means for the EAG at the posttest-pretest level (Change Scores I) were high and positive, Six of the 10 subscale means for the SAG were negative. Five subscale means were also negative for the control group. The argument used in the previous section also applies here in comparing the SAG and EAG. Since it has already been shown in the discussion of Table 7 that the feedback differential was overwhelmingly positive for both of these groups, the high positive Change Score means for the EAG suggest that the members of this group were positively influenced by feedback to a greater extent than the members of the SAG, since the EAG changed their self perceptions (as measured by posttest - pretest) to a much greater extent in the direction of the positive differential subscale mean scores.

The clinching argument in support of the claim that the group whose feedback was externally assessed (EAG) changed their self perceptions more in accordance with the feedback received than the group whose feedback was self assessed (SAG) lies in a study of the correlation coefficients summarized in Table 9. In this table the coefficients for both treatment groups are presented when feedback differential is correlated with Change Scores I or II. In 21 out of 22 instances, the coefficients for the EAG were higher than those for the SAG. Furthermore, the correlation coefficients obtained were significantly

different from zero in all 22 cases for the EAG groups as against only 10 cases for the SAG (likelihood of Type I error set at .05). These results point to a somewhat stronger and more consistent relationship between feedback differential and changes in self perceptions for the EAG. This again suggests that members of the EAG were more likely to change their self perceptions in accord with the direction and magnitude of the feedback they received.

In summary, both the comparison of mean change scores and a study of correlations between change scores and feedback differential suggest that changes in self perception were slightly higher for the EAG than for either of the other two groups. The differences were most pronounced when post-test - pretest scores served as the dependent variable.

> Question 3: Is the magnitude of change in self perceptions resulting from assessment feedback a function of empirically established levels of agreement (high, low) between feedback received and original self perceptions?

As this question suggests, the author had originally intended to identify two distinct groups of teachers: those with high feedback differential scores and those with low feedback differential scores. However, to establish these two groups empirically was an insurmountable problem. The difficulty is illustrated by the data displayed in Table 14. The greatest obstacle was in the wide and uneven distribution of scores. They spread from below -.31 to above 1.20, and

TABLE 14 FREQUENCY DISTRIBUTION OF MEANS FOR TOTAL FEEDBACK DIFFERENTIAL SCORES (FEEDBACK - PRETEST)

Group	n	$\bar{X}$ [a]	<-.31	-.30to -.01	.00 to .29	.30 to .59	.60 to .89	.90 to 1.19	1.20<
SAG	13	.55 (.52)	1	1	1	4	1	4	1
EAG	13	.59 (.70)	2	-	2	2	1	4	2
SAG & EAG Combined	26	.57 (.61)	3	1	3	6	2	8	3

were thinly and unevenly strung out among the intervening categories. Also, the vast majority of scores were positive. As a result, it was virtually impossible to generate a reasonable decision rule that would experimentally identify relatively large groups from whom the agreement between feedback received and original self perceptions was in fact either high or low.

Although the categories of agreement could not be established in the way originally required by the question, another possibility existed which seemed to meet its substantive intention. Simply put, the question essentially seeks to find out whether persons who received feedback which differed widely from their self perceptions (higher or lower), tended to change their self perceptions to a greater extent than those who received feedback that was only slightly different (higher or lower) from their self perceptions. In other words, do persons with high feedback differential scores have higher change scores than those with low feedback differential scores? To describe the question in this way is almost a prescription for correlational analysis. And so, this is what was undertaken. The scores for feedback differential were correlated with the change scores using the process already described for question one. These results were presented in Table 9, and have already been discussed in terms of their relevance to the first two research questions.

In the context of the present question, the data in Table 9 show a uniformly positive relationship for both sets of correlations. When feedback differential scores were correlated with Change Scores I (Posttest - Pretest), the correlation coefficients for the SAG ranged from .11 to .73, with only four of the 11 coefficients falling below .35. For the EAG, the coefficients ranged from .74 to .90, which are exceptionally high correlations. When feedback differential scores were correlated with Change Scores II (Delayed Posttest - Pretest), the coefficients for the SAG ranged from .38 to .63; those for the EAG ranged from .53 to .87, which again are exceptionally high. In all, only four of the 44 correlation coefficients that were computed were below .35 and well over half (29) were above .50. Furthermore, all 22 EAG correlation coefficients, and 10 of 22 SAG correlation coefficients were significantly greater than zero, when the likelihood of Type I error was set at .05.

The uniformly positive and generally high correlation coefficients at both the total scale and the individual subscale levels indicate that high feedback differential scores were typically followed by large changes in self perception; low feedback differential scores were followed by relatively small changes in self perception.

It should also be noted that coefficients were all higher when feedback differential scores were correlated with Change Scores II (delayed posttest - pretest). It is therefore tempting to conclude that changes in self

perceptions continued over time to move in the direction of the feedback received. However, subsequent analyses of change in each individual's self perceptions over time (as represented in Table 17) do not seem to support this conjecture.

Question 4: Will the changes in self perceptions revealed in the posttest immediately following the experimental intervention (seven to 10 day interval) differ from corresponding changes in perceptions revealed in a delayed posttest administered approximately six weeks after the intervention?

This question deals with the stability of changes in one's self perceptions over time. The simplest way to deal with the question of whether changes in self perceptions are stable is to calculate the differential between the total change scores (delayed posttest - posttest) for the two treatment groups and for the control group, and to test whether these differences are significantly different from zero. Small and statistically insignificant differences would suggest a high degree of stability.

Table 15 summarizes the results of these analyses. For all groups the means were very small, ranging from -.01 (for the EAG) to -.06 (for the EIIG) and .06 (for the SAG). The mean of the combined treatment groups was .03. Results of the t - tests suggest that the means for combined treatment group and for the control group were not significantly different from zero when the likelihood of Type I error was set at a relatively liberal level of .10 in order to

TABLE 15: CHANGE SCORE DIFFERENTIAL (DELAYED POSTTEST - POSTTEST): MEANS, STANDARD DEVIATIONS AND RESULTS OF *T* - TESTS.

Group	*n*	$\bar{X}$	*SD*	*df*	*t*
SAG	13	.06	.17		
EAG	13	-.01	.18		
SAG & EAG Combined	26	.03	.18	25	.714[a]
EIIG	15	-.06	.36	14	-.624[a]

[a]Not significant at $p < .10$

increase power. Thus, it would appear that the changes in self perceptions were stable, varying only slightly over time.

The stability of self perceptions can be examined in an even more stringent test. Correlations between change scores for each of the 10 subscales, in addition to the total scale, were tested for statistical significance. The results are presented in Table 16. They show very high coefficients for all three groups. With only one exception, the correlations for the two experimental groups were higher than the corresponding coefficients for the control group. Furthermore, all but two of the correlation coefficients were significantly different from zero when alpha was set at .05. High correlations between Change Scores I and II again strongly suggest that changes in self perceptions varied little over time for all groups, with the slight edge in stability favoring the groups that received feedback.

Additional Analyses

The correlational analyses done so far in response to questions 2 and 3 suggest that clear relationships exist between feedback differential scores and changes in self perception at both the total and subscale levels of analysis. In other words, those persons with high feedback differential scores were most likely to alter their self perceptions in accord with the feedback they received.

TABLE 16: STABILITY OF CHANGES IN SELF PERCEPTIONS: CORRELATIONS BETWEEN CHANGE SCORE I (POSTTEST - PRETEST) AND CHANGE SCORE II (DELAYED POSTTEST - PRETEST)

Subscales

1 = Planning
2 = Management
3 = Subject Matter
4 = Teaching Method
5 = Instructional Materials
6 = Human Growth
7 = Curriculum Development
8 = Evaluation
9 = Working With People
10 = Personal/Professional Characteristics

		Subscales										
Group	n	1	2	3	4	5	6	7	8	9	10	Total
SAG	13[a]	.75**	.92**	.81**	.82**	.89**	.95**	.83**	.87**	.87**	.91*	.96**
EAG	13	.72**	.88**	.89**	.95**	.83**	.74**	.51*	.85**	.87**	.82**	.92**
EIIG	15	.54[1]	.64*	.46[1]	.74**	.66*	.75**	.65*	.80**	.84**	.64*	.71**

[a] n = 11 for subscale 5

[1] $p < .10$ * $p < .05$ ** $p < .01$

Having found that this relation held across individuals for the 10 subscales and the total scale, a more demanding question--one not suggested by the four specific research questions--could now be asked. Given the tendency for participants to respond to feedback differential in a significantly predictable manner when data was analyzed at the group level, would a similar pattern of responsiveness hold true for individuals, taken one at a time, across the 10 subscales?

In an attempt to answer this question, the correlations between the 10 subscale feedback differential scores and the 10 subscale change scores were calculated for each individual. The aim was to discover which specific individuals were or were not responsive to the feedback they received.

The types of correlations now being pursued should first be distinguished from those summarized in Table 9 and described earlier in the discussion of question one with reference to Table 8. As explained then, measures of group responsiveness for the total scale and the individual subscales were derived by correlating each column of 26 feedback differential scores with the corresponding column of 26 subscale change scores, one pair at a time for each participant.

To get a measure of individual responsiveness, rows of scores rather than columns were correlated. This meant that for each individual, his or her row of 10 feedback

differential scores were correlated with the individual's corresponding row of 10 subscale change scores. This correlation provides an index of an individual's responsiveness to feedback. Imagine, for example, that an individual's feedback differential score is very high on one subscale (for example, management) and comparatively low on another (for example, planning). If changes in self perception are consistent with feedback received, the differences between posttest and pretest ratings should be larger for the first subscale, management, than for the second, planning. Thus, strong correlations of this type reflect high levels of responsiveness to feedback; low correlations reflect low levels of responsiveness.

In this way, as shown in Table 17, two measures of responsiveness were derived for each of the 26 individuals in the sample:

$\underline{r}_1$ = Correlations between each participant's 10 subscale feedback differential scores with that individual's 10 subscale Change Scores I.

$\underline{r}_2$ = Correlations between each participant's 10 subscale feedback differential scores with that individual's 10 subscale Change Scores II.

As an example of this procedure, consider the first individual in the sample. The pattern of feedback differential scores and subscale change scores were as follows:

	Subscales					
	1	2	3	4	5	10
Feedback Differential	1.19	1.30	1.11	1.57	-.09	1.26
Change Scores I (Post test - Pretest)	.40	.60	.20	.40	.20	.56

The correlation between these two rows of scores was .69. This coefficient is significantly different from zero when the likelihood of a Type I error was set at .05.

The two measures of responsiveness for each individual in the sample are summarized in Table 17. It is important to note that of the 26 participants, 22 had at least one statistically significant responsiveness score. This means that only four participants were not significantly affected by the feedback they received. This is a very strong indication of individual responsiveness to the feedback received.

The influence of feedback differential was about equally strong for both change scores. Eighteen of 26 correlation coefficients were significantly different from zero when feedback differential scores were correlated with Change Scores I (r_1), and 16 of 26 coefficients were significantly different from zero when feedback differential scores were correlated with Change Scores II (r_2). There seems, therefore, to be only a slight difference in

TABLE 17: INDIVIDUAL RESPONSIVENESS: CORRELATIONS BETWEEN THE PETTERN OF FEEDBACK DIFFERENTIAL SCORES (FEEDBACK - PRETEST) AND THE CORRESPONDING PATTERN OF CHANGE SCORES (POSTTEST - PRETEST AND DELAYED POSTTEST - PRETEST) FOR EACH PARTICIPANT IN EXPERIMENT I

r_1 = Correlation Coefficients: Feedback-Pretest/Posttest-Pretest Across 10 Subscales

r_2 = Correlation Coefficients: Feedback-Pretest/Delayed Posttest-Pretest Across 10 Subscales

	SAG												
SUBJECT #	1	2[a]	3	4[b]	5	6	7	8	9	10	11	12	13
r_1	.69**	.68*	.60*	.78**	.60*	.64*	.61*	.51[1]	.33	.80**	.33	.54*	-.11
r_2	.79**	.38	.60*	.78**	-.09	.49[1]	.59*	.06	.41	.70**	.40	.86**	-.11

	EAG												
SUBJECT #	14	15	16	17	18	19	20	21	23	23	24	25	26
r_1	.81**	.37	.73**	.82**	.80**	-.07	.45[1]	.80**	.40	.46[1]	.90**	.43	.23
r_2	.74**	.23	.72**	.92**	.04	.57*	.20	.89**	.71**	.23	.73**	.68*	.45[1]

[a] $n = 9$ for r_2 [b] $n = 9$ for r_1 and r_2 [1] $p < .10$ * $p < .05$ ** $p < .01$

individual responsiveness at these two levels.

Another question suggests itself at this point. Is group membership a factor where individual responsiveness to feedback is concerned? To answer this question, the means for the two sets of correlations ($\underline{r}_1$ and $\underline{r}_2$) were calculated for both treatment groups. The differences between the groups were tested for statistical significance. The results summarized in Table 18 show that where the first set of correlations ($\underline{r}_1$) is concerned, the means for the two groups were almost identical (.54 for the SAG and .55 for the EAG), producing an insignificant difference. The differences between group means were only slightly higher for the second set of correlations ($\underline{r}_2$) and were also insignificant. Where individual responsiveness is concerned, group membership seems to make very little difference.

In summary, the data suggest that individual responsiveness to feedback was quite strong for the majority of participants. Twenty-two individuals significantly changed their self perceptions in line with the feedback they received. Individual responsiveness did not vary as a function of how the feedback was received (SAG or EAG).Furthermore the influence of feedback differential on individual responsiveness did not vary over time: the level of relationship was approximately the same for the delayed posttest as it was for the posttest.

TABLE 18: INDIVIDUAL RESPONSIVENESS: CORRELATIONS BETWEEN FEEDBACK DIFFERENTIAL SCORES (FEEDBACK - PRETEST) AND CHANGE SCORES I OR II (POSTTEST - PRETEST *or* DELAYED POSTTEST - PRETEST) FOR EACH PARTICIPANT IN EXPERIMENT I: MEANS, STANDARD DEVIATIONS AND RESULTS OF ANOVA TESTS

r_1 = Correlation Coefficients: Feedback - Pretest/Posttest - Pretest - Across 10 Subscales

				Analysis of Variance			
Group	n	$\bar{X}$ r_1	SD	Source	df	MS	F
SAG	13	.54	.24	Main Effect	1	.00	.01
				Error	24	.07	
EAG	13	.55	.29				

r_2 = Correlation Coefficients: Feedback - Pretest/ Delayed Posttest - Pretest - Across 10 Subscales

				Analysis of Variance			
Group	n	$\bar{X}$ r_2	SD	Source	df	MS	F
SAG	13	.45	.33	Main Effect	1	.06	.65
				Error	24	.09	
EAG	13	.55	.29				

Summary of Part One

Part one of this chapter attempted to answer four questions about changes teachers make in their self perceptions as a result of the feedback they receive from others. In response to the first question as to whether the feedback received causes any change in the teachers' self perceptions, analysis of the data strongly suggested that the teachers in this sample were clearly influenced by the feedback they received. Despite their training in self assessment, these teachers made significant and predictable changes in their self perceptions following feedback from others. Those who received feedback higher than their own estimation of themselves raised their self perceptions in the direction of the feedback; those who received feedback lower than their own self evaluation, lowered their self perceptions.

The second question asked which of the two forms of feedback (SAG or EAG) will yield the greatest changes in teachers' self perceptions. The analysis of mean change scores and correlation coefficients between change scores and feedback differential consistently suggested that changes in self perception were slightly higher for the teachers in the EAG group (who had the feedback analyzed for them and reported to them by an external evaluator), than for the teachers in the SAG group (who had analyzed and reported on their feedback). The differences between the EAG group and the other two groups were most pronounced when posttest - pretest

scores served as the dependent variables. Also, the changes for the EAG group were strongest in the direction of the feedback received.

The third question asked whether the magnitude of change in self perceptions resulting from assessment feedback was a function of levels of agreement between feedback received and original self perceptions. The uniformly positive and generally high correlation coefficients between feedback differential and change scores resulting from analysis of the data in response to this question, strongly indicate that high feedback differential scores were typically followed by large changes in self perception. Low feedback differential scores were followed by relatively small changes in self perception. Thus, magnitude of change in self perceptions could be said to be a function of levels of high or low agreement between feedback received and original self perceptions.

The fourth question had to do with the stability of changes in self perception which result from feedback. Changes in self perception were found to be very stable, varying only slightly over the six week interval between the posttest and the delayed posttest.

An additional set of analyses to determine each individual's level of responsiveness to feedback was an exactig test of the central question in this study. Individual responsiveness to feedback was found to be strong for the majority of participants. Twenty-two of twenty-six teachers

changed their self perceptions in line with the feedback they received, somewhere during the course of the study. How the feedback was provided (SAG or EAG) did not seem to affect the level of individual responsiveness.

Part Two: Attitudinal Outcomes

Introduction

Following the delayed post test, the 26 participants who had experienced the two term evaluation program (guided self analysis in the first term, experimental treatments in the second) were asked to complete a debriefing form. As described in Chapter three, the purpose of this survey was to find out how these participants felt about the various aspects of the program, and in this way to collect formative data about the processes involved. Specifically, the debriefing survey sought answers to three questions:

1. What elements in the two term evaluation program did the participants find most valuable, or least valuable?
2. How did they feel about participating in the program?
3. What were their perceptions of the feedback they received?

Summaries of responses that relate to these questions serve as the focus of the three sections which follow.

Question 1: What elements in the two term evaluation program did the participants find most valuable or least valuable?

Table 19 presents a summary of responses to questionnaire items that focused on this question. These data suggest that the self assessment done in the first term was felt by participants to be the most beneficial part of the two term process. Where the second term was concerned, feedback from the first part of the questionnaire (TBS) was viewed as most valuable. According to the participants, the least valuable part of the experiment was the process of thinking through and writing up the summary outline at the end of the evaluation. However, even this feature was seen as having limited to moderate value.

Five open ended items at the end of the attitude scale gathered further data on this question. Items 29 and 32 asked participants to describe the features of the program they most liked or valued. Their responses to these two items included the following (the numbers in brackets indicate the number of times the feature was mentioned):

---

1. Self evaluation (14)
2. Peer evaluation (11)
3. Interaction with colleagues (8)
4. Becoming aware of strengths and deficiencies (7)
5. Being given the opportunity to participate in an evaluation program (6)

TABLE 19 MEAN SCORES FOR ITEMS CONCERNED WITH VALUES PARTICIPANTS PLACED ON CHARACTERISTIC PROGRAM ELEMENTS.

4 = high value 2 = limted value

3 = moderate value 1 = little or no value

Item Number	Items	$\bar{X}$
1	The process of deriving the lists of teacher strengths and needs in the first term	3.19
2	The self assessment done in the first term	3.54
3	Feedback from the first section of the questionnaire ("Teacher's Strengths") provided by your professional colleagues in the second term	3.00
4	Feedback from the lists of strengths cited by your colleagues at the end of the first section of the questionnaire	2.85
5	Feedback from the second section of the questionnaire ("Possible Areas for Improvement") provided by your professional colleagues	2.88
6	Feedback from the lists of your possible needs for improvement cited by your colleagues at the end of the second section of the questionnaire	2.92
7	The process of thinking through and writing up the summary/outline at the end of the evaluation	2.54
	Grand Mean = 2.99	

6. Becoming aware of the importance of evaluation (5)
7. Deriving lists through group discussion in the first term (5)
8. Being evaluated by administrators (5)

These responses confirm that self evaluation was viewed as the most valuable feature in the total program. The process of deriving lists and interaction with colleagues which were also highly valued, were among the most characteristic features of the first term when the self evaluation was done. The participants also valued the two types of assessors, peer and administrator, who contributed feedback data during the experimental portion of the program. It is also important to note that several of the participants commented on the value of the evaluation process itself (see 4, 5 and 6 above).

The features of the program that were least valued or liked, as measured by two open-ended items 30 and 33, were as follows (the numbers in brackets indicate the number of times each feature was mentioned.):

---

1. Repetitiousness (15)
2. Too much time required of participants (10)
3. Summary/Outline written at the end (7)
4. Too much time required of assessors (5)
5. Inefficient assessors.(4)

Items 1, 2 and 4 on this list indicate the participants had strong objections to the amount of time and repetition required by the program. They had spent a lot of time in the first term in activities that led to the production of the items from which the Teacher Behavior Survey (TBS) was derived, and then in the second term they were asked to fill out the TBS three times. In addition, the TBS they completed had 114 items in two sections, the second being a negative version of most of the items on the first section. As already mentioned, this second section proved not very helpful in collecting data, and was eventually discarded.

The summary/outline again proved to be unpopular, perhaps partly because it was also time consuming and came at the end of term when the participants were under pressure to hand in their other end-of-term assignments. But also, these summaries were never discussed with nor returned to the members of the externally assessed group (EAG), who had written their summaries after their Interview in which the feedback had been given them. On the other hand, the summary was the basis of the report given by the self assessed group (SAG) in their interview; but even here the protocol of the experiment precluded discussion.

Question 31 on the attitude scale asked the participants to recommend changes that should be made before the program is used with other students. The main responses are listed below:

1. Shorten the process, decrease repetitions (14)
2. Make assessors more efficient (10)
3. Give additional academic credit (6)
4. Condense the TBS (4)
5. Do earlier in the year, not in Spring term (3)

The first and fourth items are consistent with the strong message already received in response to questions 31 and 33. The third item also indicates a degree of frustration participants felt in having to undergo the experimental treatments in the second term in addition to other course requirements. Item 5 also refers to the time required of them and their assessors in what they described as the busiest term of the year.

Where item 2 was concerned, "make assessors more efficient," the participants gave a number of more specific and useful suggestions:

-clearer instructions should be given to assessors so their evaluations could be "more helpful and honest".

-more time should be spent on the in-classroom assessment by the assessors, and a "more nearly equal amount of time" spent by all.

-peer assessment should be done "over a longer period so as to be more valid".

-peer assessors should be "<u>required</u> to observe"

-peer assessors should be chosen from other schools

-peer assessors should be members of the evaluation program

-opportunity should be provided for participants "to go over the assessments with the assessors"

-assessors should not "guess"

In summary, participants seem to have valued most those activities done in the first term--self evaluation, interaction with colleagues, and deriving lists of teacher strengths and deficiencies. Where the second term is concerned, they also felt that they benefited greatly from the feedback they received from the first section of the TBS. They also valued the opportunity to receive feedback from peers and administrators. They felt that they achieved a growing awareness of their abilities, and were grateful for the opportunity to be involved in the evaluation process.

They least appreciated the amount of time and repetition required by their participation, and strongly suggested that the program be condensed. They also did not favor the summary they were asked to write, although it was only mentioned by one person as something to be changed. Finally, a number of suggestions were given to improve the efficiency or effectiveness of the assessors.

Question 2: How did the teachers feel about participating in the program?

Table 20 summarizes responses to questions that focused on participants' attitudes toward the experience. It is

TABLE 20: MEAN SCORES FOR ITEMS INDICATING HOW PARTICIPANTS FELT ABOUT TAKING PART IN THE TWO-TERM PROGRAM

4 = strong agreement
3 = moderate agreement
2 = moderate disagreement
1 = strong disagreement

No. on Scale	Items	$\bar{X}$
8	Because members of the control group had less to do, I initially felt some frustration/displeasure in being assigned to an experimental group	2.81
9	The benefits I received from being a member of an experimental group adequately compensated for the extra work I had to do	2.31
10	The experiment procedure was too long and repetitive	3.35
11	I made considerable gains in my ability to assess my own performance as a result of this experience	2.73
12	I made considerable gains in my ability to assess a colleague's performance as a result of this experience	2.42
13	I am more sensitive to shortcomings in my classroom performance as a result of participating in this project	3.23
14	I am more sensitive to strenghts in my classroom performance as a result of participating in this project	3.08

No. on Scale	Items	$\bar{X}$
15	The benefits teachers might derive from some form of systematic evaluation in the MACT Assessment I course will depend in large part on the interpersonal/communication skills of the instructor who directs the experience	2.88
16	The two-term evaluation procedure in which I participated should become a required part of the MACT Assessment I course	2.50
17	All teachers should undergo some form of professional assessment at least once every five years	3.65
18	I felt comfortable and relaxed during the classroom sessions in which I was being observed by my professional colleagues	3.13
19	The process of being observed and assessed by professional colleagues was a valuable experience in and of itself	2.85
20	I would have taken the feedback I received more seriously if the assessors had spent more time in the classroom observing my performance	2.88
21	The assessors took this job seriously and did it conscientiously	2.92
22	There are few if any significant disagreements between my own ratings and those provided by my professional colleagues	2.96
23	The fact that the assessors tended to be very generous in their ratings reduced the value of the feedback experience	2.35
24	I have a higher regard for my own abilities as a teacher as a result of this experience	2.92
	Grand Mean = 2.88	

important to note that at the end of the two terms of evaluation--a program which many must have felt to be arduous and sometimes irksome--the participants gave their strongest agreement rating ($\overline{X}$ = 3.65) to the following item:

> 17. All teachers should undergo some form of professional assessment at least once every five years.

This response confirms the finding that participants had a high positive attitude toward the main experience provided by the program. This positive feeling was also conveyed by their responses to items 13 and 14. Participants felt that they had benefited significantly from the program, for, as they reported, they had become more sensitive to both the strengths and shortcomings in their classroom performance. On balance, they also felt higher regard for their abilities as a result of their experience in the program (See item 24).

Item 9 confirms the teachers' feelings that additional graduate credit should have been given for the second term of the course. Participants also reported some disagreement with the statement that the benefits from being a member of the experimental group adequately compensated for the extra work they had to do.

Responses to items 18, 21, 22 and 23 reveal a high positive attitude toward the processes and results of the assessment they experienced. Participants felt comfortable and relaxed while being observed (item 18). Although they recognized shortcomings in their assessors, as already reported, they agreed that these individuals in general took

their job seriously and did it conscientiously (item 21). They saw few significant disagreements between their own ratings and those of their assessors (item 22), and denied that the generosity of their assessors' ratings reduced the value of the feedback experience (item 23). As already mentioned in discussing item 24, they reported having a higher regard for their abilities as teachers as a result of this experience.

Briefly, then, participants seemed to have had favorable attitudes towards the program and felt they had benefited from it. However, they also felt that in its present form it was too long and repetitive.

Question 3: What were the participants' perceptions of the feedback they received?

Items 18, 21, 22 and 23 of the attitude scale provide a partial answer to this question. The responses to these items have already been summarized in the section above, where it was reported that participants responded favorably to the processes and results of the feedback assessment procedures.

Some additional items on the scale were specifically concerned with participants' perceptions of the level and effects of the feedback they received. Responses to items 25 and 28 are summarized in Table 21. Only one participant felt that the assessors' overall rating was lower than his or her own (see item 25). Ten participants felt that the assessors ratings generally coincided with their own.

TABLE 21: NUMBER OF RESPONSES FOR ITEMS RELATED TO PERCEPTIONS OF FEEDBACK RECEIVED

No. on Scale	Items	Number
25	Overall, the assessors tended to rate me...I rate myself	
	1. higher than	15
	2. at about the same level as	10
	3. lower than	1
26.	The extent to which the assessors agreed among themselves in rating my performance of specific skills was	
	1. very high	6
	2. high	15
	3. low	5
	4. very low	0
27.	I was most sensitive to these differences between my ratings and those of my assessors when they rated me	
	1. higher than I rated myself	6
	2. lower than I rated myself	19
28.	The most significant outcome of the feedback I received was	
	1. identification of some weaknesses I did not know I had	9
	2. identification of some strengths I did not know I had	2
	3. confirmation of the perceptions I had of my abilities prior to this experience	15

Fifteen felt the assessors' ratings were higher, but, as reported earlier (see item 23), this tendency of the assessors to rate generously did not reduce the value of the feedback experience for most of them.

Another positive attitude to the feedback data received is suggested by the responses to item 26. Only six participants felt that the extent to which the assessors agreed among themselves was low.

With regard to item 27, the great majority of participants (19 out of 25) said they were most sensitive to these differences between their ratings and those of their assessors when the assessors rated them lower than they rated themselves. It bears repeating here that, as reported in Part One of this chapter, the differences between feedback and participants' initial self perceptions were almost always positive, that is, feedback scores were most frequently higher than initial self perception scores. The second point is that although participants claim greater sensitivity to negative differentials, they nevertheless changed their perceptions significantly in the direction of the positive feedback they received.

Finally, the data collected in response to item 28 show that most participants (15 of 26) felt that the principal outcome of the feedback they received was to confirm their original perceptions of their abilities. Slightly over one third (9 of 26) felt that the primary outcome was an identification of some weaknesses they did not know they

had; only two of the participants believed the main result was an identification of previously unrecognized strengths. Nevertheless, the teachers felt the feedback helped them to raise their estimations of their abilities. This sense of positive reinforcement might also explain why the participant were generally positive toward the experience at the end of the demanding two term evaluation program.

Summary of Part Two

The principal outcome of the two-term experience in the eyes of most participants was a confirmation of the perceptions they already held regarding their own abilities. Although they judged that the level of feedback they received from colleagues and supervisors was generally higher than their initial self perceptions, they nevertheless, felt that this did not reduce its usefulness. In keeping with these perceptions, the participants generally expressed strong, positive feelings toward the program, even though it lasted two terms, and was seen by some as being too long and repetitious.

Where specific aspects of the program were concerned the opportunity to engage in self analysis in the first term was seen as the most valuable feature of the two term program. Peer groups interactions, which were a part of their analysis, were also seen as particularly valuable.

The first half of the TBS questionnaire was felt to have provided profitable feedback. The second half, however,

was regarded as less useful. As suggested elsewhere, the second section may have contributed to some feelings about repetitiousness and the amount of time required by the whole process. Another aspect of the program had a consistently negative rating. This was the final summary/outline which came at the end of a very busy term.

The assessors who provided the feedback, were judged to have done their job conscientiously and well. This sensitive and positive attitude to the assessors resulted in participants giving some very useful suggestions for improving the assessors' performance.

CHAPTER V: SUMMARY AND IMPLICATIONS

Purpose and Design of the Investigation

The main purpose of this study was to investigate whether external feedback causes teachers to change self perceptions of their professional behavior in those areas where feedback has been given. It was decided to test this question in very exacting circumstances. The investigation was therefore done with teachers who had recently undergone an extensive program of self analysis, and in this process had used and helped define the categories of teacher behavior which were later used to provide them with feedback.

Two practical considerations motivated this decision. First, from an experimental point of view, it seemed within the realm of common sense to predict that feedback would make some impact on the self perceptions of teachers who had not been asked to reflect on their own abilities in any systematic way. The literature gave strong indications that this would be so. On the other hand, it was not so certain that teachers would be affected by external feedback once they had firmly established their self perceptions through a systematic process of self assessment. The second practical consideration was concerned with the fact that self assessment is becoming increasingly important in teacher education,

especially in inservice teacher education where teachers are increasingly being involved in determining their own needs and programs. In this context, self assessment by itself is being seen by some as providing teachers with adequate insights on which to base decisions about the course of their professional development.

Putting these two practical considerations together, the question of interest became: after engaging in an extensive process of self assessment, does external feedback change one's self perceptions to any significant extent?

Another question explored was concerned with which of two forms of feedback--externally analyzed and reported, or self analyzed and reported--caused the greatest changes in teachers' self perceptions. Also considered were questions about (a) the effect of the magnitude of differences between self perceptions and the perceptions of others, and (b) the stability of changes in self perceptions.

The design of the experiment took into account the need for all participants to have experienced a preparatory stage involving a rigorous term of self assessment. Next, since the two distinct modes of giving feedback were to be tested, three groups of participants were selected, one for each of the two feedback treatments, and one to act as a control group for the experiment. The groups were designated SAG (the self assessed group), EAG (the externally assessed group) and EIIG (the control group). The members of the SAG were given their assessment data and had to analyze them and

make an oral report. The members of the EAG had the data analyzed for them and were given the results in an evaluator's report. The assessment data took the form of performance ratings supplied by three assessors who were given precise written instructions on when and how to complete the feedback survey. The assessors were chosen by each participant. In order to provide a wider range of data sources, two were peers, and the third had to be an administrator. The control group (EIIG) received no feedback at all. Neither this group nor its members were randomly selected; hence, the EIIG group was strictly a quasi-control group, chosen because its members had all undergone a one term self evaluation program similar to that experienced by the two treatment groups.

The main instrument used in the investigation was the Teacher Behavior Survey (TBS), which the participants in the treatment groups had helped to develop. Responses to the TBS served as pretest, posttest, and delayed posttest for all the participants, and it was also the instrument used by the three assessors to rate teacher performance.

A second instrument, an attitude scale, was given to the members of the experimental group after they had completed the experiment. The scale was intended to find out how participants felt about the two term program and its constituent parts: what they valued and did not value,and what they felt ought to be altered. It was an attempt to collect formative data about this sort of evaluation program.

Feedback and Self Perceptions

A. Major Findings

A number of important conclusions were suggested by the analysis of the data focusing on the relation between feedback received and alterations in self perceptions. The three most significant findings were:

1. Teachers' self perceptions change as a result of feedback, even after they have previously rigorously assessed their own behavior.
2. Changes in self perceptions are a clear function of feedback differential. Thus, teachers change their self perceptions in the direction of the difference between their original self perceptions and the feedback they receive. Furthermore, the magnitude of changes in self perception are a function of the size of the difference between initial self perceptions and feedback received.
3. When feedback is externally analyzed and reported, it appears to be a slightly stronger motivator of change in self perception than when it is self analyzed and reported to an external evaluator.

B. Discussion of Findings

1. Change in Self Perception as a Result of Feedback

It is a very significant finding that teachers, even after they have analyzed themselves in specific areas and established fairly precise perceptions, are nonetheless

malleable to feedback that is given in those same areas. However, once these self perceptions are changed as a result of the feedback received, they remain stable over time when no additional treatments are applied.

How can this malleability be explained? At least three different explanations might be proposed. Common sense suggests one practical explanation. Teachers spend their professional lives locked away in their private classrooms, and professional politeness and the social structure of the school tend to reinforce this isolation. Principals do their required evaluations reluctantly and with little in-class observation. Their feedback presents not only no challenge but little useful information to teachers. As a result of this situation, teachers crave helpful informative, professional feedback on their teaching performance. They desire the opportunity to get this type of feedback especially when it is not done for the sole purposes of retention, tenure and promotion. And they respond to it when they are given it.

Interactionists would explain the malleability of self perceptions, where external feedback is concerned, in terms of the "looking glass self". They claim that an individual's self concept is very much determined by the way others view that person, and the extent to which their view is communicated to him or her. If the "others" are "significant" to the person, their views are more influential. Also, the clearer and more detailed the perception of these views are, the more they are likely to be influential. In this investigation, since the assessors were chosen by the participants,

it is very likely that they were "significant" to them. Next, the feedback was detailed and clear due to the fact that the participants had used the same categories as those in the TBS for their earlier self assessment, and the TBS itself made for a clear and detailed presentation of the data. In these circumstances, even after self assessment, the teachers were likely to be influenced by external feedback.

In addition to sharing these views, Combs (1975) was very dubious about the efficacy of self assessment. He was quoted earlier as saying that "the objective analysis of self" is "vastly overrated as a device for personality change". Thus, Combs would explain the responsiveness to others' opinions even after detailed and prolonged self assessment as due to the inefficacy of self assessment. For him the explanation would be simple: the self assessment made little enduring impact. As already indicated in the review of the literature, however, there is experimental evidence supporting the effectiveness of the use of self assessment as a valuable tool in behavior change.

So, in brief, it does seem that the notion of the "looking-glass self", in addition to the fact of their habitual isolation, is the best explanation for the malleability of teachers' self perceptions even after they have undergone long and detailed self assessment.

2. Change as a Function of Feedback Differential

When teachers are given feedback that matches closely their own perceptions of their performance, there is little subsequent change in their self perceptions. This one would expect. On the other hand, when feedback is widely different from initial self perceptions, these self perceptions change in the direction of the feedback, and to a greater or less extent depending on the extent of the difference between feedback and initial self perceptions. In other words, the magnitude of change in self perceptions appears to be a clear function of feedback differential.

This behavior is consistent with Festinger's theory and experiments cited in the literature review. Elsewhere there was a concern expressed that too great a difference between a teacher's perceptions and those of the assessors might cause the teacher to become less responsive to the feedback or more impervious to change. This does not seem to have happened to any noticeable degree in the present investigation. Results of this experiment suggest that teachers do respond to cognitive dissonance caused by differences between their own perceptions and those of their chosen assessors by changing their self perceptions. The results also indicate that the greater the differences, the more the teachers seem to change.

3. Feedback Data: Externally Analyzed vs. Self Analyzed

The results of this investigation suggest that data feedback which is externally analyzed and reported to teachers (EAG) influences their self perceptions to a somewhat greater extent than the same sort of data when it is analyzed and reported on by the teachers themselves (SAG). As pointed out earlier, both forms of data feedback have great influence on teacher self perceptions, even after teachers have engaged in protracted self analysis. But when the question of a comparison is raised as to which form of feedback (EAG or SAG) is the more potent, then the results show only slight differences between the two, with the differences consistently favoring the external mode.

Studies discussed in the literature review in Chapter Two show that both modes of data feedback have been used with success. Those studies concerned with supervisor or peer evaluation support the value of externally analyzed feedback. The literature on "educational linking agents" also indicates that a "support system" seems necessary to ensure the efficacy of feedback. On the other hand, research described in connection with microteaching and computer assisted programs give some indications that self analyzed data feedback is a potent force for change. With regard to microteaching, studies reported by Cooper and Allen (1971) support the practice of teachers confronting their videotaped performance with or without supervisor or peer assistance. Computer

printouts, self analyzed by teachers, also effect changes in teacher behavior, as reported by Pohlmann (1976), Semmel and Olson (1977) and Hail (1978). In addition, recent teacher education practice, promoted by many writers including Curwin and Fuhrmann (1975) and Rubin (1978), stress the importance of teacher self determination and the discovery of one's teaching self through processes of self analysis.

The present investigation tested these two alternative modes of feedback in a one-on-one experiment. Since previous evidence was that both forms of feedback do bring about change in teacher behavior, big differences between the two were not expected. It should be remembered, too, that the EAG treatment was restricted to one function of external evaluation, that of analysis and transmission of data. All the other functions of an external evaluator commonly exercised in supervision, which are described as "helping," "support," or "interaction," were deliberately suppressed in the experiment so as to compare more exactly the two different modes of analysis and reporting. These experimental conditions constituted a very conservative test of the effects of external evaluation. Nevertheless, external evaluation, so narrowly exercised in this experiment, proved to be more potent than self analysis. This seems to present a powerful argument in favor of the more normal uses of external evaluation found in teacher field-based experience, as compared to self analysis. For, if an external evaluator limited to one function can influence teacher self perceptions

consistently, even if only to a slightly greater extent than self analysis in the tests conducted in this investigation, then it seems likely that an external evaluator, freely interacting, helping and supporting the teacher, will be an even more powerful influence.

There are at least two possible explanations for the greater efficacy of external evaluation. First, there are arguments and research findings of interactionist psychologists reported earlier in Chapter Two, which support the theory of the "looking-glass self". In this theory, the opinions of others--particularly those of "significant others"--communicated to an individual, present him with a mirror of his self which determines the development of that person's self concept to a very large extent. One's self concept, though it becomes fully developed, is never finally formed, and is therefore open to change. The most powerful agent for changing one's self perceptions is new information about one's performance from other people who one holds in esteem. In this experiment the significant others were, first of all, the assessors chosen by the teachers themselves to be the sources of their feedback. For those teachers who were given the EAG treatment, another significant person reinforced the process of other-reflection. As co-instructor for the two-term program, the external evaluator can be expected to have achieved a "significant" status with these teachers. The effect of the feedback was, therefore, magnified by his intervention.

Another explanation for the greater strength of external evaluation as revealed in this conservative test may be found in the deference customarily shown to people in authority. This seems the most straightforward explanation. Deference shown to persons in authority, including one's university instructor, is common human behavior. A set of opinions about one's professional behavior, analyzed and reported by one's instructor is likely to appear more authoritative and, therefore, persuasive than similar opinions which an individual weighs and sifts in the privacy and warmth of one's living room or study.

The power of opinions when they are communicated by a significant other, to whom, in addition, one shows professional deference, would seem to go a long way in explaining the greater potency of external evaluation as compared with self assessment in giving feedback.

Additudinal Outcomes

General

The main benefit participants said they derived from the two graduate courses that focused on the evaluation process was the confirmation of their perceptions of their strengths and weaknesses. In fact, the feedback they received was higher than their initial self perceptions, and most participants raised the level of their self perceptions as a result. Nevertheless, the feeling they were left with was one of support for their professional self perceptions.

It is also significant that participants expressed strong positive attitudes towards the program, even though it was long and demanding, continuing over two terms. They placed high value on the program in general, and were able to identify specific aspects from which they felt they benefited as well.

Taking the program stage by stage, the outstanding characteristics identified by participants were as follows:

Term 1: Guided Self Analysis

The highest value of all was placed on the self analysis done in the first term. This process was characterized by a great deal of peer group interaction as exemplified by the derivation of lists of teacher strengths and deficiencies and review of one another's needs assessments.

Term 2: The Experiment

1. The Teacher Behavior Survey (TBS)

Participants rated highly the feedback they received from the first half of the TBS --the section concerned with the assessment of the teacher's strengths. Teachers preferred this mode of receiving feedback, and it also proved an adequate basis for needs-based program planning.

On the other hand, the second section of the TBS was not found to be particularly useful. Assessors communicated their antipathy to this section through messages or written notes. But, most of all, they restricted themselves to two

rating responses. They either disagreed or disagreed strongly with the items in this section, which all suggested specific areas in which the teacher might need to improve.

In brief, the first half of the survey was an adequate instrument, framed as it was in positive terms requiring the assessment of relative levels of a teacher's strengths. The second half of the survey, which was framed in more negative terms requiring an assessment of what some perceived as the teacher's "weaknesses", was neither fruitful nor well received.

2. The Assessors

The participants appreciated their assessors and the feedback they gave. They thought them to be conscientious, felt comfortable when observed by them, and valued their feedback. They made a number of specific suggestions for improving their assessors' performance. Basically, they felt that in the circumstances, the assessors did a good job, but should be given training and supervision in future programs.

3. Feedback

Participants perceived that the level of feedback was higher than that of their own initial self perceptions. However, they did not think that this reduced its effectiveness. Generally, they felt that the effect of the feedback was to confirm perceptions they already had of their abilities. In fact, as already mentioned, feedback raised the levels of their

own assessment of their abilities. Nevertheless, they also reported that they responded most to these differences between their ratings and their assessors' when the assessors rated them lower than they rated themselves.

Implications of the Investigation

A. Professional Development of Teachers

This investigation has a number of implications related to teacher education, and more specifically, to the professional development of teachers. The recent trend is increasingly to involve teachers in assessing their needs and in planning and even developing their own professional development programs. For teachers to be in the best position to do this they need to base their assessments on the best possible foundations. Objective self analysis is one factor, but by itself it may not be enough. Certainly this investigation strongly suggests that teachers' insights into their behavior are greatly influenced by the perceptions and opinions of others, even after these insights have been originally established by detailed and prolonged self assessment. Granted the validity of external data, this potent external influence, so strongly rooted in the interactionist psychological tradition, ought to be utilized in building the soundest foundations for professional development. Teachers, as evidenced by the participants in this study, are attracted to and feel rewarded by self assessment and peer interaction. But they

also respond very significantly to additional feedback from external sources, and the self perceptions so formed appear to be stable over time.

Where the question of modes of feedback is concerned, this investigation suggests that it is slightly more effective to transmit the feedback through an external agent. It should be noted that the person acting as the external agent in this investigation performed the barest functions usually associated with an evaluator in that interaction was reduced to a stark minimum. It is therefore possible that the difference between external and self analysis of feedback data may be even more pronounced under "natural" conditions. However, teacher educators will want to consider these options very carefully, since the evidence is not conclusive on either side of the question, and the influence of age and experience were not successfully excluded.

In the professional development of teachers, various types of dissonance do seem to be potent motivators of teacher change. This investigation suggests that dissonance can be induced between self and other perceptions by using a teacher developed questionnaire, such as the first part of the Teacher Behavior Survey (TBS) developed for this study. Such an instrument with a positive focus, allows for the collection and communication of clear and detailed perceptions of assessors. Feedback in this form is essential for the development of dissonance.

Another factor affecting dissonance is the "attractiveness" of the assessing group. It does appear from this study

that if teachers choose their assessors they are most likely to choose persons who for them are "significant others". As a result, they will not easily derogate these assessors' opinions even when they differ widely from their own self perceptions. Another implication from this investigation having to do with dissonance is that it confirmed Festinger's idea that teachers are motivated to change their perceptions in proportion to the magnitude of the dissonance they experience. The limits of teachers' tolerance of dissonance were not tested. In the absence of this evidence, sensitivity and prudence would counsel against excess.

It also appears that assessors chosen by teachers will tend to rate them more generously than the teachers rate themselves. Even so, it does appear that teachers want and value this feedback as a strengthening experience. Teachers also value the opportunity both to engage in self evaluation and to receive external feedback. They seem, too, to welcome the opportunity to be observed, certainly in an environment where they do not feel threatened, but perceive it as part of an information gathering process contributing solely to their professional development.

B. Suggestions for Further Research

At the end of the literature review in Chapter Two, a number of outstanding research issues were defined. Some of these issues were specifically addressed by the primary research questions framed for this investigation. A number

of issues, nevertheless, remain outstanding. These include:

(1) So many of the feedback differential scores were positive that the evidence concerning directionality was one-sided. Positive differential induces positive change, but the effect of negative differential needs further investigation.

(2) The question of magnitude of dissonance needs to be further looked into. What are the effects of "maximum possible dissonance" on the perceptions and behavior of a much wider range and larger number of teachers?

(3) The attributes of significant others operating as assessors in the context of professional development needs further definition. Also, granted that teachers given the choice might well choose significant others as their assessors, what, nevertheless are the criteria on which these choices are based? From the literature, expert assessors also seem very effective as motivators of change, and may be said to operate as "salient" others. What are the differential characteristics of these two groups and do they influence teachers in different ways? And are expert assessors more reliable and valid in their perceptions of teachers' behavior?

(4) The nature and effect of different forms of feedback needs further investigation. Optimum levels

of feedback specificity, for example, different categories of teacher behavior and other modes of transmission suggest themselves as fruitful areas for further research. In particular, since this study chose to limit the role of the external evaluator in transmitting the feedback, the question of the effect of greater interaction, which is usually associated with a supervisory or peer assessor, could serve as the subject of important additional research in this area.

(5) The first section of the TBS survey used in this study seems to have been an effective instrument. What would be the effect of using an instrument not developed by the participants, which might then not reflect their value system?

(6) Some variables that were dealt with as potential confounding variables in this investigation could usefully be explored as treatment variables in other experiments. What are the effects of sex, other geographical locations, and level of school taught on any of the research questions asked in this study? Also, the effects of age and years in teaching were not totally accounted for in this investigation. They are certainly variables that require further investigation.

(7) The stability of self perceptions seems well documented in the literature and by the results of

this investigation. The obvious area for further investigation is the effect of longer periods of time.

(8) The final question, however, that needs urgent exploration is the effect of changed self perceptions on specific and general aspects of teacher behavior. It was suggested earlier that in the absence of a valid and supported model of teacher behavior, feedback differential affects perception but not behavior. Is this true only in the short term? If not, then efforts to change self perceptions, the assurances of Combs (1965) and the other perceptual psychologists notwithstanding would seem somewhat sterile or at least, limited exercises in improving teacher effectiveness. This is an area of research that needs extensive exploration.

(9) Because the question addressed and the findings obtained seemed to have dramatic implications for teacher education, there is a clear need to replicate this study. Those who believe that inservice teacher education ought to be built on a firm foundation of teacher insights into their own professional strengths and needs should be particularly sensitive to these results. The findings strongly suggest that Combs (1965) is right in arguing that self assessment by itself provides

an adequate base for personality change. If this is so, then the conclusion is inescapable that teachers should systematically process feedback from others as an essential part of those experiences that constitute their professional development.

Traditional methods of inservice training have requently included some process of external assessment. However the impact of external feedback on teachers may have been underrated by practitioners in this field. The findings of this study suggest that external feedback exerts a stronger and more dramatic influence on teachers than many who have worked in this area may have expected.

The implications of these findings are far reaching not only for inservice teacher education but also for coordinators of student teaching and for in-school evaluators of teachers. They, too, hold a most potent instrument in their hands: the effect of their feedback on teachers and student teachers may also be more profound than has been imagined.

Because of the critical nature of the implications of this study for teacher education, it is important that it be replicated. Other groups of teachers with different demographic features, and larger numbers of teachers ought to be used in order to ensure greater generalizability.

APPENDICES

APPENDIX A

CALENDAR OF EVENTS

APPENDIX A

CALENDAR OF EVENTS

The steps in conducting the experimental investigation may be briefly summarized as follows:

I. Assigning Participations

April 16, 1979 : Selection and assignment of participants

II. Administering Pre-, Post-, and Delayed Posttest

(These took place during three periods of approximately one week each, six weeks apart).

April 16-25, 1979 : Pretest given to first year City A group on April 16, to the first year City B group on April 18, and to the second year City A group (EIIG) on April 24.

May 28-June 6, 1979 : Posttest completed by participants in Experiment I between May 28 and June 6. Participants in Experiment II (EIIG) completed theirs on June 5.

July 10-19, 1979	:	Delayed Posttest and Attitude Scale were handed out in the last class of the term during the week of June 3, with instructions that they be kept until July 10 and completed before July 17.

III. Experimental Treatments

April 16, 1979	:	Questionnaires were given to the participants in Experiment I to be distributed to their administrators and peers. Two and a half weeks was the time suggested for the data collection.
May 10, 1979	:	All questionnaires had been returned by the assessors.
May 14, 1979	:	Analysis of data began for City A group
May 16, 1979	:	Analysis of data began for City B group
May 24 and 29	:	Interviews for City A group
May 30 and 31	:	Interviews for City B group

APPENDIX B

TEACHER BEHAVIOR SURVEY (TBS)

APPENDIX B

TEACHER BEHAVIOR SURVEY (TBS)

I. Questionnaire on Teacher's Strengths

PLEASE READ THE INSTRUCTIONS AND SUGGESTIONS IN THE ACCOMPANYING LETTER BEFORE FILLING OUT THIS QUESTIONNAIRE

Teacher's name_____

Grade Level or Subject Taught_____

Please assess the teacher named above in terms of each of the following statements, and assign a mark from 5 to 1 in each case.

5 = Exceptional (This teacher is in the top 5% of teachers at this level or in this subject)
4 = Strong (This teacher is in the top 15% of teachers at this level or in this subject).
3 = Above Average
2 = Below Average
1 = I have had no opportunity to observe or know this.

	This Teacher:	5	4	3	2	1
A. Planning	1. Makes plans that are flexible					
	2. Plans to meet the needs of individual students					
	3. Aims to involve all students in meaningful work					
	4. Can integrate different materials, methods and interests in a plan					
	5. Makes good short term and long range plans					

Teacher's Name: _____

	This teacher:	5	4	3	2	1
B. Organization and Management	6. Firmly establishes the rules of the classroom so that students know expectations and consequences					
	7. Recognizes needs and problems early, and deals with them before they develop into major matters					
	8. Attends to the special needs of indidivual students					
	9. Is consistent in establishing the day to day patterns of the class					
	10. Is flexible and adaptable					
		_____				
		_____				
C. Classroom Climate and Control	11. Helps students to build positive self-concepts					
	12. Is fair and consistent in the classroom					
	13. Communicates clearly and effectively with students					
	14. Is flexible and adjusts to the needs of individual students, and to the differences between groups and classes					
	15. Is pleasant and relaxed in dealing with students					
	16. Organizes classroom space in an attractive and stimulating manner, for example: bulletin boards, interest centers, materials available to students, etc.					
		_____				
		_____				

Teacher's Name: _____

	This Teacher:	5	4	3	2	1
D. Command of Subject Matter	17. Adapts knowledge to the ability level of students					
	18. Keeps up with current ideas and information in subject (s)					
	19. Has personal interest in subject matter and communicates this to students					
	20. Has inclusive knowledge of the subject matter including the content of last year's and next year's syllabuses					
	21. Can admit ignorance and is willing to learn along with the class					
		_____				
		_____				
E. Teaching Method	22. Uses a variety of methods and materials appropriate to different students, subjects and situations					
	23. Includes all children, regardless of ability					
	24. Communicates clearly and effectively with students in both speech and writing					
	25. Communicates a sense of the purpose of what is being taught					
	26. Keeps up with new developments, and is willing from time to time to give a careful try to a new approach					
		_____				
		_____				

Teacher's Name: _____

This Teacher:	5	4	3	2	1
F. Use of Audio-Visual and other materials					
27. Uses audio-visual and other suitable materials to enrich the curriculum, help students understand, or excite their interest					
28. Uses a variety of these materials					
29. Uses audio-visual materials to engage more fully the students' sense in the learning process					
30. Has clear goals and sticks to them in using audio-visual and other instructional aids and materials					
31. Prepares students before using audio-visual or other instructional material in class, and follows up on its use					
	_____				
	_____				
G. Understanding Human Growth & Development					
32. Understands and cares about human feelings					
33. Helps students to develop self-confidence					
34. Has expectations that neither exceed nor fail to challenge the students' levels of development					
35. Is familiar with the unique cognitive and social needs of students at the age he/she is dealing with and adjusts instruction accordingly					
36. Is aware of the home environment of students and its effects on them					
	_____				

Teacher's Name: _____

	This Teacher:	5	4	3	2	1
H. Understanding Curriculum & Curriculum Development	37. Knows in detail the objectives for the grade level above and below for each subject (or grade) he/she teaches					
	38. Modifies curriculum guidelines to suit the needs of particular class (or classes)					
	39. Uses pretesting to find where the students are					
	40. Sets clear, reasonable objectives for each grade level (or subject)					
	41. Knows federal, state and local guidelines, and establishes priorities congruent with their goals and objectives					
		_____				
		_____				
I. Evaluation Procedure	42. Uses a variety of evaluation procedures appropriately and is not restricted, for example, to written tests					
	43. Consistently evaluates own teaching					
	44. Interprets results of student evaluation for the benefit of students, for example, by putting comments on papers and offering suggestions for improvement					
	45. Makes sure students understand the evaluation procedures and what is expected of them					

Teacher's Name: _____

This Teacher:	5	4	3	2	1
46. Sees and uses tests as teaching instruments					
	_____				
	_____				
J. Working With People					
47. Is open and honest in his/her attitudes and dealings					
48. Avoids negative preconceived ideas about students or parents					
49. Is a good listener.					
50. Respects others--students, staff, and parents					
51. Is sensitive and tactful					
	_____				
	_____				
K. Personal & Professional Characteristics					
52. Has a positive attitude toward teaching					
53. Relates to students, parents, community members and other teachers in a professional manner					
54. Respects professional confidentiality in dealing with both students and parents					
55. Is sincere and enthusiastic					
56. Has a good sense of humor					
57. Is open to new ideas					
58. Seeks and/or accepts constructive criticism of his/her professional ideas, attitudes and hehavior					
	_____				
	_____				

ON COMPLETING THIS QUESTIONNAIRE:

List in rank order what you consider the 5 most significant strengths of this teacher. (Please be as specific as possible, but you need not restrict yourself to the categories or statements used in this questionnaire).

1.

2.

3.

4.

5.

II. Questionnaire on Teacher's Possible Areas for Improvement

PLEASE FILL OUT THIS QUESTIONNAIRE AFTER COMPLETING THE QUESTIONNAIRE ON TEACHER'S STRENGTHS

Teacher's Name_____

Grade Level or Subject taught_____

Assess the teacher named above in terms of each of the following statements, and assign a mark from 5 to 1 in each case.

5 = strongly agree
4 = agree
3 = disagree
2 = strongly disagree
1 = I have had no opportunity to observe or know this

	This teacher needs to:	5	4	3	2	1
A. Planning	1. Plan more adequately to meet the needs of the students					
	2. Have clearer objectives and priorities					
	3. Plan less rigidly and be more flexible					
	4. Plan more carefully the use of his/her own and the class's time					
	5. Keep an adequate lesson plan book for substitute teacher to use					
		_____				
		_____				
B. Organization & Management	6. Plan for and establish the sort of classroom routine that makes students feel more secure					
	7. Be more consistent in discipline and daily work patterns					
	8. Use fewer seat-work assignments and other types of busy work					

Teacher's Name: _____

This teacher needs to:	5	4	3	2	1
9. Adapt materials and/or methods to the needs of individual students to a greater extent					
10. Have a more poised and relaxed manner of interacting with students					
	_____				
	_____				
C. Classroom & Climate Control					
11. Build up the dignity and self-worth of students to a greater degree					
12. Recognize and understand the needs of individual students					
13. Be more consistent and strive to find the happy medium					
14. Relate to all students in a more positive manner					
15. Organize the classroom space in a more attractive and stimulating manner, for example: bulletin boards, interest centers, materials more readily available for students, etc.					
	_____				
	_____				
D. Command of Subject Matter					
16. Be better able to present subject matter at students' level					
17. Have a more adequate background in subject matter					

Teacher's Name: _____

This teacher needs to:	5	4	3	2	1
18. Be willing to cover all aspects of the subject including areas of the curriculum where he/she might feel insecure					
19. Be more curious about new methods and developments in his/her area					
20. More adequately understand and model the ideas or skills being taught					
E. Teaching Method					
21. Be careful in practice not to expect the same from every child					
22. Be less rigid and inflexible, and use a wider range of methods and materials					
23. Make learning more enjoyable					
24. Be better able to integrate a variety of techniques into his/her own style					
25. Focus more on teaching the students and not exclusively on teaching the books (or materials)					
F. Use of Audio-Visual and other materials					
26. Be careful not to overuse audio-visual materials, especially films					

Teacher's Name: _____

This teacher needs to:	5	4	3	2	1
27. More carefully correlate the use of audio-visual, or other aids and materials, with the aims of the lesson					
28. Prepare the class beforehand for the use of audio-visual or other instructional material, and follow up on its use					
29. Use more audio-visual or other aids and materials, and/or a wider range of these materials					
30. Be more aware of the potential of audio-visual materials and other aids in the teaching process					
	_____				
	_____				
G. Understanding Human Growth & Development					
31. Be more aware of growth and development as important factors in the learning process					
32. Be more aware of individual needs and differences and treat children more as individuals					
33. Adjust his/her teaching techniques to the level of the students					
34. Be able to cope more adequately with special learning or behavior problems in class					
35. Have more empathy for the emotional changes students are undergoing					
	_____				
	_____				

Teacher's Name: _____

	This teacher needs to:	5	4	3	2	1
H. Understanding of Curriculum and Curriculm Development	36. Be more aware of and pay more attention to the requirements of what is being taught above and below the subject (or grade) level he/she teaches					
	37. Pay more attention to curriculum objectives and/or more adequately integrade them with his/her own perceptions of subject needs					
	38. Have a better knowledge of relevant guidelines and objectives					
	39. Set objectives or define priorities more clearly					
	40. More consistently evaluate the curriculum and make modifications in it					
		_____				
		_____				
I. Evaluation Procedure	41. Make more adequate provisions to evaluate own teaching, and/or be more responsive to evaluation feedback					
	42. Know better how to interpret evaluation results and how to give helpful feedback					
	43. Evaluate students more frequently and completely					
	44. Be less affected by whether he/she likes a student or not					

Teacher's Name: _____

	This teacher needs to:	5	4	3	2	1
	45. Be more familiar with a wider variety of evaluation procedures and instruments					
		_____				
		_____				
J. Working with People	46. Show more tact or concern for the feelings of others					
	47. Be less negative and apathetic					
	48. Have a more open mind and be able to admit mistakes					
	49. Avoid showing favoritism and/or unfairly labelling students					
	50. Be more reliable					
		_____				
		_____				
K. Personal and Professional Characteristics	51. Be more careful not to reveal confidential information without proper need					
	52. Show less concern for the paycheck and/or other benefits					
	53. Be less narrow minded and lacking in concern for others					
	54. Have less difficulty working with other teachers and/or parents					
	55. Avoid coming down to the level of students, for example, by arguing with them or trying to be a "pal"					

Teacher's Name: _____

This teacher needs to:	5	4	3	2	1
56. Comply more completely with rules set by administration and/or staff					
	_____				
	_____				

ON COMPLETING THIS QUESTIONNARIE:

List in rank order 5 specific areas in which you think this teacher might most need to improve. (Please be as specific as possible, but you need not restrict yourself to the categories or statements used in this questionnaire).

1.

2.

3.

4.

5.

THANK YOU VERY MUCH FOR YOUR HELP!

APPENDIX C

DEBRIEFING FOR PARTICIPANTS IN THE TWO-TERM MACT PROFESSIONAL EVALUATION

APPENDIX C

DEBRIEFING FOR PARTICIPANTS IN THE TWO-TERM MACT PROFESSIONAL EVALUATION

N.B: When completed please return separately in the stamped self-addressed envelope provided!!

1-7: To what extent have each of the following activities in the two-term evaluation program influenced your relative strengths and shortcomings as a classroom teacher?

1 = strong influence
2 = moderate influence
3 = limited influence
4 = little or no influence

Please circle the appropriate number in the box on the right

1. The process of deriving the list of teacher strengths and needs in the first term	1	2	3	4
2. The self assessment done in the first term	1	2	3	4
3. Feedback from the first section of the questionnaire ("Teacher's Strengths") provided by your professional colleagues in the second term	1	2	3	4
4. Feedback from the lists of strengths cited by your colleagues at the end of the first section of the questionnaire	1	2	3	4
5. Feedback from the second section of the questionnaire ("Possible Areas for Improvement") provided by your professional colleagues	1	2	3	4
6. Feedback from the lists of your possible needs for improvement cited by your colleagues at the end of the second section of the questionnaire	1	2	3	4

7. The process of thinking through and writing up the summary/outline at the end of the evaluation	1	2	3	4

8-24 Please indicate your level of agreement with each of the following statements by circling the appropriate number in the box to the right.

1 = strongly agree
2 = agree
3 = disagree
4 = strongly disagree

8. Because members of the control group had less to do, I initially felt some frustration/displeasure in being assigned to an experimental group	1	2	3	4
9. The benefits I received from being a member of an experimental group adequately compensated for the extra work I had to do	1	2	3	4
10. The experimental procedure was too long and repetitive	1	2	3	4
11. I made considerable gains in my ability to assess my own performance as a result of this experience	1	2	3	4
12. I made considerable gains in my ability to assess a collegaue's performance as a result of this experience	1	2	3	4
13. I am more sensitive to shortcomings in my classroom performance as a result of participating in this project	1	2	3	4
14. I am more sensitive to strengths in my classroom performance as a result of participating in this project	1	2	3	4
15. The benefits teachers might derive from some form of systematic evaluation in the MACT Assessment I course will depend in large part on the interpersonal/communication skills of the instructor who directs the experience	1	2	3	4

16. The two-term evaluation procedure in which I participated should become a required part of the MACT Assessment I course	1	2	3	4
17. All teachers should undergo some form of professional assessment at least once every five years	1	2	3	4
18. I felt comfortable and relaxed during the classroom sessions in which I was being observed by my professional colleagues	1	2	3	4
19. The process of being observed and assessed by professional colleagues was a valuable experience in and of itself	1	2	3	4
20. I would have taken the feedback I received more seriously if the assessors had spent more time in the classroom observing my performance	1	2	3	4
21. The assessors took their job seriously and did it conscientiously	1	2	3	4
22. There were few if any significant disagreements between my own ratings and those provided by my professional colleagues	1	2	3	4
23. The fact that the assessors tended to be very generous in their ratings reduced the value of the feedback experience	1	2	3	4
24. I have a higher regard for my own abilities as a teacher as a result of this experience	1	2	3	4

25-28 Please select the alternative that best describes your response to each of the following statements. <u>In each case circle the number to the left of your answer</u>

25. Overall the assessors tended to rate me...I rated myself

1. higher than
2. at about the same level as
3. lower than

26. The extent to which the assessors agreed among themselves in rating my performance of specific skills was

    1. very high
    2. high
    3. low
    4. very low

27. I was most sensitive to those differences between my ratings and those of my assessors when they rated me

    1. higher than I rated myself
    2. lower than I rated myself

28. The most significant outcome of the feedback I received was

    1. identification of some weaknesses I did not know I had
    2. identification of some strengths I did not know I had
    3. confirmation of the perceptions I had of my abilities prior to this experience

---

29-33 Please complete the following statements. PLEASE TYPE OR PRINT YOUR COMMENTS IF POSSIBLE.

29. The most valuable aspect of the two-term evaluation program was

30. The least valuable aspect of the evaluation program was

31. I would recommend that the two-term evaluation procedure should be required of all MACT candidates if, and only if, the following change(s) is (are) made:
    a)
    b)
    c)
    d)

32. I like

33. I did not like

THANK YOU VERY MUCH!!!

APPENDIX D

LETTER TO ASSESSORS WITH INSTRUCTIONS FOR COMPLETING THE TEACHER BEHAVIOR SURVEY

C/o Dr. John Cragun
Battle Creek Teacher Education Center
Miller-Stone Building
77 Capital Avenue, N.E.
Battle Creek, MI 49016

April 16, 1979

TO: _____

School:_____ For (Teacher's name)_____

Thank you for agreeing to fill out the two attached questionnaires. We greatly appreciate your cooperation.

You are participating in an exercise that has two purposes. First of all, you will be helping the teacher named above to complete the needs assessment of professional abilities which is part of the Master of Arts in Classroom Teaching Program which this teacher is currently involved in. The teachers on this program rely on the results of their needs assessment to help them to choose courses and to select their own special professional development projects. Secondly, you will be helping a research project which is being done by this researcher at the Battle Creek Area Teacher Education Center, School of Education, Michigan State University. The aim of this project is to compare the relative efficacy of two kinds of feedback--self and external--in assisting teachers to evaluate their needs for the purposes of their own professional development.

Please follow the following procedure in filling out the questionnaires:

1. Read the questionnaires carefully to see the sort of information required. Then, <u>during the next two and a half weeks</u>, which is the recommended length for the period of this assessment:

2. Visit this teacher and observe him/her teaching for <u>about one hour</u>, or, for high school teachers, <u>one full class period</u>. This observation may be done over two or three visits, but should total the time specified above.

3. Look at teacher's lesson plan book, samples of tests given, samples of students' work, teacher's records and reports, teacher prepared materials, and other teacher and student products which can assist you.

4. Have informal talks with the teacher

5. When you have completed your period of assessment, complete the questionnaires as quickly as possible.

Please note that in answering the questionnaires you should draw an overall knowledge of this teacher over the period you have known him/her as a teacher.

Please also note that the results of these questionnaires are to be kept annonymous. Only the researcher will know who has filled out each set of questionnaires. To ensure this very important condition, please enclose the completed questionnaires in the stamped addressed envelope provided and return to me. I will then transcribe your responses onto another questionnaire form which will then be seen by the subject of the questionnaires. You are one of three evaluators for this teacher whose questionnaires will be treated in this way.

Finally, may I ask you to be certain to complete and return these questionnaires by:_____

THANK YOU VERY MUCH INDEED!!!

Yours Sincerely,

Donald G. Wilson

APPENDIX E

DESCRIPTION OF ASSESSOR

APPENDIX E

DESCRIPTION OF ASSESSOR

Please fill out and return this sheet with the completed questionnaires.

1. Name_____

2. School_____

3. Position_____

4. Have you done a formal evaluation of a colleague before?_____If yes, about how many times?_____

5. How long have you known the teacher being evaluated since he/she has been a teacher?_____

6. Do you teach/work in the same building?_____If yes, for about how long?_____

7. Have you had the opportunity to observe this teacher in the classroom before?_____If yes, please briefly describe the nature and extent of the previous observation:_____

_____

_____

8. How many hours/class periods did you spend in the last three weeks observing this teacher prior to completing these questionnaires?_____

9. How many times have you had discussions with this teacher for the purpose of the present exercise?_____ About how long did these discussions last?_____

10. What teacher and student products (e.g. lesson plans, tests, records, students' work, instructional material, etc.) did you look at in the course of the present evaluation? (Please list these)_____

_____

11. How long have you been in teaching?_____

12. How long have you been in your present position?_____

THANK YOU VERY MUCH!!!

APPENDIX F

SUMMARY OF DATA ON ASSESSORS

SUMMARY OF DATA ON ASSESSORS-ADMINISTRATOR

Teacher/Student Products: LP= Lesson Plans, IM=Instructional Materials, T=Tests, SW=Student's Work, R=Teacher's Records, CR=Classroom BB=Bulletin Boards

Teacher Evaluated	Assessor's Position	Times Assessed Before	Years Teacher Known	Years in Same Building	Observed Teacher Before	Periods spent in this observation	Discussions Held	Time Spent in Discussion (Minutes)	Teacher/ Student Products Seen	Years in Teaching	Years in Present Position
1	Principal	200+	5	5	5	-	1	5	LP,IM,R	14	5
2	Principal	Many	2	2	7	2	3	20	all as listed	21	13
3	Intern Consultant	10+	3	2	Many	1	1	20	LP,Tests, SW,BB	17	3
4	Department Chairman	Many	6	6	Annually	6	2	20	LP,T,R	15	7
5	Principal	Many	8	7	Many	0	0	0	all as listed	26	15
6	Principal	Many	1	1	3	1	2	15	LP,T,SW	13½	11½
7	Principal	Many	4	4	4+	1	1	10	all as listed	22	12
8	Principal	Many	2	2	Few	3	2	Brief	BB,SW,LP,R,	24	6
9	Asst. Prin.	Many	2½	2½	Few	1	1	10	R,IM,SW	19	6
10	Principal	400	1½	1½	Many	-	5	10-25	LP,T,SW,IM	22	7
11	Asst.Prin.	Many	6	6	3	2	1	15	All as listed	30	7
12	Principal	400	3½	3	Many	4	6	10-15	All as listed	22	7
13	Principal	Many	Many	5	0	½	2	5 -10	LP,T,R,SW,IM	24	6+

SUMMARY OF DATA ON ASSESSORS-ADMINISTRATOR (Page 2)
Teacher/Student Products: LP=Lesson Plans,IM=Instructional Materials, T=Tests, SW=Student's Work, R=Teacher's Records,CR=Classroom,BB=Bulletin Boards

Teacher Evaluated	Assessor's Position	Times Assessed Before	Years Teacher Known	Years in Same Building	Observed Teacher Before	Periods Spent in this Observation	Discussions Held	Time Spent in Discussion (Minutes)	Teacher/ Student Project Seen	Years in Teaching	Years in Present Position
14	Prin	200+	5	5	5	-	1	5	LP, IM,R	14	5
15	Prin	Many	1	1	Many	1	2	30	LP, T,IM	10	5
16	Prin	Many	8	8	16	1	1	30	LP, SW, R	21	13
17	Prin	Many	13	13	Many	2	2	15	All as listed	21	13
18	Prin	Many	1	1	3	4	2	10	LP, SW	13	5
19	Prin	Many	1	1	0	1	0	-	All as listed	15	7
20	Prin	200	1	1	5	3	3	45	LP,T,R,IM	7	1
21	Prin	Many	1	1	Many	2	1	3 -4	LP,T,R,IM	24	6+
22	Asst.Prin	6+	7+	7	8	1	1	15	All as listed	30	8
23	Prin	12+	10	10	Many	2+	1	20	LP	19	11
24	Prin	Many	15	15	Many	-	1	brief	-	21	11
25	Prin	Many	7	7	Many	2	1	12	All as listed	16	13
26	Prin	Many	15	5	Many	2	2	5	All as listed	24	6

SUMMARY OF DATA ON ASSESSORS - PEER TEACHER[1]

Teacher/Student Products: LP=Lesson Plans, IM=Instructional Materials, T=Tests, SW=Student's Work, R=Teacher's Reocrds, CR=Classroom, BB=Bulletin Boards

Teacher Evaluated	Assessor's Position	Times Assessed Before	Years Teacher Known	Years in Same Building	Observed Teacher Before	Periods spent in this observation	Discussions held	Time spent in discussion (Minutes)	Teacher/ Student Products Seen	Years in Teaching	Yeras in Present Position
1	Counselor	-	3	3	yes	2	2	5 -10	-	4	3
2	Teacher	6	2	2	yes	2	4	30	LP,IM,SW,BB	15	14
3	Teacher	-	2	2	yes	3	3	20	LP, SW,IM	23	16
4	Coach	2	4	4	yes	3	1	5 -10	T	6	4
5	Teacher	-	8+	8	yes	4	4	10 -20	SW,IM	8	8
6	Teacher	-	1	2/3	yes	1	4	15	LP,T	6	4
7	Teacher	-	3	3	No	1	1	-	-	5	5
8	Teacher	4	2	2	No	-	-	-	LP,T,R,	25	13
9	Dept.Head	8	3	2	Yes	1	2	10	LP,SW,IM	11	11
10	Teacher	-	1	1	No	1+	4	10	LP,R,SW,IM	6	1
11	Dept.Head	-	8	8	Yes	0	3	90	LP,R,	9½	1
12	Teacher	-	1½	1½	Yes	3	3	10 -15	LP,IM,SW,R,BB	1½	1½
13	Teacher	5	14	14	Yes	3	2	40	T,R,SW,games	19	16
14	Coordinator of Special Serv.	8	13	No	Yes	1	1	30	R,Texts, IM	22	5
15	Teacher	-	1	1	Yes	1½	1	30	LP,IM	3	3

SUMMARY OF DATA ON ASSESSORS -PEER TEACHER[1] Page 2

Teacher/Student Products: LP=Lesson Plans, IM=Instructional Materials, T=Tests, SW=Student's Work, R=Teacher's Records, CR=Classroom, BB=Bulletin Boards

Teacher Evaluated	Assessor's Position	Times Assessed Before	Years Teacher Known	Years in Same Building	Observed Teacher before	Periods spent in this observation	Discussions held	Time spent in discussion (Minutes)	Teacher/ Student Products Seen	Years in Teaching	Years in Present Position
16	Teacher	6	8	8	Yes	2	4	25	All as listed	15	14
17	Teacher	-	15	8	Yes	2	3	20	All as listed	8	8
18	Reading Con. Consultant	-	3/4	3/4	Yes	2	3	10 -20	Students' Reading + LP,T,BB,R,SW	7	4
19	Teacher	5	1	1	Yes	-	1	5	LP,T,R,C,R,	30	12
20	Teacher	-	2	2	No	4	2	20	LP, SW,IM	12	4
21	Teacher	2	2	1½	Yes	1	1	15	LP,T,IM	17	14
22	Dept.Head	6+	9	7	Yes	-	1	90	all as listed	22	5
23	Teacher	2	8	8	Yes	Many	1	10	All as listed	8	4
24	Teacher	-	16	14	Yes	Many	Many	60	LP,T,IM,A.V. mat.	19	14
25	Teacher	-	9	9	Yes	3	1	30	LP,SW	9	9
26	Teacher	3	7	7	Yes	1	6	30+	LP,R,SW,ditto sheets, art mat.	28	10

SUMMARY OF DATA ON ASSESSORS - PEER TEACHERS (2)

Teacher/Student Products: LP=Lesson Plans, IM=Instructional Materials, T=Tests, SW=Student's Work, R=Teacher's Records, CR=Classroom, BB=Bulletin Boards

Teacher Evaluated	Assessor's Position	Times Assessed Before	Years Teacher Known	Years in Same Building	Observed Teacher Before	Periods Spent in this observation	Discussions held	Time spent in discussion (Minutes)	Teacher/Student Products Seen	Years in Teaching	Years in Present Position
1	Teacher & Coordinator	-	7	6	Yes	Few	1	5	IM,T,SW,LP	8	3
2	Teacher	-	2	2	Yes	2	3	20	All as listed	8	8
3	Reading Con.	-	2	-	No	1	3	15	LP,SW,IM	9	3
4	P.E.Teacher & Football Coach	-	1	1	Yes	Few	3	15	LP,T,R,	20	1
5	Teacher	-	7	7	Yes	4	1	10	LP,T,R,SW,IM	8	7
6	Teacher	-	1	1	Yes	2	8	15	LP,SW	3	3
7	Teacher	3	3	3	Yes	1+	3	15 -20	LP,T,SW,	10	7
8	Teacher	12	2	2	Yes	2	2	30	LP,SW,IM	24	14
9	Teacher	-	2	2	Yes	3/4	3	3 -8	LP,T,IM,VTR	13	12
10	Teacher	-	1½	1½	Yes	Many	3	15	LP,IM	2½	2
11	Teacher	70	7	7	Yes	-	-	-	R,LP,SW,CR	16	15
12	Teacher	-	1	1	No	1+	4	10	All as listed	6	1
13	Teacher	-	14	14	Yes	1½	2	15	All as listed	17	14
14	Teacher	3	19	16	Yes	1	2	30	SW	19	19

SUMMARY OF DATA ON ASSESSORS - PEER TEACHERS (2) (page2)

Teacher/Student Products: LP=Lesson Plans, IM=Instructional Materials, T=Tests, SW=Student's Work, R=Teacher's Records, CR=Classroom, BB=Bulletin Boards

Teacher Evaluated	Assessor's Position	Times Assessed Before	Years Teacher Known	Years in Same Building	Observed Teacher before	Periods Spent in this observation	Discussions Held	Time spent in Discussion (Minutes)	Teacher/ Student Products Seen	Years in Teaching	Years in Present Position
15	Teacher	-	3/4	3/4	Yes	2½	2	10	LP,SW,BB,IM	15½	1
16	Teacher	-	2	2	No	2	3	20	All As listed	2	2
17	Teacher	-	2	2	No	2	3	20	All As Listed	2	2
18	Teacher	-	1	1	No	-	3	20	R, LP,BB,T,SW	3	3
19	Teacher	-	2/3	2/3	Yes	1	1	5	LP,T,IM	6	6
20	Teacher	-	2	2	Yes	3	3 10	-15	LP,T,IM,SW	12	6
21	Teacher	4	1½	1½	Yes	2	2	30	T,R,SW, Game, newspaper	28	14
22	Teacher	-	9	9	Yes	0	3	A lot	LP,T,SW,Tests, Plays, speeches	9	9
23	Teacher	-		1	No	6	1 15	-20	T,SW,R	10	1
24	Dept.Head	100+	17	17	10+	1	3 5	-30	T,LP,R,	25	13
25	Teacher	4	3	1	Yes	3	4	40	LP,T,SW	6	3
26	Teacher	12	7	7	Yes	2	2	30	LP,SW,R,	24	14

APPENDIX G

INSTRUCTIONS FOR SELF ASSESSED GROUP (SAG)

APPENDIX G

INSTRUCTIONS FOR SELF ASSESSED GROUP (SAG)

I am going to give you two sets of things as a result of the evaluations which have been carried out. The first is a set of transcribed copies of the questionnaires which were filled out by the three evaluators you selected. The second is a copy of the lists of strengths and suggestions for your improvement made up by your evaluators.

The aim is for you to get the most meaningful insights you can from these sources of information regarding yourself as a teacher. Look closely at the data. Sift them, compare them, analyze them to find out what they can say to you that will help you in your own personal and professional development.

There is no special method for you to use in your analysis. What you should do is make the most careful and meaningful analysis of these data, keeping in mind that what you want from them is help in making your decisions about your future development as a teacher.

When you have completed your analysis, you are being asked to make a written summary, and an oral report in an interview to be arranged. Write a summary/outline of how you went about processing the data and what you learned from doing

this analysis--both what you learned about yourself from the point of view of your professional development, and also what you learned about evaluation and feedback as a process, within this context. Do not spent time making up a statistical table or doing a statistical analysis, since you are not concerned with making a quantitative or comparative study at this time.

This summary/outline need not be typewritten. But please make two copies, one for you to keep. It should be at least 2 typewritten pages (3 handwritten) pages long, and not more than ten typed pages.

Your summary outline should adequately cover the following points which you should note down:

(1) How did you go about processing the data?

(2) Give a summary of your findings, that is, what were the main messages about yourself that you got from the data?

(3) From the point of view of your professional development--from the point of view of your future as a teacher--what did you find that should be helpful to you?

(4) What was there about this evaluation and feedback <u>process</u>, that you found helpful, not helpful or needing improvement?

At the interview (which you are being asked to sign up for on the sheet being passed around), use your written summary/outline to make an oral report of between twenty minutes and half an hour.

APPENDIX H

RESEARCHER'S SCHEDULE FOR REPORTING TO GROUP II (EAG)

APPENDIX H

RESEARCHER'S SCHEDULE FOR REPORTING TO GROUP II (EAG)

(Give candidate transcribed lists of strengths and needs for improvement provided by assessors. Allow time for them to be read. Give candidates blank questionnaire to make notes on)

I. You have read the lists of your strengths and needs for improvement as suggested by your colleagues. Let us now fill out the picture by looking at the data from the questionnaires your colleagues sent us.

I am going to begin by giving you an overview of what they had to say on the first section of the questionnaire which dealt with your strengths.

I have worked out the mean score you got for each category. You will remember that the categories were Planning, Organization, and Management, and so on.

I am going to tell you which 3 (or 4) categories you scored highest in, which 4 (or 5) you got your middle scores in, and which 3 (or 4) you got your lowest scores in.

*Your top strengths are:

Category----(name) You got an average of_____, which means you were seen by your colleagues as being_____in_____.

(Repeat for other top strengths)

*Your middle strengths are:

Category----(name) You got an average of____, which means you were seen by your colleagues on the average as_____ in_____.

(Repeat for other middle strengths).

*Your bottom strengths are:

Category----(name). You got an average of_____, which means you were seen by your colleagues on the average as_____ in_____

(Repeat for other bottom strenghts)

So, in summary, your top strengths are_____,_____, and _____. Your middle strengths are_____, _____, _____ and _____. And you are seen to be least strong in_____, _____, and _____.

II. Now, I'm going to go through each of these 3 sets of categories, one category at a time, and give you your individual

ratings for each item in the category. I'll begin with your top categories, and deal with the items in these categories first; then move on to the middle categories; and then to your least strong categories. Is that clear? (Repeat, using examples and referring to the blank questionnaire if necessary).

First, I'm going to tell you which items in a category at least 2 raters gave you a 5 for, which means they thought you were exceptional, in the top 5%, in these aspects of the category.

Then I'll tell you which items in the same category at least 2 of the raters gave you a 4, which means they thought you were strong, in the top 15%, in these aspects of the category.

For each item, I will always tell you what each rater gave you, if they do not all agree.

And I will always tell you specifically when you have been given a 3 (or 2, or 1), since relatively speaking they are likely to be your lowest marks. And you might wish to think what each particular 3 (or 2, or 1) means in assessing your own needs for professional development.

Let me give you an example of the procedure.

Look at the first category on your blank questionnaire.

(Give examples of procedure using this category, and one or two others if necessary. Make sure procedure seems to be understood before continuing.)

O.K.? Let's begin.

Your top strengths are_____, _____ and_____.

In category _____ (name):

*2 (or 3) raters gave you a 5 (thought you were exceptional) in item _____. (Read item). (The other raters gave you a_____.

(Repeat for each item where 2 or 3 raters gave a 5).

In this same category:

*2 (or 3) raters gave you a 4 (thought you were strong) in item_____. (Read item) The other raters gave you a _____(Repeat for each item where 2 or 3 raters gave a 4).

Also:

*1 (2, or 3) raters gave you a 3 (2, or 1) in item _____ (Read item). The other rater(s) gave you a_____(and a _____).

(Repeat for other items of this sort).

Since that (those) 3 (or 2, or 1) (s) is (are) your lowest mark(s), you might want to think what it (each one of them) means in assessing your needs for professional development.

(Repeat procedure for middle and bottom strengths).

III. So, to sum up:

You top strenghts are in_____, _____ and _____.

Your middle strengths are in _____,_____,_____, _____, and _____.

And you seem least strong in _____,_____, and _____.

You might also wish to keep in mind the 3 (s) (2s and 1s) you got in item_____ (read it), item_____(read it), etc. in assessing your particular needs for professional development.

IV. Now let us turn to the second section of the questionnaire which has to do with possible areas for improvement.

In this section the intention was for your raters to suggest some specific areas where you might need some improvement.

Let me identify some of their suggestions.

*In category_____, (name), 1 (2 or 3) rater(s) suggested that your might need some improvement in item _____(read). 1 (2 or 3) rater(s) gave you a_____,(and the others gave you a_____(and a_____).

(Repeat for categories and items of this sort).

So, in summary, you should consider whether you have some needs for improvement in the following areas:

*Category_____(name), item_____(read item).

(Repeat as necessary).

OR

So, in summary, your raters did not make any specific suggestions for your improvement in this section of the questionnaire.

APPENDIX I

DESCRIPTION OF PARTICIPANTS

APPENDIX I

DESCRIPTION OF PARTICIPANTS

Self Assessed Group (SAG)	Geographical Location	Sex	Level of School Taught	Age	Years Taught
1	City A	F	Post Elem.	40+	15-19
2	City A	F	Elem.	20-24	1-4
3	City A	F	Elem.	20-24	1-4
4	City A	M	Post. Elem.	25-29	5-9
5	City A	F	Elem.	35-39	10-14
6	City A	F	Elem.	25-29	1-4
7	City A	F	Elem.	25-29	1-4
8	City B	F	Elem.	25-29	1-4
9	City B	F	Post Elem.	20-24	1-4
10	City B	F	Elem.	20-24	1-4
11	City B	M	Post Elem.	30-34	5-9
12	City B	F	Elem.	25-29	1-4
13	City B	F	Elem.	35-39	10-14
Externally Assessed Group (EAG)					
14	City A	M	Post Elem	40+	20-24
15	City A	F	Elem.	25-29	1-4
16	City A	F	Elem.	30-34	5-9
17	City A	F	Elem.	40+	15-19
18	City A	F	Elem.	20-24	1-4
19	City A	F	Post Elem	20-24	1-4
20	City A	F	Elem	35-39	5-9
21	City B	F	Elem.	20-24	1-4
22	City B	F	Post Elem	30-34	5-9
23	City B	F	Elem	40+	20-24

Self Assessed Group (SAG)	Geographical Location	Sex	Level of School Taught	Age	Years Taught
24	City B	M	Post Elem	35-39	15-19
25	City B	F	Elem.	40+	10-14
26	City B	F	Elem.	40+	5-9

EXPERIMENT II

Experimental II Group (EIIG)	Geographical Location	Sex	Level of School Taught	Age	Years Taught
27	City A	F	Elem.	20-24	1-4
28	City A	F	Elem.	25-29	5-9
29	City A	F	Elem.	40+	20-24
30	City A	F	Elem.	40+	20-24
31	City A	F	Elem.	20-24	1-4
32	City A	F	Elem.	30-34	10-14
33	City A	F	Elem.	35-39	1-4
34	City A	F	Elem.	35-39	5-9
35	City A	M	Post Elem	40+	15-19
36	City A	F	Elem.	25-29	1-4
37	City A	F	Elem.	25-29	1-4
38	City A	M	Post Elem.	40+	20-24
39	City A	F	Elem.	40+	10-14
40	City A	F	Elem.	35-39	1-4
41	City A	M	Elem.	35-39	10-14

APPENDIX J

RELIABILITY ANALYSIS OF TOTAL SCALE

APPENDIX J

RELIABILITY ANALYSIS OF TOTAL SCALE

Item	Corrected Item-Total Correlation	Alpha if Item Deleted	Item	Corrected Item-Total Correlation	Alpha if Item Deleted
Q1	.61	.97	Q31	.68	.97
Q2	.48	.97	Q32	.72	.97
Q3	.70	.97	Q33	.94	.97
Q4	.73	.97	Q34	.28	.97
Q5	.37	.97	Q35	.62	.97
Q6	.32	.97	Q36	.62	.97
Q7	.47	.97	Q37	.41	.97
Q8	.42	.97	Q38	.66	.97
Q9	.30	.97	Q39	.48	.97
Q10	.59	.97	Q40	.49	.97
Q11	.66	.97	Q41	.36	.97
Q12	.70	.97	Q42	.65	.97
Q13	.44	.97	Q43	.71	.97
Q14	.64	.97	Q44	.56	.97
Q15	.68	.97	Q45	.55	.97
Q16	.43	.97	Q46	.59	.97
Q17	.49	.97	Q47	.62	.97
Q18	.71	.97	Q48	.63	.97
Q19	.69	.97	Q49	.72	.97
Q20	.44	.97	Q50	.74	.97
Q21	.71	.97	Q51	.74	.97
Q22	.60	.97	Q52	.74	.97
Q23	.60	.97	Q53	.74	.97
Q24	.57	.97	Q54	.70	.97
Q25	.63	.97	Q55	.63	.97
Q26	.78	.97	Q56	.60	.97
Q27	.58	.97	Q57	.73	.97
Q28	.66	.97	Q58	.64	.97
Q29	.62	.97			
Q30	.62	.97			

REFERENCES

Abramson, P., When teachers evaluate each other. Scholastic Magazine, September, 1972, pp. 26-28.

Allport, G., Pattern and growth in personality. New York: Holt, Rinehart and Winston, 1961

Amidon, E.J. and Flanders, N.A. The effects of direct and indirect teacher influence on dependent-prone subjects in learning Geometry. Journal of Educational Psychology, 1961, 52, 286-291.

Amidon, E.J., and Giamatteo, M.C. The behavior of superior teachers. The Elementary School Journal, 1965, 65 283-285.

Amidon, E.J., and Hough,J. (Eds.) Interaction analysis: Theory, research and application. Reading, Mass: Addison Wesley, 1967.

Amidon, E.J. and Hunter, E. Improving teaching: The analysis of classroom verbal interaction. New York: Holt, Rinehart and Winston, 1966

Amidon, E.J., Kies, K.M. and Palisi, A.T. Group supervision: A technique for improving teaching behavior. The National Elementary School Principal, 1966, 45 (5) 54-58.

Argyris, C. Interaction theory and method: A behavioral science view, Reading, Mass: Addison Wesley, 1970.

Argyris, C. and Schon, D. Theory in practice: Increasing professional effectiveness. San Francisco: Jossey Bass, 1974.

Blumberg, A. Supervisors and teachers: A private cold war. Berkeley, Calif. McCutchan, 1974.

Blumenthal, P. A faculty self-study group approach to improving instruction quality. Paper presented at the Annual Meeting of the American Psychological Association, San Francisco, California, August 26-30, 1977. (ERIC, ED 147715).

Bradley, B.T., Cragun, J., DeMarte, P., and Scrivens, R. Inservice Education through a field based Master of Arts Program Paper presented at the ATE Annual Convention, Washington, D.C.: February 16-20, 1980.

Brehm, J.W. and Cohen, A.R. Explorations in cognitive dissonance. New York: John Wiley and Sons, 1962.

Brooks, E.D. Effect of alternative techniques for modifying teacher behavior. Unpublished doctoral dissertation. Stanford University, 1967.

Brooks, E.D. Teacher self appraisal: A professional approach to evaluation. The Arizona Administrator. 1967, 7, 7-8.

Brophy, J.E. Using observation to improve your teaching. (Occassional Paper, No. 21), Michigan State University, The Institute for Research on Teaching, 1979.

Bryant, G. and Haack, F. Appraisal: Peer centered and administrator centered. Education Leadership May 1977, 34, 608-612.

Campbell, D.T. and Stanley, J.C. Experimental and quasi-experimental designs for research on teaching in Gage, N.L. (Ed), Handbook on Research on Teaching Chicago: Rand McNally, 1963.

Chissom, B.S. and Morris, J.E. Development of an automated system for evaluation of student teacher effectiveness (Final report), Washington, D.C. : National Institute of Education, 1976 (ERIC, ED 117 156).

Claus,K. The effects of cueing during modeling procedures on the learning of a teaching skill. Stanford University: Stanford Center for Research and Development in Teaching, 1969.

Cogan, M.L. Rationale for clinical supervision. Journal of Research and Development in Education, 1976, 9, (2), 3-19.

Combs, A.W. The professional education of teachers. Boston: Allyn and Bacon, 1965.

Combs, A.W., Richards, A.C., and Richards, R. Perceptual psychology: A humanistic approach to the study of persons. New York: Harper and Row, 1976.

Cooley, C.H. Human nature and the social order. (First published in 1902 by Scribners) New York: Schocken Books 1964.

Cooper, J.M. and Allen, D.W. Microteaching: History and present status. Microteaching: Selected Papers ATE Research Bulletin, 9, September, 1971 (ATE & ERIC, SP 004 978).

Cragun, J.J. and Wilson, D.G. Developing professional autonomy through individualized needs assessment. East Lansing, Michigan: Michigan State University., 1980.

Curwin, R.L. and Fuhrmann, B.S. Discovering your teaching self: Humanistic approach to effective teaching. Englewood Cliffs, N.J: Prentice Hall, 1975.

DeMarte, P.J., Kelly, M.K., and Freeman, D.J. A report on the Master of Arts in Classroom Teaching (MACT) program. East Lansing, Michigan: Michigan State University, 1979. (Adapted by Bradley et al., 1980).

Doyle, W. and McNally, J. Effect of intent-action discrepancy and student performance feedback on teacher behavior change. Journal of Educational Psychology 1974, 66, 5, 750-755.

Emrick, J.A., and Peterson, S.M. A synthesis of findings across five recent studies in educational dissemination and change. (Rev. ed.) San Francisco: Far West Laboratory for Educational Research and Development, June 1978.

Fages, S.F. The effect of videotaped feedback and dissonance on the change in four dimensions of self perception of student teachers. Unpublished doctoral thesis, University of Miami, 1978.

Festinger, L. A theory of cognitive dissonance. Evanston, Ill.: Row, Peterson & Co., 1957.

Flanders, N.A., Teacher influence, pupil attitudes, and achievement. Washington, D.C.: Government Printing Office, 1965. (Cooperative Research Monograph No. 12 Office of Education, U.S. Department of Health, Education and Welfare).

Flanders, N.A. Interaction analysis and inservice training. Journal of Experimental Education. Fall, 1968, 37, 126-133.

Flanders, N.A. Analyzing teaching Behavior. Reading, Mass: Addison Wesley, 1970.

Froman, R.D. and Owen, S.V. Influence of different types of student ratings feedback upon later instructional behavior. Paper presented at the Annual meeting of the American Educational Research Association, Boston, Mass., April 7-11, 1980 (ERIC, ED 187 724).

Fuller, F. and Baker, H. Counseling teachers: Using video feedback of their teaching behavior. (Report #0010). University of Texas at Austin, Research and Development Center for Teacher Education, 1970. (ERIC #ED 058760).

Gage, N., Runkel, P., and Chatterjee, B. Equilibrium theory and behavior changes: An experiment in feedback from pupils to teachers. University of Illinois, Bureau of Educational Research, 1960.

Glassberg, S. Student teaching supervision designed, implemented and evaluated from a developmental perspective. Paper read at American Educational Research Association, Toronto, Canada, March, 1978.

Goldhammer, R. Clinical supervision: special methods for the supervision of teachers. New York: Holt, Rinehart and Winston, 1969.

Good, T., and Brophy, J. Changing teacher and student behavior: An empirical investigation. Journal of Educational Psychology, 1974, 66, 390-405.

Hail, G.E. Computer processing and feedback of interaction analysis data for teachers. Journal of Classroom Interaction, 1978, 13, (2), 32-37 (ERIC, EJ 195 492)

Hamachek, D.E. Encounters with the self. (2nd ed) New York: Holt, Rinehart and Winston, 1978.

Hamachek, D.E. Psychology in teaching, learning and growth. (2nd ed.) Boston: Allyn and Bacon, 1979.

Harris, B.M. Supervisory behavior in education, Englewood Cliffs,N.J.: Prentice Hall, 1963.

Hartman, C.L. A longitudinal look at self-appraisal strategies. Journal of Teacher Education, September-October, 1978, 29 (2), 11-12.

Howey, K.R., Yarger, S.J., and Joyce, B.R. Improving teacher education. Washington, D.C.: Association of Teacher Educators, 1978.

Joyce, B.R., Howey, K.R., and Yarger, S.J. Issues to face: ISTE Report 1. Syracuse University, National Dissemination Center, 1976.

Karnos, A. and Jacko, C. The role of significant others during the student teaching experience. Journal of Teacher Education. September-October, 1977, 28, (5), 51-55.

Kash, M.M. and Borich, G.D. Teacher behavior and pupil self concept. Reading, Mass.: Addison-Wesley, 1978.

Koran, J.J., Jr. Supervision: An attempt to modify behavior Educational Leadership, May, 1969, 26, 754-757.

Kotula, J.R. Affecting community college teaching behavior through feedback. Unpublished doctoral dissertation. Rutgers University, 1975.

Litwack, M. A study of the effects of authority feedback, peer feedback, and self feedback on learning selected indirect-influence teaching skills. Unpublished doctoral dissertation, Indiana University, 1974.

Medley, D.M. The language of teacher behavior: Communicating the results of structured observations to teachers. Journal of Teacher Education, Summer, 1971, 22, 157-165.

Moore, J., Schaut, J., and Fritzges, C. Evaluation of the effects of feedback associated with a problem-solving approach to instruction on teacher and student behavior. Journal of Educational Psychology , 1978, 70, 200-208.

McDonald, F.J. The New York City Teacher Center, Year One: Description, research evaluation. (Volume Three: The teacher specialists: Their training and its effects). Princeton: Education Testing Service, 1980.

McDonald, F.J. and Allen, D.W. Training effects of feedback and modeling procedures on teaching performance. Stanford University, Stanford Center for Research and Development in Teaching, 1967. (Technical Report No. 3, ERIC ED 017 985).

Mead, G.H. Mind, Self and Society. Chicago: University of Chicago Press, 1934.

Olivero, J.L. Helping teachers grow professionally. Educational Leadership. December 1976, 34, 194-200.

Piele, P.K. Review and analysis of the role, activities and training of educational linking agents. Eugene, Oregon: University of Oregon, 1975. (ERIC Clearinghouse on Educational Management.)

Pohlmann, J.T. Evaluating instructional effectiveness with the Instructional Improvement Questionnaire. Southern Illinois University, Carbondale, Counseling and Testing Center, 1976 (ERIC, ED 121 836).

Porter, A.C. Comparative experiment in educational research. Draft of a script for an audiotape prepared as one in the series, "Alternative methodologies for Research in Education," Richard M. Jaeger (Ed.) American Educational Research Association, 1980.

Purkey, W.W. Inviting school success--a self concept approach to teaching and learning. Belmont, Calif: Wadsworth, 1978.

Roper, S.S. et al. A pilot test of collegial evaluation for teachers. (Research and Development Memorandum No. 142). Washington, D.C.: National Institute of Education, 1976 (ERIC ED 126 120).

Rosenberg, M. Conceiving the Self. New York: Basic Books, 1979.

Rosenshine, B. Teaching behaviors and student achievement. London: National Foundation of Educational Research in England and Wales, 1971.

Rubin, L. Professional development, Vol.1: Perspectives on preservice and inservice education. Syracuse University, National Dissemination Center, National Council of States on Inservice Education, 1978.

Ryans, D.R. Characteristics of teachers, their description, comparison and appraisal: A research study. Washington, D.C.: American Council of Education, 1960.

Semmel, M.I., et al. The effects of CATTS' feedback in a preservice special education teacher trainng program. (Final Report 53.32). Indiana University, Bloomington Center for Innovation in Teaching the Handicapped, 1976, (ERIC, 162 466).

Semmel, M.I. and Olson, J. System documentation manual. Indiana University, Center for Innovation in Teaching the Handicapped, 1977 (ERIC, ED 162 465).

Sergiovanni, T.J. (Ed.) Professional supervision for professional teachers. Washington, D.C.: Association for Supervision and Curriculum Development, 1975.

Sergiovanni, T.J. Toward a theory of clinical supervision Journal of Research and Development in Education, Winter, 1976, 9, (2), 20-29.

Sergiovanni, T.J. Reforming teacher evaluation: Naturalistic alternatives. Educational Leadership, 1977, 34, 602-607.

Shuck, R.F. Microteaching in teacher education programs. Microteaching: Selected Papers . ATE Research Bulletin. No. 9, September 1971 (ATE & ERIC, SP 004 978).

Simon, A.E. Videotapes illustrating concepts of the Argyris and Schon model in instructional supervisory situations. Unpublished doctoral dissertation, University of Illinois, 1976.

Simon, A.E. Analyzing educational platforms: A supervisory strategy. Educational Leadership, 1977, 34, 580-584.

Spencer, M. A study to evaluate the impact of structured and unstructured feedback to teachers: Unpublished doctoral dissertation, Rutgers University, 1973.

Sullivan, H.S. Conceptions of modern psychiatry. Washington, D.C.: W.H. White Psychiatric Foundation, 1947.

Trank, D.M. The effect of student feedback on the improvement of instruction. University of Iowa, 1978 (ERIC ED 177 642).

Tuckman, B.W. and Oliver, W.F. Effectiveness of feedback to teachers as a function of source. Journal of Educational Psychology, 1968, 59, 297-301.

Tuckman, B.W. Feedback and the change process. Phi Delta Kappan, 1976, 57, 341-344.

Tuckman, B.W., McCall, K., and Hyman, R. The modification of teacher behavior: Effects of dissonance and coded feedback. American Educational Research Journal, 1969, 6, 607-619.

Walencik, V.J. An experimental study to determine the effectiveness of feedback as a means of changing student teachers' humanistic behavior. Unpublished doctoral dissertation, Rutgers University, 1973.

Walencik, V.J. and Tuckman, B.W. The effectiveness of feedback for changing student teachers' humanistic behavior . Paper presented at meeting of the American Educational Research Association, Chicago, 1974.

Webster, M., Jr. and Sobieszek, B. Sources of self evaluation: A formal theory of significant others and social influence. New York: John Wiley and Sons, 1974.

Whetstone, R.D. Innovation in faculty-course evaluation and feedback. Paper presented at the Annual meeting of the American Personnel and Guidance Association, New Orleans, Louisiana, April 1974 (ERIC, ED 103 447)

Withall, J. An objective measurement of a teacher's classroom interaction. Journal of Educational Psychology, 1956, 47, 203-212.

Wylie, R.C. The self concept: A critical survey of pertinent research literature. Lincoln: University of Nebraska Press, 1961.

Wylie, R.C. The self concept Volume One: A review of methodological considerations and measuring instruments. (Rev. ed.) Lincoln: University of Nebraska Press, 1974.