ADMINISTRATORS' RATINGS OF TEACHER PERFORMANCE AND PERCEPTIONS OF RATINGS ON THE NORTH CAROLINA PERFORMANCE APPRAISAL SYSTEM

by

Virginia Stone Myers

Dissertation submitted to the Faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of DOCTOR OF EDUCATION in Educational Administration

APPROVED:

Jimmie C. Fortune, Co-chairman

Robert R. Richards, Co-chairman

Houston Conley

David Mutter

Wayne M. Worner

August 1989
Blacksburg, Virginia
The purpose of this study was to examine school administrators' observations, ratings, and perceptions of rating scores on the North Carolina Teacher Performance Appraisal Instrument. The process of evaluating teaching behaviors and reaching agreement on the ratings was also addressed. Attention was directed to three research questions: Do administrators rate teaching performance consistently using the six-point rating scale of the North Carolina Teacher Performance Appraisal Instrument? How do administrators perceive the ratings on the performance appraisal instrument? Do administrators have confidence in their ratings of teaching performance?

A teacher evaluation workshop was conducted with building level principals and central office administrators in a small school system in North Carolina. Administrators
viewed videotaped teaching episodes and rated five observable teaching behaviors included on the North Carolina Teacher Performance Appraisal Instrument. Individual and group consensus ratings were made. Participants were interviewed following the workshop to record perceptions of the ratings and the evaluation process.

Findings in this study revealed that administrators rated within an acceptable limit for the five observable teaching behaviors. There was closer agreement on the teaching behaviors which were rated below standard than on those rated satisfactory and above. Differences in the perceptions of what a particular rating means were evident. Discriminating among the four ratings which were At Standard or higher was more difficult than deciding if a teaching behavior was acceptable or unacceptable. Administrators willingly changed ratings to reach agreement with other raters, indicating little confidence in their own ratings.
ACKNOWLEDGMENTS

This study was completed through the assistance and encouragement of many people to whom the writer is indebted. Sincere appreciation is extended to the writer's committee: Jimmie C. Fortune, co-chairman; Robert R. Richards, co-chairman; Houston Conley; David Mutter; and Wayne M. Worner.

The writer thanks her husband, , whose patience, understanding, and encouragement made possible the realization of this study.
# TABLE OF CONTENTS

ABSTRACT ........................................... 11
ACKNOWLEDGMENTS ............................... iv
LIST OF TABLES ..................................... lx

## CHAPTER ONE

- Introduction ..................................... 1
- The Need for the Study .......................... 4
- Significance of the Study ...................... 8
- Purpose of the Study ............................ 9
- Assumptions ..................................... 9
- Limitations of the Study ....................... 10
- Definitions .................................... 12
- Organization of the Study ..................... 14

## CHAPTER TWO

- Review of the Literature ...................... 16
  - Section 1: Instructional Leaders as Teacher Evaluators ........................................... 16
  - Section 2: Teaching Evaluation ................ 23
  - Section 3: North Carolina Performance Appraisal ................................................. 40
- Summary ......................................... 55

## CHAPTER THREE

- Introduction .................................... 57
- Design of the Study ............................ 58
Comments........... 182
Findings and Conclusions........... 184

Question 1: Findings ........... 184
Question 1: Conclusions........... 184
Question 2: Findings ........... 185
Question 2: Conclusions........... 186
Question 3: Findings ........... 186
Question 3: Conclusions........... 187
Other Findings ........... 187

Recommendations ........... 188
Implications........... 189

LIST OF REFERENCES ........... 193

APPENDICES

A. Public Law — G. S. 115C-326........... 203
B. Teacher Performance Appraisal Instrument and Rating Scale ........... 204
C. Formative Observation Data Instrument (FODI) . 211
D. Formative Observation Data Analysis (FODA) . 212
E. Professional Development Plan . 214
F. Principal Performance Appraisal Instrument . 215
G. "The Brandt Report" — Executive Summary . 224
H. Teacher Evaluation Workshop Agenda and Script . 233
I. Interview Guide ........... 242
J. Rater Information Form ........... 243
K. Teacher Questionnaire and Responses........244
VITA........................................250
# LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Tape 1: Means and Standard Deviations for Function Ratings.</td>
<td>80</td>
</tr>
<tr>
<td>2. Individual Ratings for Tape 1</td>
<td>81</td>
</tr>
<tr>
<td>3. Tape 2: Means and Standard Deviations for Function Ratings.</td>
<td>86</td>
</tr>
<tr>
<td>4. Individual Ratings for Tape 2</td>
<td>87</td>
</tr>
<tr>
<td>5. Tape 2: Ratings by Two-member Teams</td>
<td>90</td>
</tr>
<tr>
<td>6. Number of Times and the Direction That Participants Changed Their Scores to Reach Group Consensus</td>
<td>92-93</td>
</tr>
<tr>
<td>7. Tape 3: Means and Standard Deviations for Function Ratings.</td>
<td>95</td>
</tr>
<tr>
<td>8. Individual Ratings for Tape 3</td>
<td>96</td>
</tr>
<tr>
<td>9. Tape 3: Ratings by Three-member Teams</td>
<td>98</td>
</tr>
<tr>
<td>10. Tape 4: Means and Standard Deviations for Function Ratings.</td>
<td>101</td>
</tr>
<tr>
<td>11. Individual Ratings for Tape 4</td>
<td>102</td>
</tr>
<tr>
<td>12. Tape 4: Ratings by Four-member Teams</td>
<td>105</td>
</tr>
<tr>
<td>15. Tape 5: Ratings by Two-member Teams</td>
<td>112</td>
</tr>
<tr>
<td>16. Tape 6: Means and Standard Deviations for Function Ratings</td>
<td>118</td>
</tr>
<tr>
<td>17. Individual Ratings for Tape 6</td>
<td>119</td>
</tr>
<tr>
<td>18. Tape 6: Ratings by Four-member Teams</td>
<td>122</td>
</tr>
</tbody>
</table>
19. Tape 7: Means and Standard Deviations for Function Ratings ............... 128
20. Individual Ratings for Tape 7. .................................. 129
21. Tape 7: Ratings by Six-member Teams. .......................... 133
22. Tape 8: Means and Standard Deviations for Function Ratings ............... 138
23. Individual Ratings for Tape 8. .................................. 139
24. Differences in Participants' Rating and Group Mean Ratings for First and Last Observations. ..................... 147-148
25. Comparison of Standard Deviations and Means Between Tape 1 and Tape 8. .......................... 149
26. Comparison of Individual Ratings to Group Means .......................... 166
CHAPTER ONE

INTRODUCTION

As part of a coordinated effort to improve the education of students in the state of North Carolina, high quality teacher performance is essential. Through the evaluation of teachers, educators at all levels seek to enhance the quality of teaching to which the public school students are exposed: as a result, education improves. As evidenced by the formation of public task forces, through legislation, and through citizens' demands for accountability at all levels, this interest in educational improvement is shared by educators, political leaders, business and industry representatives, and citizens.

The North Carolina Department of Public Instruction (DPI) developed a system for evaluating teacher effectiveness as one part of a response to the concern for improving the instructional program for students. The Teacher Performance Appraisal System (TPAS) was developed and adopted by the State Board of Education in 1982. The system which had been developed, field tested, and revised
the previous year, was mandated for implementation in all school systems across the state immediately (G.S.115C-326 - See Appendix A). In the 1982-83 school year, local education agencies (LEA's) began using the system, including the Teacher Performance Appraisal Instrument (TPAI - See Appendix B).

The responsibility for assisting teachers in improving effectiveness lies with administrators and supervisors at the local level. These are the same people who are responsible for using the TPAS to evaluate teacher performance. With this responsibility comes the challenge of consistent and fair evaluation for all employees.

In a survey of 32 school districts (Wise, 1984), most respondents identified two major problems with teacher evaluation: the principal's lack of competence to evaluate accurately and the inadequacy of training for evaluators. The identification of such problems is the first step to improving the evaluation process and to producing useful results for the professionals involved. Further, it was found that the school districts with successful evaluation programs provided means to guarantee competent evaluators. As early as 1978, other research showed concern over the
variability in teacher ratings by administrators using new methods of evaluation (Holley, 1980).

The North Carolina Career Development Program (CDP) which was begun in pilot units in 1985 as outlined in G.S.115C-363, is dependent on fair and consistent evaluations by building administrators and their designees. Central office personnel including superintendents, supervisors, directors, and observer/evaluators are involved in these evaluations also. Advancement through the career development program is dependent upon satisfactory performance as measured on the TPAI. To move from one level to another in a teaching career, one must receive ratings at set minimums. Thus, teacher advancement on the career ladder which leads to higher salary, more status and responsibility, and job promotion is determined by the results of the evaluations conducted by administrators, central office supervisory personnel, and observer/evaluators. As reported to the State Board of Education (North Carolina Department of Public Instruction, October, 1988, February, 1989), an analysis of evaluations of participants in the North Carolina Career Ladder Program pilot showed that most rating errors have resulted in scores that are too high. The decisions made by evaluators about advancement on the career ladder will determine salary and
special duties of teachers. Similarly, decisions to demote, to fire, and to deny tenure will rely on these evaluations for evidence of performance.

The Department of Public Instruction (DPI) provides a four day (24 hours) training program on the use of the TPAS. All administrators, supervisors, and observer/evaluators who use the TPAI are trained using the DPI model (North Carolina Department of Public Instruction, 1985-86). Following the completion of this training, administrators and the other observers are considered qualified to implement the evaluation system. Teachers are trained in effective teaching methods in a five day (30 hours) module which has been taught to administrators previously (1985). With the completion of the training modules for teachers and their evaluators, the TPAS is used for performance appraisal.

THE NEED FOR THE STUDY

One problem which exists in the North Carolina teacher evaluation system is the variation in ratings by different evaluators. Differences in the meanings of the rating scores on the Teacher Performance Appraisal Instrument lead
to ratings which are not consistent from one evaluator to another.

Data collected at the North Carolina Department of Public Instruction (NCDPI) from the pilot studies for the Teacher Performance Appraisal System and from the Career Development Plan (CDP) indicate that problems continue to exist in the understanding of the criteria and standards of the evaluation system (NCDPI, Analysis of Teacher Performance Ratings in Career Development Program Pilot Units, 1987-88, 1988). Appeals by teachers in the CDP units who feel that their ratings are unfair, have been filed as allowed by legislation for the career development pilots. During the first two years of the pilot, 69 appeals were filed, the majority of which ended with the upholding of the principals' decisions (NCDPI, Performance Appraisal -- Cornerstone of Career Development Plan, 1987). According to a consultant in the Department of Public Instruction (Barbara Kuligowski, 1989), appeals in the last two years of the pilot program have been decided in a balanced manner, about half in favor of the teacher, about half supporting the principal's decision. Administrators' confidence in their rating decisions are always questioned in the appeals procedure. Indications from these actions show a need for
further training in the use of and perceptions regarding the
teacher evaluation instrument and the rating scores.

Studies carved out by other researchers have contributed to the body of knowledge and recommendations for further study of the appraisal system. A study by Phillip Scott Riner to test for correlation between the TPAI function ratings and student achievement, suggests that study should be continued to explore the evaluation instrument and evaluators' perceptions of the rating scores (1988). Problems such as rater reliability and ineffective use of the TPAI by administrators have been addressed in a doctoral dissertation by Daniel Massey at The University of North Carolina at Chapel Hill (1988). Judith Wolfe's 1988 study (Virginia Polytechnic Institute and State University) recommends that on-going training/remediation "be implemented to ensure that objective, defensible data are consistently collected and processed by observers." Henderson (UNC-Chapel Hill, 1987) and Jewell (North Carolina State University, 1989), in reports on the North Carolina system of teacher evaluation, recommend that concerns with the perceptions of rating scores should be addressed. According to sources at several North Carolina universities, other studies are in progress but have not yet been completed.
At the local level, administrators in one small rural school system in North Carolina decided there was a need to examine TPAI rating skills and patterns in that system. During a staff meeting in the winter of 1988, sixteen administrators in this school system viewed and rated a single videotaped lesson of a kindergarten teacher using the North Carolina Teacher Performance Appraisal Instrument. In tabulating the ratings, it was found that the administrators had recorded different scores for the same function. The discussion which followed the rating activity revealed that the justifications for the ratings varied, and that administrators were not confident of their rating skills. From that activity, it was determined that a need existed for further practice in observing and evaluating teaching behaviors in order to achieve consistency and to practice the skills used in the TPAS. The instrument had been used for one year at the time the administrators determined the need to examine evaluator consistency in rating, reasons for assigning particular scores to the teaching functions, and rater confidence in scores.
SIGNIFICANCE OF THE STUDY

This study provides insight as to how evaluators select specific ratings including the influence of the observers' backgrounds and the impact of team discussion and decision-making on those ratings.

Information on how raters perceive specific teaching behaviors can be used to develop follow-up training programs for observers. Continuous monitoring of observation and rating skills can improve the evaluation process and training programs for evaluators. This study provides information needed to plan further training and staff development for administrators using the system.

Administrators place importance on possessing skills needed to evaluate teachers fairly. There are two concerns related to the outcomes of the evaluation process. One deals with the goal of improving instruction in the schools. The second addresses a practical set of concerns for teachers: decisions based on evaluations determine advancement on the career ladder, promotions, diversity of professional responsibilities, status, re-employment or dismissal, and salary. Evaluators must be aware of these
concerns. This study examines ways of addressing these concerns.

PURPOSE OF THE STUDY

The purpose of this study was to examine the process of rating teaching performance using the North Carolina Teacher Performance Appraisal Instrument. The following research questions were addressed:

1. Do administrators rate teaching performance consistently using the six-point rating scale of the North Carolina Teacher Performance Appraisal Instrument?

2. How do administrators perceive the ratings on the performance appraisal instrument?

3. Do administrators have confidence in their ratings of teaching performance?

ASSUMPTIONS

Two sets of assumptions were made for this study. Assumptions inherent in the evaluation system were: 1) student learning increases when effective teaching practices
are used; 2) teacher evaluation leads to improved teaching performance; 3) teachers perform as "expected and inspected" by superiors; 4) evaluation is conducted by raters who are competent and knowledgeable in effective teaching practices and in teacher evaluation. Participants in this study brought with them certain assumptions including the following: 1) acceptance of the assumptions of the evaluation system; 2) desire to improve skills as an evaluator; 3) staff development activities can improve one's evaluating skills; 4) the opportunity to interact with other administrators in professional activities is desirable.

LIMITATIONS OF THE STUDY

The methodology used in this study has an inherent limitation. Ethnography is an attempt to observe and record everything that takes place in a particular setting. While observers of the consensus process attempted to see, hear, and record the interactions of participating evaluators, each observer made choices about what was most important to record because everything that happened could not possibly be recorded. The interpretation of non-verbal interactions of participants was also made by the observer during the study.
Participants in the study were administrators in the same small school system. The desire to cooperate with peers with whom they would continue to work may have influenced their willingness to agree on a different rating rather than insisting on the correctness of one's own rating. Although consensus rating was used successfully by the participants in this study, administrators may have hesitated to say things which were contrary to expressed views of other raters with whom they have built respect and rapport, leading to less than honest ratings.

Videotaped lessons provided a limited view of the classroom activities and interactions. Actual classroom observing enables the evaluator to scan the room and see more than one area at a time. No checking for rater preferences or ability to observe and record from a videotape lesson was conducted prior to the rating activity. The taped lessons rated were not equally representative of a cross-section of teacher characteristics of race and gender. Videotapes of minorities and male teachers at all grade levels were not readily available.

Attitudes of the administrators may have influenced ratings. Empathy with the teacher being observed because of
past experiences as a teacher could have been present. Rater objectivity was not measured beforehand.

All 18 participants did not view and rate every taped lesson because of the schedule and setting of the workshop. Some administrators missed sessions to check on matters at their job sites, although the rating workshop was held away from schools or central office buildings. It is impossible to verify the effect that distractions such as telephone calls to or from job sites and taking care of unexpected job responsibilities may have had on an administrator's ability to focus full attention on rating teaching behaviors.

DEFINITIONS

For the purpose of this study, the following definitions apply:

ADMINISTRATOR - Building-level principal or assistant principal, superintendent, directors, and supervisors with administrative responsibilities as identified by the local school board.
AGREEMENT OF RATINGS - Multiple ratings which have a standard deviation (SD) of .70 or lower. Close agreement is any SD < .70.

CONSISTENCY - Plus (+) or minus (-) one point on the six-point rating scale used in the NCTPAS. This range is considered acceptable for the NCTPAS.

EFFECTIVE TEACHING PRACTICES - The thirty-eight practices, grouped under eight functions, which have been identified through research as directly related to successful instruction, and which are included on the NCTPAS.

EVALUATORS - Administrators, supervisors, directors, and observer/evaluators who have been trained and are responsible for rating teacher performance on the NCTPAS.

FUNCTIONS - The eight categories of the act of teaching under which thirty-eight practices have been identified as effective teaching practices for the North Carolina Teacher Performance Appraisal System. The first five functions that relate directly to classroom teaching are: 1) Management of Instructional Time; 2) Management of Student Behavior; 3) Instructional Presentation; 4) Monitoring of Student Performance; 5) Instructional Feedback.
NORTH CAROLINA TEACHER PERFORMANCE APPRAISAL SYSTEM (NCTPAS) - The system used to evaluate eight function areas, five of which are observable for formative purposes, adopted by the North Carolina Department of Public Instruction. As required by law, the system must be used for formative and summative evaluation of all teaching personnel in North Carolina. Teacher Performance Appraisal System is defined in Chapter 16 of the North Carolina Administrative Code section 2f.0600.


ORGANIZATION OF THE STUDY

Chapter one provides an introduction to the study, including background on the North Carolina teacher evaluation system, and the need for the study.

A review of recent literature on the role of instructional leaders as evaluators, teacher evaluation
efforts, and performance appraisal in North Carolina public schools is contained in Chapter two.

Chapter three describes the methodology used in this study. Methods, procedures, and data of the study are discussed.

Findings are presented and analyzed in Chapter four.

Chapter five includes a summary, findings, conclusions, recommendations, and practical implications of this study.
CHAPTER TWO

REVIEW OF LITERATURE

In order to provide the reader with pertinent background information for this report, three areas of the literature are reviewed. The first section of this chapter deals with effective schools research, with emphasis on the role of instructional leaders as related to evaluating teachers. Section Two reports on teacher evaluation from a general perspective, followed by the third section which is a review of the North Carolina teacher evaluation system.

Section 1: INSTRUCTIONAL LEADERS AS TEACHER EVALUATORS

The term "effective schools" has been used in the literature for the past two decades. A study which has had a strong influence on the present effective schools movement was done by Ron Edmonds (1977). Characteristics of effective schools were identified and named in that project. It disputed the results of earlier studies such as the 1966 Coleman report, Equality of Educational Opportunity, and the 1973 Jencks report entitled Inequality, which attributed
school success to social factors, family background, and IQ, rather than factors within the control of the school.

The characteristics of effective schools have been identified as those that emphasize the basic skills of communication and computation, and as schools in which all students, regardless of social background or family economics, achieve academically (Edmonds, 1977; Brookover, 1981; Squires et al., 1985; Mann, 1989; Browne, 1989). It is further noted that factors characteristic of an effective school cannot be examined independently, but must be considered in the environment of the school. In looking at schools, more than achievement of students is examined. Patterns of behavior, instructional practices, and strong leadership are indicative of a school's effectiveness.

Throughout the research conducted in the past few years, there is evidence that specific aspects of a school's organization can determine effectiveness. Administrative leadership, personnel, and school climate are closely related to achievement. Squires, Hultt, and Segars in Effective Schools and Classrooms: A Research-based Perspective (1985), define the six elements of an effective school as leadership, school climate, teacher behaviors, supervision, student behaviors, and student achievement.
Active strong leadership which emphasizes high expectations for success, academics, and an orderly school environment is necessary. These are some of the same characteristics which were identified earlier.

It appears that any number of management styles can be conducive to the process of establishing effective schools. Whatever the particular style, there must be a strong principal who helps define and maintain the environment that promotes achievement and discipline. The principal is the most important person in the school when it comes to providing leadership. The principal as an instructional leader provides teachers with direction, feedback on performance, and incentive to produce greater student achievement. As Squires states, "Through supervision, teachers are aware of how their planning, instruction, and management patterns affect their students and their students' achievement," (1985, p.2). It is on this basis that teacher evaluation is promoted as an integral part of an effective school.

Numerous studies address the need for a principal to demonstrate instructional leadership skills by working with teachers to analyze teaching behaviors which lead to greater student achievement. A sense of vision about human
potential must be conveyed to the staff by the effective educational leader, Tanner states in a report by the NASSP Curriculum Council (1986, p.11). Principals who offer support to teachers have a more positive effect on classroom instruction. Rutherford (1985, p.33) found that the more effective principals continuously monitor progress, supply specific details about teaching performance, and provide insights into why the teachers performed as they did. Studies reported by Rothberg and Buchanan (1981), Bossert et al. (1982), and Hallinger and Murphy (1985) further evidence the importance of evaluating teacher behavior and working with teachers to conceptualize instructional goals. The role of the school administrator as the teacher's supervisor, Wise notes in *Teacher Evaluation: A Study of Effective Practices* (1984), includes the responsibility for evaluating performance. Encouraging the teacher's efforts is recognized as part of the leader's role.

The 1986 journal article "Excellent Schools: The Leadership Function of Principals" by Mangieri and Arnn, and a later study by the same authors, report their findings that the principal who is an effective instructional leader evaluates teacher performance regularly, expecting to see teachers teaching. These principals observe classroom performance looking for enthusiasm, effective communication
skills, involvement, and productive modeling (Arnn and Mangieri, 1988, p.6). As early as 1963, there was strong agreement that two of the most important aspects of the principal's job were assisting teachers and initiating professional growth opportunities, responsibilities which go along with teacher evaluation (Ploughoft and Perkins, 1988, p.26).

Principals as leaders must set instructional goals and determine how to evaluate them. Where clear school goals do not exist, there is little basis for teacher evaluation, and the principal does not take time to observe. According to Rosenholtz in writing about teaching reform (1984, p.17), providing teachers with information to help them, and making judgments about promotion and tenure are the two roles that the principal must address in evaluation. Donald Reyes, in the report *Applying Teacher Effectiveness Research in the Classroom*, continues with a similar recommendation for the principal's role in evaluation as a developmental approach to observation and the improvement of instruction (1986, p.17).

As a school becomes more effective, the leadership of the principal becomes more important. The strong leader knows which teachers are good and which are not, Dale Mann
stated in an address to members of the North Carolina Association for Supervision and Curriculum Development (1989). He continues by pointing out that the principal who externalizes failure and rates all teachers outstanding guarantees a school to fail, for teaching behaviors determine the effectiveness of schools. Bad teachers become comfortable and good teachers become frustrated. The effective principals are in classrooms observing and evaluating teacher performance. Mann proposes that there is not a single intervention that will improve schools, but a comprehensive strategy of which purposeful teacher evaluation is a part.

The basic assumption of school effectiveness that teacher evaluation and accountability improve instruction and learning is touted by another current educational spokesperson. Instructional leaders, Art Costa stated in an address to North Carolina educators in January, 1989, must continue to devote time and energy to evaluating teacher performance. The administrator's role is interpreted to mean that teachers must be held accountable; check lists of competencies and skills should be used to evaluate teachers; and inadequate teachers must be "fixed." Regarding teacher assessment, the 1989 president of the Association for Supervision and Curriculum Development continued by focusing
on team evaluation, coaching, goal clarifying, and more time spent envisioning the broad goals of instruction rather than on singular evaluation of teaching behaviors. The instructional leader in effective schools collaborates with teachers to evaluate performance in the classroom.

This same attention to the leader's role in promoting collaboration in evaluation is addressed in articles for *Educational Leadership* by Crandall (1983, p.7) and by McCormack-Larkin (1985, p.33). In a 1988 interview with the editors of the *NASSP Bulletin*, John Gardner states that cooperation and collaboration between teachers and principals should be a part of the evaluation process ("Principals and Leadership - An Interview with John Gardner," 1988, p.77). The principal should view teacher evaluation as a shared responsibility, just as the practice of teaching is a matter of mutual interest, according to Smith and Blase in an article written for the National Association of Secondary School Principals (1988, p.7).

One characteristic of effective schools is that the principal communicates high expectations of the staff and students. That teachers should be regularly reminded of expected standards, evaluated, and rewarded when they achieve them has been recognized throughout the effective
schools research (Duke, 1982, p.6; Sweeney, 1982; Lamb and Thomas, 1984, p.21). The nationally recognized report A Nation At Risk: The Imperative for Educational Reform called for attention to teacher performance and reward (1983). Writing about the principalship, Lamb and Thomas state that evaluation becomes a means of quality control for an orderly environment, high expectations, monitoring of progress, and quality instructors. The caring principal, "does not tolerate mediocrity...expects excellence from oneself, from one's colleagues, and from students" (1984, p.21). In a 1983 study by Hallinger and Murphy, promoting and maintaining high expectations for teachers was a characteristic found in effective schools (Ploghoft and Perkins, 1988, p.25). Larry Lezotte (1988) echoes this sentiment by emphasizing the importance of the principal as a manager of the process to develop schools into effective schools.

Section 2: TEACHER EVALUATION

Teacher evaluation has taken many forms in past years. From the perfunctory checklist of personal characteristics such as attitude, intelligence, dress, and physical appearance, all of which were termed "presage," to the
currently prevailing "process" and "product" methods (Soar, 1983, p.7-8), evaluation continues to challenge administrators and teachers. The definition of teacher evaluation includes a task analysis approach which deals with the process of collecting and analyzing information about past activities and events in order to plan and control future activities and events (Hawley, 1982, p.9). Evaluation is "the set of required or recommended policies, procedures, processes, and instrumentation that directs the attitudes and actions of the participants" (McGreal, 1983, p.8). Evaluation is the set of processes and activities which take place to enable a judgment to be made about the worth of the teaching behavior. This generally accepted definition of teacher evaluation is based on the purposes of the evaluation system.

In 1983, the report *A Nation At Risk: The Imperative For Educational Reform* called for effective evaluation systems as a basis for salary, promotion, tenure, and retention decisions for teachers in this country. This report triggered reform movements in the teacher evaluation systems being used. Among the recommendations made in the report was a call for accountability for the quality of the educational services offered to our students. The National Commission on Excellence in Education asked that teachers be
required to meet rigid standards and "demonstrate competence in an academic discipline" (p.30). Evaluation should improve or terminate poor teachers, encourage average ones, and reward superior ones. These recommendations provided impetus for local and state units to develop effective teacher evaluation systems.

Performance appraisal or teacher evaluation may be used for many purposes. Formative evaluation encourages individual development, while summative evaluation increases accountability for individuals and the institution (system). These two general purposes of formative and summative evaluation have been agreed upon in the literature (Bolton, 1971, 1973; Abbott, 1975, p.51; Redfern, 1980; Knapp, 1982; Krajewski, Martin, and Walden, 1983, p.207; Cresap, McCormack and Paget, 1984, p.30; Barber, 1985, p.1; Parker, 1985, p.11; Ellis, 1986, p.1). Specific purposes of evaluation systems are designated as they are developed and implemented.

Evaluation serves to improve teaching and specific skills of the teacher. At the same time, it facilitates personnel or job status decisions. The mixture of the two purposes causes difficulty for teachers and school systems. Although the most effective plan for evaluation leads to the
improvement of learning conditions for students, it should also contribute to higher teacher morale. Research in the early 1970's indicated the purposes for teacher evaluation as follows: to improve teaching through the identification of ways to change teaching systems, environments, and behavior; reward superior performance; supply information for the modification of assignments; protect students from incompetence and to protect teachers from unprofessional administrators; to validate the school system's teacher selection process; and provide a basis for the individual teacher's career planning and development (Bolton, 1971). Manatt, in referring back to the 1976 School Improvement Model approach to teacher evaluation, names the same six general purposes (1988, pp.83-84). In addition to the above list, Abbott included: the gathering of information that may lead to promotion or termination. He acknowledges the formative and the summative functions of evaluation (1975, p.51). The evaluation process identifies areas in which improvement is needed or areas in which existing strengths can be enhanced. Improvement of instruction as the primary goal of evaluation is addressed by Foley in a document about teacher evaluation. He recommends that only those teachers most in need of improvement be identified and evaluated (1981, p.2).
Four basic purposes served in teacher evaluation as identified by Wise et al. in the report of the Rand teacher evaluation study (1984, p.11), represent improvement and accountability concerns. For individuals, improvement through staff development is approached; while personnel decisions for individuals (such as promotion or termination) relate to the accountability function. For the organization, school improvement is a purpose; school status decisions such as accreditation serve the accountability purpose. In deciding the types of data the system will use in the evaluation process, the Wise report continues, the district must consider the goals of the system to ensure that conflict will not occur.

As the basis for diagnosis, alternative assessment, and prescription, evaluation is a powerful tool. According to Buser and Pace in an article entitled "Personnel Evaluation: Premises, Realities, and Constraints," (1988, pp.7-8), evaluation provides opportunities for communication, observation, and coaching. Standards, expectations, and the mission of the institution can be communicated through the evaluation system. These two university professors continue that the primary purpose of evaluation should be instructional improvement and staff development. Evaluation purposes include staff development, acknowledgement of job
performance, determination of employment status, inservice training, and carrying out of mandates or laws.

Four goals of teacher evaluation as classified by Crenshaw and Hoyle are: to assess the overall school program, to motivate teachers to perform at their highest level, to provide a basis for improving instruction, and to provide a basis for making administrative decisions. Emphasis is placed on the idea that evaluation is a method which should be done "with" rather than "to" an individual. All the goals should be directed toward the improvement of instruction (1981, pp.38-39). A few years later in the book Improving Teacher Performance: Formative Evaluation, Barber notes that there are two broad categories of evaluation: a system that rewards and/or punishes, and a system designed to improve teaching performance. Teachers want to know how they are doing, he continues, to help them improve their performance (formative); yet, evaluation is used as a tool by administrators to judge the net worth of the teacher's performance (summative). Barber insists that the two systems of formative and summative do not mix (1985, p.1).

In promoting the cognitive development approach to evaluation, four purposes of teacher evaluation are recognized. Two focus on the individual teacher: improve
teacher performance and inform personnel decisions. Two focus on the organization or district: improve organizational performance and inform organizational decisions (Costa, Garmston, Lambert, 1988, p.147). These authors concur with Glickman in his 1986 report to the National Curriculum Study Institute that there is virtually no evidence that teacher evaluation practices have improved instruction for students (p.169). Emphasis on one purpose may limit another, just as some purposes are inconsistent with others.

The purposes of teacher evaluation have not changed significantly in the last two decades, but the procedures for conducting performance appraisal have changed. Methods which are peer-oriented have been used. Self-evaluation, team evaluation, quantitative performance ratings, and qualitative systems are available. Clinical supervision is a method used when the purpose is to improve instruction. It is generally agreed that the purpose of the evaluation should determine the criteria and standards used for the process. The person who evaluates should know and understand the purpose for the evaluation so that the performance appraisal can be conducted professionally. If the purpose is to determine promotions, terminations, or
change in teaching assignments, the evaluator should view the process with that particular end result in mind.

There are many models of teacher evaluation systems in use today. In some of these, characteristics of more than one system are combined to meet the school system needs. Often the system evolves from dissatisfaction with previous attempts. More and more school districts are finding the need for using a combination of data sources in evaluating their teachers. Some models being implemented today are outlined according to the purpose (McGreal, 1983). Common law models, goal-setting models, product models, clinical supervision models, and artistic or naturalistic models are described in McGreal's Successful Teacher Evaluation. In the descriptions, he provides characteristics, advantages, and problems or issues.

The types and quantity of data used to evaluate teachers are issues. Multiple data sources are recommended in most systems today. Classroom observation, used as the basis for many formative evaluations, is criticized by Michael Scriven in "Teacher Personnel Policies: Equity, Validity, and Productivity," as inadequate to fairly judge performance. He proposes using a new system that would only use facts about the teacher, such as: estimate of the amount
learned by students (pre- and post-testing, and/or student evaluations); a judgment of the quality of the content learned by students; and the use of a legitimate teaching process (1980). A few years later Scriven offered a duties-based approach as a better method (1988).

Sources of data recommended for effective evaluation systems in addition to classroom observations, include parent evaluation, student evaluation, peer supervision, self-evaluation, student performance, and artifact collection (McGreal, 1983, pp.125-140). McGreal notes that parent evaluation is likely the least useful, with peer supervision, student performance, and self-evaluation providing important information. He continues that the use of student evaluation and artifact collection have potential also. Wise et al., reporting on a McNeil and Popham 1973 study, stated that the "product" of the teaching process, student achievement test scores, should be a source of data for teacher evaluation, but a National Education Association survey conducted in 1979 found that teachers disagreed with the use of this information (1984, pp.8-9). As early as 1975, House suggested that a group of professionals should work with the teacher to provide multiple outcome data to use for evaluation (p.7/7).
The research as presented, overwhelmingly agrees that the two major purposes of teacher evaluation are the improvement of instruction and the basis for personnel decisions. For an evaluation system to be successful, there must be congruence and compatibility between the purpose and the process of the system.

Current issues identified in the well-known Rand report Teacher Evaluation: A Study of Effective Practices are being addressed in varying degrees by individual school districts. The issues concern disagreement in regards to instrumentation, frequency of evaluation, the role of teachers in the process, and how to use the information for other district activities (Wise et al., 1984, p.21). The commonalities of successful evaluation systems identified by the same report are: the necessity for commitment by the leaders of the school district; competence of the evaluators; collaboration of the teachers and the evaluators/administrators in the implementation and the use of the process; and the compatibility of the purpose and the process of the particular system (p.26).

Problems which have occurred in teacher evaluation systems recently envelop process, the evaluators, and the use made of the results. Gathering credible data, the use
of a structured observation form, and the training in the use of the particular observation form were found to be problems in current systems (Soar et al., 1983, p.241; Yap and Capie, 1985, p.7; Buser and Pace, 1988, p.86; Wolfe, 1988, p.82). The criteria on which an instrument is based are not clearly defined in terms which will withstand legal challenges, according to a current description of on-going issues in the field of teacher evaluation (French, 1988). The legal problems which continue to develop are characterized in a recent court case in Georgia in which a teacher questioned the Teacher Performance Assessment Instruments that were used to determine his teaching competency (White, 1988). In this case, the teacher was concerned because he was being terminated for reasons which the Board of Education would not state, but which were based on the adopted instrument. The criteria which the instrument addresses were being questioned by the teacher. In the report Career Ladder/Master Teacher Programs: Implications for Principals, James Parker stresses that the criteria and standards of evaluation should be differentiated by career level, coordinated with the purpose of the evaluation, include meaningful teaching behaviors, be free of favoritism, and should reflect levels of teaching proficiency. Acceptable levels of validity, reliability,
and utility are required (1985, p.11). Describing the Behaviorally-Anchored Rating Scales (BARS) method for establishing performance criteria, Conley emphasizes that clear performance standards must be established for a system to work (1988, p.78).

Validity and reliability are ethical issues, a 1983 study by Donovan Peterson reports. The test of validity is whether the teacher evaluation criteria are job-related and whether the evaluation system actually measures the attributes it claims to measure. Reliability is based on the consistency of judgments over time among separate observers. These tests must be considered at each of three levels for the observations: experienced teachers, using formative only; new teachers and the ones needing improvement, with the emphasis on formative using several trained observers to identify problems and develop strategies for improvement; and teachers considered incompetent judged on summative evaluations. In the last group, multiple observers are used, remediation is offered, and provisions for assistance are in place (Peterson, 1983).

The collection of data remains a problem with evaluation processes. Actual classroom observation is a common method of collecting data for the formative
evaluation. Script-taping, especially useful for the purposes of clinical supervision, is one technique that has been adopted for use in many systems. In addition to actual observations, which have been claimed to be disruptive, artificial, and subjective (Andrews and Knight, 1988, p.2), videotaped lessons have been used in the collecting of data. In 1984 using the Florida Performance Measurement Systems' Summative Observation instrument, videotapes of actual classroom performance were used for the data for ratings. In that test, it was found that performance was reliably rated using videotapes when at least two separate observations by at least two observers were conducted. Using single observations, the system was not deemed reliable for summative purposes (Micerri, 1984, p.25). The methods for collecting data, whether through artifact collection, classroom observation, self-evaluation, or evaluation by parents or students, is an issue that continues to be addressed by researchers (Rose and Huyah, 1984, p.13; Andrews and Knight, 1987, p.2-3; Herrmann, 1987, p.29; Buser and Pace, 1988, p.87; McGreal, 1988, pp.19-21; Manatt, 1988, p.80; Leslie, 1989, p.20).

Evaluators themselves present problems to the system of teacher performance appraisal. Evaluators must be trained in the purpose and use of the system. Observation
skills can be learned, but evaluators must develop the skills to a level that will assure accuracy and reliability (Manatt, 1985, p.16). There is a continual concern with the generalizability across observers, or interobserver agreement (Holley, 1980, p.6; Leese, 1981, p.23; McGreal, 1983, p.viii; Manatt, 1983, p.6; Wise et al., 1984, p.26; Yap, 1985, p.7; Holdzkom, 1987, p.43; Smith, B. Othanel et al., 1987, p.18; Watson et al., 1987, p.22). The question of reliability of the instrument when used by different observers is addressed in the development and training phases of the system. Evaluator competence is a difficult element of the process. An evaluator must not only be able to use the instrument selected by the district, but must be able to make judgments based on the collected data. Using the data to make recommendations for improvement is an expectation for evaluators. In a study of four successful evaluation systems, it was found that each system had mechanisms for verifying the accuracy of the evaluators' reports about teachers. Concrete, precise terms had to be used to justify ratings (Wise et al., 1984, p.vii).

Role conflict of the evaluator continues to be a problem. In most systems, the principal is the primary evaluator. This principal is also the supervisor who has the responsibility for being an instructional leader. As an
instructional leader, the principal works to build collegial relationships which can be cumbersome in an evaluator's role. McGreal offers a solution for the role conflict problem by recommending that administrators learn to "tilt" their administrative hat and act more as instructional supervisors than building administrators (1983, p. 38). The tendency for principals to be more lenient with teachers who will be working with them is illustrated by the action that school systems have had to take. In one, principals were put into remediation programs themselves when they failed to refer any teachers for remediation when it was well known that incompetent teachers existed in the system. In a second, principals' evaluation skills are viewed in the principal's evaluation (McLaughlin and Pfelfer, 1986, pp.67-68). A teacher evaluation study done in 1984 revealed that building level principals tended to rate higher than district-level observers in many cases. That principals rated experienced teachers more consistently than in-experienced teachers, and that district staff rated teachers within their own field with higher scores than teachers outside their own field were found. Overall, the findings concluded that there were minor differences in the ratings of district observers and principals. Any differences tended to show principals to be more lenient in
their ratings, possibly because of the conflicting evaluative and support roles (Rose and Huyah, 1984, pp.11-13). In North Carolina, ratings by local observers and principals compared with ratings of outside state evaluators in the Outside Evaluator Project, revealed consistently higher ratings by the local evaluators than the outside state evaluators (North Carolina Department of Public Instruction [NCDPI], *Certified...,* 1989, pp.12-13).

Teacher resistance and apathy was cited as a problem in the Rand report. Some of the resistance is a result of the perceived incompetence of the evaluator to be consistent and uniform with all evaluations (Wise et al., 1984, p.22). More has been written about teachers' lack of confidence in the system. There is a concern with rater bias. Human error in rating performance is categorized in five forms by Manatt: 1) leniency/severity/central tendency; 2) halo effect; 3) rater characteristics; 4) rater position; and 5) personal bias (1988, p.101). The method used to resolve these problems is usually more training for the raters. In the SIM project results of a study of a large southern school district, all five types of rater bias were found. Among the results were: females rated male and female teachers significantly lower than did male evaluators; minority raters had lower average ratings than the average
ratings of white principals; evaluators with more than a master’s degree rated teachers higher than raters with less than a master’s degree; evaluators with 11 to 15 years experience rated tougher than raters with more or less experience; gender and race were found not to be significant; ratings were not effected by the distance of the appraisers (from the same campus or a different campus); and the first semester appraisals predicted the second semester appraisals accurately (Manatt, 1988, p.102). A study by Tadlock in 1984 found that the evaluation system in Jackson, Mississippi, did not have the problems with years of experience, race, certificate field and level, and the school level. None of these factors proved significant in ratings over a two-year period. The inconsistencies which were revealed were associated with variances with ratings at different schools and the lack of understanding of the standard score concept (pp.3-6). Similar concerns with variability across schools have been reported by other researchers (Holley, 1980, p.13; Soar et al., 1983, p.214; Ferguson, 1985).
Section 3: NORTH CAROLINA PERFORMANCE APPRAISAL

The North Carolina General Assembly, in 1979, enacted legislation which set the stage for the "development of uniform performance standards and criteria to be used in evaluating professional public school employees" (Section 35, Appropriations Act, 1979 General Assembly, Second Session, 1980). Following this legislation, the State Board of Education and the North Carolina Department of Public Instruction (NCDPI) initiated a study to develop a system of professional evaluation. Through the combined efforts of the NCDPI, the State Board, universities, and professional associations of educators, the construction of a performance appraisal system has continued since 1978 (Holdzkom, 1987, p.41).

In the statute enacted by the General Assembly, local boards of education were required to evaluate the performance of teachers annually using criteria established by the State Board of Education. Local boards could adopt rules and regulations which went beyond the requirements of the statute, but could not do less than required or countermand the State Board's criteria. The State Board asked the NCDPI to construct an evaluation instrument that
would meet the criteria set forth to measure teacher performance (Holdzkom, 1987, p.41).

Subsequent legislation amended the original act and allowed for pilot field-testing in various systems across the state before statewide implementation. The pilot field-testing allowed time for testing the standards and criteria in 24 local school administrative units and for training of personnel involved in the implementation. During the two years 1981-83, a series of three-day training sessions were developed and implemented for staff in every school unit in the state. Research studies were completed during that period to determine the validity of the standards and the reliability of the process. The instrument as developed at that time, contained 33 standards and was implemented statewide, July, 1982 (NCDPI, Teacher Performance Appraisal System: The Standards and Process for Use [TPAS: Standards], 1986, pp.3-4).

As the effective schools research began to evolve, the Department of Public Instruction contracted with the Group for the Study of Effective Teaching, based at the University of North Carolina at Chapel Hill, to review the research on effective teaching. The results of the study were presented in 1983 as the final report of Phase I of the Carolina
Teaching Performance Assessment System. This project resulted in the identification of 28 specific practices that could be divided among five basic teaching functions (White, Wyne, Stuck, and Coop, *Teaching...,* 1983). The five functions were Management of Instructional Time, Management of Student Behavior, Instructional Presentation, Monitoring of Student Performance, and Instructional Feedback. The NCDPI, using the collected research, determined that the goal of personnel evaluation would be to improve performance, not terminate or eliminate individuals from the teacher ranks (Holdzkom, 1987, p.41).

To assure that whatever practices the teacher was being held accountable for were applicable to the teaching situation, the practices had to meet the following criteria: the practices must be generic across grade levels and subject areas; practices must have been identified in more than one research study; and the practice must be alterable (the teacher could change to learn it). Twenty-eight practices were identified that relate directly to classroom teaching under five major teaching functions: Management of Instructional Time (4 practices); Management of Student Behavior (5 practices); Instructional Presentation (11 practices); Instructional Monitoring of Student Performance (4 practices); and Instructional Feedback (4 practices).
Three additional functions including Facilitating Instruction (5 practices), Communicating Within the Educational Environment (2 practices), and Performing Non-Instructional Duties (3 practices), were identified as job-related but not necessarily part of daily instruction. The Teacher Performance Appraisal Instrument is made up of these 38 practices grouped under 8 functions (see Appendix B) (NCDPI, TPAS: STANDARDS, 1986, pp.5-6). In evaluating teacher performance, the functions rather than individual practices, are rated to measure the teacher's competence.

Based on the results of a pilot test of the instrument in 40 local school units during 1984-85 and 1985-86, the State Board of Education adopted the policy in May, 1986, that "Beginning July 1, 1987, all local education agencies will be required to use the new research-based criteria, procedures, and processes for teacher evaluation." The Department of Public Instruction developed and implemented training workshops for preparing educators for the use of the performance appraisal system (NCDPI, TPAS: STANDARDS, 1986, p.6).

Several training modules have been designed to assist local school districts in implementing the Teacher Performance Appraisal System.
Training (ETT), Professional Development Plans (PDP), and Mentor/Support Team Training (M/STT) form the core of the training programs. Additional modules continue to be developed to offer assistance to the local agencies.

Effective Teaching Training (ETT), a 30-hour program that explains the 38 critical practices on the Teacher Performance Appraisal Instrument, uses videotapes, mini-lectures, role-plays, and other strategies to prepare teachers to use the evaluation plan (NCDPI, North Carolina Effective Teaching Training Program, 1985). The next training module in the series, Teacher Performance Appraisal System Training (TPAS), teaches observation, recording, and evaluation skills to principals and other observer/evaluators in a 24-hour program. Experience in making appropriate evaluations is provided through the use of videotaped teaching episodes (NCDPI, North Carolina Teacher Performance Appraisal System Training, 1985-86). In early training experiences, 75% of the staff trained were able to accurately rate teachers' performance, according to NCDPI sources (NCDPI, Teacher Performance Appraisal System Training, 1986, p.16).
Another module, **Mentor/Support Team Training (M/STT)**, is a 24-hour program aimed at helping teachers who will work with initially certified teachers in the first two years of teaching after graduation from an accredited college. This training, based on effective teaching practices and adult learning models, develops skills needed to assist new teachers (NCDPI, *Mentor/Support Team Training*, 1986). The **Professional Development Plan (PDP)** training teaches principals how to develop growth plans for personnel improvement in a six-hour session (Holdzkom, 1987, p.44). Other training programs for practicing and developing observation and evaluation skills have been designed and are in various stages of development.

The process of teacher evaluation in North Carolina has been outlined for local school units. The effective teaching practices which have been correlated by research with positive student achievement are the basis for the skills assessed in the evaluation process. The rating scale for the functions which are observed in the classroom and as related teaching factors, has six points. The State Board of Education adopted the scale and definitions as follows: 6. Superior; 5. Well Above Standard; 4. Above Standard; 3. At Standard; 2. Below Standard; 1. Unsatisfactory (NCDPI, *TPAS:Standards*, p. 15) (see Appendix B). In the training
for TPAS, practice in using the rating scale and achieving inter-rater reliability are stressed.

The principal at the building level has the responsibility for evaluation of teachers. General statutes and State Board policy place this responsibility on the principal or the designee, as approved by the local superintendent. The principal is assessed on the newly adopted Principal's Performance Appraisal Instrument in at least three functions for supervising and evaluating staff (see Appendix F). Under the function of Instructional Leadership, the principal "Provides supervision to promote growth and the attainment of program goals" (IC,1), and "Evaluates instructional staff, using criteria and processes adopted by the board of education" (ID,2). As a Resource Manager, the principal "Provides supervision to promote growth to attain goals of resource management program" (2C,3). In the area of Communications, the principal "Interprets and carries out the policies established by the local board, State Board of Education, North Carolina School Law, and federal law" (3B,4). The State Board of Education states that all persons who use the Teacher Performance Appraisal Instrument must have "successfully completed" the Teacher Performance Appraisal Training (NCDPI, TPAS:Standards, p. 17).
Multiple observations are required for valid use of the Teacher Performance Appraisal Instrument (TPAI). Using the forms suggested by the North Carolina Department of Public Instruction (NCDPI), observers proceed with observation and evaluation of teacher performance. Observations may be conducted by persons other than the principal, but the final evaluation is the responsibility of the principal. According to NCDPI guidelines, observers do not complete a teacher evaluation form (TPAI) at the end of each observation. They use a Formative Observation Data Instrument (FODI) and Formative Observation Data Analysis (FODA). After multiple observations, the TPAI is completed using the rating scale for each of the eight functions (NCDPI, TPAS:Standards, p.17) (see Appendix for forms).

In the Standards and Process for Use booklet on the Teacher Performance Appraisal System, the North Carolina Department of Public Instruction provides guidelines for conducting performance appraisal. Each teacher should be observed on at least three occasions during each school year. This requirement was modified in 1987 to require that career status teachers could be placed on an alternating year basis in which three observations are conducted every other year, with one observation required on alternating years as long as the teacher received "At Standard" or above
ratings on each of the eight function areas. Probationary and initially certified teachers must be observed a minimum of three times annually. There are more inclusive guides to follow for initially certified teachers with a support team, to ensure that assistance will be given to these inexperienced teachers. Specific dates for completion of observations to ensure interval timing of observations are provided. Of the three minimum observations, one should be announced with a pre-conference and at least one should be unannounced. All observations should be followed by a post-conference to discuss the classroom performance. For summative observation and the development of a professional growth plan, the principal and the support team, in the case of an initially certified teacher, rate the teacher using the six-point scale on the Teacher Performance Appraisal Instrument. Copies of the observations/recommendations and the professional development plan should be included in the teacher's file at the school and in the superintendent's office. The teacher and the principal sign the instrument indicating that the observation has been reviewed and discussed (NCDPI, TPAS: Standards, pp.18-20). Procedures and laws have been established for due process for the appeal or request for re-consideration of actions taken as a result of the evaluation process.
Effective July 1, 1985, the State Board of Education approved a process requiring that every certified person have a Professional Development Plan (PDP) which will help the person keep current, acquire better skills, and plan for improvement in the field (NCDPI, *TPAS:Standards*, p.23). The local administrative unit is responsible for maintaining the PDP with each certified educator (see Appendix E). This PDP is an integral part of the North Carolina performance appraisal process.

Legislation in 1985 established the North Carolina Career Development Plan which was designed to improve the quality of classroom instruction through the recognition and retention of high quality teachers (G.S. 115C-363). Sixteen local school systems across the state were selected to participate in the Career Development Pilot/Field Test for a period of four years. Teachers who take part in the pilot study will have the opportunity to receive higher pay for classroom performance as evaluated by local and state evaluators, according to the NCDPI *Career Development Plan Guide* ([CDP Guide] (1985). Evaluation is based on indicators associated with effective classroom practices included on the TPAI and other criteria. Movement within the four differentiated levels of the career ladder for teachers is determined by evaluations. The administrators' Career
Development Plan was scheduled to run simultaneously (NCDPI, CDP: Guide, 1985). Reports on the progress of the Career Development Pilot are made to the General Assembly each year, providing up-dates on evaluation ratings, status of participants, and changes made in the program as the field test progresses. These reports summarize and analyze the progress of the CDP units in moving toward the goal of implementing a valid and reliable system of evaluating professional educators on a statewide basis. The most recent reports on the Career Development Program analyze teacher performance ratings and student achievement (NCDPI, Analysis of Teacher Performance Ratings in Career Development Program Pilot Units 1987-88, December, 1988; NCDPI, Student Achievement in Career Development Program Pilot Units 1985-88, November, 1988). These reports indicate that rater error continues to exist, and the error tends to be in the direction of positive leniency. Specific recommendations include: reliability checks of raters should be conducted by the staff of the Department of Public Instruction; evaluators should conduct joint observations to check reliability; evaluation data should be reviewed periodically; summative ratings should be decided on through a consensus process; superintendents should review summative ratings; DPI should develop additional training for the last
three functions (6, 7, and 8); periodic assessments of rater reliability should be conducted; and wide dissemination of the results of studies should be done.

As a continuation of the development of valid and reliable evaluations of teachers and administrators, the 1985 General Assembly established the Outside Evaluator Project as a pilot program in nine local school systems (G.S. 115C-362). Data comparing summative performance appraisal ratings for two groups of teachers and one group of principals are collected in this pilot project. Evaluations performed by persons employed by an agency other than the school units to which they are assigned (state evaluators) and evaluations performed by locally-employed personnel are conducted and compared annually for the three-year pilot (NCDPI, Outside Evaluator Project, 1988; NCDPI, Certified School Personnel, 1989.) Results of this project show that locally-employed evaluators consistently tend to rate teachers higher than state evaluators. For more than 80 percent of the teacher comparisons, however, ratings were identical or differed by only one point.

The direction that North Carolina takes in the future for its teacher performance appraisal system has been addressed by educators and non-educators. In 1988, a TPAS
Review Panel Report was submitted to the North Carolina General Assembly. Among the conclusions made by the third party program review referred to as "The Brandt Report," was that the TPAI "is a quality instrument, one that is highly suited to its purpose" (Brandt, 1988, p.1). Twenty-four recommendations to further improve the instrument and enhance its effectiveness were presented (see Appendix G for Executive Summary). The NCDPI reviewed and began taking action on the recommendations. The recommendations related to concerns addressed in this study are: 6) Maintain the six-point rating scale; 7) Follow a step-by-step procedure in making summative judgments; 11) Certification of evaluators - require a performance-based test of competency, demonstrate competence on a regular basis, must be certified to function as evaluators; 12) Career Ladder Decisions; 13) Summative Ratings by Consensus; 15) Joint Observations/Analyses - Conduct joint observations, gather data, and share analyses to enhance the consistency of instrument use and interpretation; 16) Conduct reliability studies to assess consistency across raters; and 23) Assign greater weight to performance of principals as teacher evaluators.

Questions about the consideration of content knowledge in evaluating teacher performance have been addressed by
NCDPI officials and noted educators throughout the state. In a February, 1988, memorandum to all principals, the Assistant State Superintendent for Personnel Services Area, Robert Boyd, stated that procedures for ensuring content knowledge of teachers are in place through the monitoring of academic programs and the requirement of a passing score on the National Teacher Exam. The presentation of content is addressed in functions 6 and 3, he asserts. Content knowledge has also been addressed by the North Carolina Association for Supervision and Curriculum Development in its Fall, 1988, publication *Teacher Evaluation and the Treatment of Content*. David Holdzkom, Director of the Personnel Relations Division of the North Carolina Department of Public Instruction which oversees the Career Development Program in the state, concludes that evaluation must include assurances of adequate content knowledge as well as accountability for educating children. There is not a "clean division between bureaucratic responsibility and professional responsibility" (Holdzkom, 1988, p.17). The researchers who did the original study of effective teaching practices for the North Carolina system, Group for the Study of Effective Teaching, reiterate that the five observable functions of the North Carolina Teacher Performance Appraisal Instrument, serve only as one part of the complex
skills necessary for effective teaching (Stuck and White, 1988, p.3).

Public groups of business and community leaders continue to express concern with the evaluation of North Carolina teachers. A 1986 report by the Southern Growth Policies Board asks for higher standards for the teaching profession, provision for better teacher education, rewards for successful performance, and public recognition of the costs for quality education (p.12). Another group of civic leaders endorsed and supported the Career Development Plan with specific recommendations in a widely published "Position Statement" in 1988 (League of Women Voters of Charlotte-Mecklenburg). The Public School Forum, a group of business and education leaders in North Carolina, have presented a plan for educational growth that includes rewarding superior teaching performance and making the evaluation process "more goal-oriented and professionally enhancing." Individual performance and student achievement would both be considered in the recommended Career Development Program (Public School Forum of N. C., 1988, p.19).

The North Carolina program for evaluating teacher performance continues to develop as more is learned through
pilot studies, research, input from practitioners, dialogue among professionals, and implementation of a quality system. Indications from educators and other state leaders are that teacher performance will remain an important component for providing a quality education for the students in the state.

SUMMARY

The review of the literature establishes the need for administrators to be competent evaluators of teaching performance. The section on instructional leaders directs attention to the effective administrators' responsibility to evaluate teaching and to help teachers improve the delivery of instruction.

Teacher evaluation in recent years is concerned with the processes used to rate teaching performance through various methods, especially observation and rating of specific effective teaching practices. Problems with administrators' perceptions of rating scales and interpretation of teaching practices throughout education systems is reviewed.
The review of the North Carolina teacher performance system of evaluation indicates on-going concerns as well as progress made to date.
CHAPTER THREE

INTRODUCTION

This chapter provides an overview of the methodology used in this study. Procedures, instrumentation, data treatment, and data analysis are described.

The purpose of this study was to answer the following research questions:

1. Do administrators rate teaching performance consistently using the six-point rating scale of the North Carolina Teacher Performance Appraisal Instrument?

2. How do administrators perceive the ratings on the performance appraisal instrument?

3. Do the administrators have confidence in their ratings of teaching performance?

To answer these questions, a research model based on describing a series of rating activities and the data obtained from the activities was used.
DESIGN OF THE STUDY

The study was conducted in a small rural school system of 4300 students in northeastern North Carolina. The study involved administrators at the central office and building levels in an activity to examine their use of the process and their perceptions of the rating scale employed in the evaluation of teachers using the North Carolina Teacher Performance Appraisal System (TPAS). Administrators viewed eight videotaped teaching episodes, evaluated five observable teaching behaviors or functions for each lesson, discussed their ratings in groups, and assigned consensus ratings for each teaching function. Consensus rating was used for several reasons. It had become evident through research on the NCTPAI and through discussions with leaders of teacher organizations that teachers were more accepting of consensus ratings than individual ratings to assure equity. Teachers reject the idea of one person determining ratings on evaluations. Disagreements about the quality of teacher performance had led to recommendations for consensus ratings in teaching evaluation, and teacher trust was greater than with individual rating. Administrators desired the opportunity to affirm their own rating skills through the discussions and sharing that are a part of the consensus
process. During the consensus ratings, an observer recorded interactions among the participating administrators. The administrators were interviewed following the activity to record their individual perceptions and understanding of the total evaluation process and of the group rating activity. While the administrators were participating in the group rating activity, teachers in the school system responded to a written questionnaire concerning the implementation of the Teacher Performance Appraisal System. The same administrators in the rating activity workshop had evaluated the teachers who responded to the questionnaire about the evaluation process.

Descriptive methodology was selected because of the need for flexibility to include many types of data as it evolved. Using exploratory techniques including group discussions, consensus decision-making, open-ended questions, and observation of participant interactions, encouraged the development of research concerns and results which could not be anticipated before the study began. Searching for perceptions of the administrator/evaluators in using the teacher appraisal system required a margin of freedom not allowed in many research methods.
DESCRIPTION OF PARTICIPANTS

The participants in the study were employed as full-time administrators or supervisors/directors of the board of education of the small school system. Each of the eighteen participants was certified by the State Board of Education for the position held. Nine of the participants were central office supervisors/directors; nine were building level principals. Observers who recorded interactions of the rating teams were school administrators from other school systems who had been trained in observation skills and in group dynamics. The 161 teachers who responded to a written questionnaire were employed in the school system in which the participating administrators worked, and had been evaluated by these same administrators the previous year.

INSTRUMENTATION AND PROCEDURES

The instrumentation used in this study is described in this section. An outline of the procedures used in the study is included also.
Nine observers to record the interactions of the raters during consensus activities were trained following the script in Appendix H. These observations were used to record the raters' verbal interchanges and body language which would indicate perceptions of the rating scores and the evaluation process. These observers were trained during the morning of the first day of a workshop in which school administrators (evaluators) practiced their teacher observation and evaluation skills.

Following the observer training, a three-session workshop in which the participants viewed and rated eight videotaped teaching episodes was conducted. The agenda and script for the workshop is located in Appendix H. A Rater Information form was completed by administrators in the participating school system when they arrived at the workshop. The form was used to obtain demographic data which is presented in Chapter 4.

During the rating activities, instruments used for the collection of ratings for the five observable teaching behaviors were those used with the North Carolina Teacher Performance Appraisal Instrument to evaluate teachers. The three forms were utilized in the order described. These
forms are available in the appendix (see Appendices B, C, and D).

The Formative Observation Data Instrument (FODI) was used to record notes and to categorize teaching behaviors during the lesson observations. The FODI lists the five observable teaching behaviors and the 28 practices which were included in the state-mandated evaluation process (TPAS). Space was provided on the form for evaluators to write notes.

The Formative Observation Data Analysis (FODA) form was used after the viewing of each teaching episode to categorize and summarize each of the five major teaching functions that are a part of the Teacher Performance Appraisal System: Management of Instructional Time; Management of Student Behavior; Instructional Presentation; Monitoring of Student Performance; and Instructional Feedback. Strengths and weaknesses were listed on this form also. Notes from the Formative Observation Data Instrument (FODI) were used to write the narratives on the Formative Observation Data Analysis (FODA).

Following the completion of the FODA, the individual evaluators rated each of the five teaching behaviors using the rating scale of the Teacher Performance Appraisal.
Instrument. This six-point scale consists of these scores: (6) Superior; (5) Well Above Standard; (4) Above Standard; (3) At Standard; (2) Below Standard; and (1) Unsatisfactory. The complete rating scale with definitions is found in Appendix B.

Rating scores obtained during the evaluation activities for the eight videotaped teaching episodes were collected at the end of the rating activity for each tape. Ratings were compiled in three categories: "All Participants," "Principals," and "Central Office" administrators. Means and standard deviations of ratings were calculated for each category. The mean score was calculated to obtain the average rating for each of the five teaching functions for each lesson. The standard deviation (SD) was calculated to determine the degree of agreement among raters. These data were used for comparison and analysis throughout the study. Data were placed on tables for visual examination in the report of findings found in Chapter 4.

The next instrument used in the study was a set of interview questions asked the participating administrators. These 16 questions, designed by the researcher to obtain perceptions of the rating process and to explore the perceived confidence of the raters, were posed in individual
interviews conducted during the two weeks following the teacher evaluation workshop. Participants' responses were recorded in written and audio form. The audio tapes were used to clarify written notes and to record exact quotations for the data analysis.

A Teacher Questionnaire about the teacher evaluation system was distributed to all teachers in the school system who had taught in that system the previous year and who had been evaluated using the North Carolina Teacher Performance Appraisal Instrument (see Appendix K). The purpose of this questionnaire which was designed by the researcher, was to report the views of teachers who had been evaluated by the participating administrators in the workshop. The teacher responses were used to explore reactions which may conflict with or confirm administrators' perceptions of the evaluation process. The questionnaire was distributed, collected, and responses compiled while the teacher evaluation workshop and the participant interviews were being conducted.

The sequence of events in the study were as follows:

I. Distribute Teacher Questionnaire (April 25)
II. Observer Training (April 26)
III. Teacher Evaluation Workshop for Administrators
   A. Session I (Day 1- April 26)
      1. Morning Session
a. Complete Rater Information form  
b. View and rate Tape 1 (Individually)  
2. Afternoon Session – Individual and team ratings for each lesson  
a. View and rate Tape 2  
b. View and rate Tape 3  

B. Session II (Day 2– May 3)  
1. Morning Session – Individual and team ratings for each lesson  
a. View and rate Tape 4  
2. Afternoon Session – Observed team ratings  
a. View and rate Tape 5  
b. View and rate Tape 6  

C. Session III (Day 3– May 4)  
1. Morning Session – Observed team ratings  
a. View and rate Tape 7  
2. Afternoon Session  
a. View and rate Tape 8 – same as Tape 1 (Individually)  
b. Workshop closure  

IV. Interview Participants (May 5–27)  
V. Collect Teacher Questionnaires (May 20)  
VI. Analyze Data (begin May 27)  

The data collection part of this study began when the Teacher Questionnaires were delivered to every school. Instructions were left with the written questionnaires for the principals who would distribute them during a regularly scheduled staff meeting within the next three weeks.  

Observers for the consensus rating activities were trained using the script in Appendix H. Because all observers had experience observing and recording group interactions, the training was scheduled to familiarize them with the specific procedures for the teacher evaluation workshop they would be observing. Instruction and practice
in the observation process emphasized the recording of body language, facial expressions, changes in voice tones and speech patterns, as well as the recording of ratings, comments, and reasons for ratings.

The teacher evaluation workshop for the 18 administrators began later the same morning that the observer training was conducted. During the workshop, videotaped lessons were viewed and rated using the North Carolina Teacher Performance Appraisal Instrument. Eight teaching episodes were viewed: primary class, grade 12 English, grade 7 literature, grade 7 math, grade 9 English, kindergarten, grade 8 literature, and the primary class a second time. The workshop was organized into three sessions.

Session I (Day 1) was divided into a morning session and an afternoon session. The morning session began with the administrators completing the Rater Information form. The first tape, a teaching episode of a primary teacher, was viewed and rated by the individual administrators using the procedures and forms from the North Carolina Teacher Performance Appraisal System. A group discussion of the lesson, ratings, reasons for rating scores, and differences
in evaluators' ratings completed the morning session. Data were collected from the administrators.

The afternoon session of the first day included the viewing and rating of two teaching episodes, Tape 2 and Tape 3. Procedures for individual observation and rating of these tapes were the same as for Tape 1 with the exception of the group discussion. In place of the total group discussion, an additional activity was added for these tapes. After each evaluator viewed and rated the lesson, teams were assigned to reach consensus on ratings for each of the five teaching behaviors. Teams were assigned by drawing names from the two groups of administrators—principals and central office administrators. The mixture of principals and central office administrators provided an opportunity for all administrators to interact with participants they did not work with on a daily basis. In order to provide the raters with experience working with different people and different sized teams, the composition of the teams was changed for each tape. For Tape 2, two-member teams were assigned. The teams met, discussed observations and ratings, and decided on team ratings. Team consensus ratings were collected, but no total group discussion was held. All data were collected. For Tape 3,
individual ratings were completed, followed by three-member team consensus ratings, and the collection of the data.

The morning session of Day 2 consisted of the viewing and rating of Tape 4. The individual ratings were completed following the same procedures as Tapes 2 and 3. For the team consensus ratings, four-member teams were assigned to rate the teaching functions. Data were collected.

The afternoon of Day 2 added a new element to the procedures. The individual ratings followed the pattern used for the previous tapes. The team ratings were observed by administrators brought in from other school units who had been trained the morning of Day 1 of the workshop. As the teams met to reach consensus on the ratings, an observer recorded interactions of team members in each group. Following the individual rating process for Tape 5, seven teams of two administrators were formed to reach consensus on the five teaching behaviors while observers recorded their actions. At the end of the rating activity, data were collected from the evaluators and the observers of each team. Tape 6 was viewed and rated the same afternoon, using the process described for Tape 5. For Tape 6, teams of four administrators, two principals and two central office administrators, met to make a consensus decision on the
ratings. The observer of each team recorded interactions during the team activity. Data were collected at the end of the session.

Session III (Day 3) began with the last of the observed team rating activities. Tape 7 was viewed and rated individually by the administrators, following the procedures used for the previous tapes, with no group discussion. Six-member teams were formed for the consensus rating. An observer recorded interactions of team members in each group. Data were collected at the end of the activity.

The afternoon session began with the viewing and rating of the last tape, Tape 8. This lesson was the same tape used for Tape 1. The first tape was repeated to check for any changes in ratings that the individual raters may have made during the course of the workshop. Tape 8 was viewed and rated individually only. No team ratings were made. Data were collected. The session ended with a closure which summarized the three-day workshop, gave participants an opportunity to discuss their feelings and reactions to the workshop, and an overview of follow-up activities such as the individual participant interviews and the analysis of the data.
In the three weeks following the workshop activity, individual participants were interviewed using the set of 16 questions contained in Appendix I to obtain their reactions to the workshop and to the evaluation process. All teacher questionnaires were collected for analysis by the end of the time interviews were held. All data were compiled and examined.

DATA ANALYSIS AND REPORTING

Data were collected and analyzed based on the three research questions. The reporting of data is in the following sequence and forms:

1. Participant characteristics - narrative report of demographic information on administrator/raters.

2. Responses to individual lessons - rating scores, standard deviations, means. Individual ratings and team ratings are provided.


4. Raters' perceptions - narrative, using observations from the team consensus activities and responses to the workshop participant interviews.

5. Characteristics match - narrative and tabular report of administrator/evaluator characteristics and those of the videotaped lessons.

6. Teacher perceptions - narrative report of responses to the teacher questionnaire.
The research questions and specific procedures which address them are presented.

To answer the first question, "Do administrators rate teaching performance consistently using the six-point rating scale of the North Carolina Teacher Performance Appraisal Instrument?", data are used from the individual videotaped lesson ratings, ratings summary, standard deviations and means of rating scores, and characteristics matches. Responses from teachers who were evaluated by the administrators are used to report perceived consistency and success with the process.

The answer to the second question, "How do administrators perceive the ratings on the performance appraisal instrument?" uses information from the observed team rating discussions and the participant interview. Administrators' own descriptions of the ratings are used.

Interview responses and reports from the observers during the team consensus rating activity are used to answer the question, "Do administrators have confidence in their ratings of teaching performance?" Examination of comments and responses of the participating administrators which may be contradictory or which express the certainty of a rating is made.
CHAPTER FOUR

PRESENTATION AND ANALYSIS OF DATA

This chapter reports the results of a study of teacher evaluation and the perceptions of participating administrators in a small rural school system who used the North Carolina teacher evaluation system to assess teaching performance. The study was conducted to answer these three questions: 1) Do administrators rate teaching performance consistently using the six-point rating scale of the North Carolina Teacher Performance Appraisal Instrument? 2) How do administrators perceive the ratings on the performance appraisal instrument? and 3) Do the administrators have confidence in their ratings of teaching performance?

The school administrators viewed and rated teaching behaviors on eight videotaped teaching episodes. The evaluators used the procedures and instruments that are a part of the North Carolina Teacher Performance Appraisal System to rate the first five functions of the instrument. The functions rated were: 1) Management of Instructional Time; 2) Management of Student Behavior; 3) Instructional
Presentation; 4) Instructional Monitoring of Student Performance; and 5) Instructional Feedback. These five functions are observable in a classroom teaching situation, and are the functions based on effective teaching practices. The other three functions on the appraisal instrument, Facilitating Instruction, Communicating Within the Educational Environment, and Performing Non-Instructional Duties, deal with behaviors which must include information obtained from sources other than a classroom lesson; therefore, they are not addressed in this study.

Following the rating activities, the administrators, including building-level principals and central office supervisors/directors in a small, rural North Carolina school unit, were interviewed to record their reactions and perceptions of the rating scale and the evaluation process. Through the use of a questionnaire, teachers in the same school district were asked about the evaluation process and their experiences and reactions to it. The teacher responses were used to explore reactions which may affect administrators' perceptions of the value and use of the evaluation process.

Data gathered during the study included: demographic data on the participants; individual and team ratings on the
videotaped lessons; written notes taken during the observation of the videotaped lessons; written analyses of the five teaching functions; observer notes on the interactions of the participants during the team ratings; participant interviews with responses to questions about the appraisal process; and responses to a teacher survey of seven questions which dealt with the evaluation process. These data are used to examine the similarities and differences in ratings on the five functions and to discover the reasons for the various ratings by the administrators.

PARTICIPANT CHARACTERISTICS

Nine participants in the study were central office supervisors/directors and nine were building level principals in a small rural school system in North Carolina. Information gathered in the Rater Information forms and in the individual interviews was used to construct the participant profile.

The central office personnel had received training in using the North Carolina Teacher Performance Appraisal System (NCTPAS), but were not responsible for specific teacher evaluations at the school level on a regular basis.
These people did observations as requested or needed to assist the principals in the state-mandated annual evaluations of teaching personnel, or to provide additional help for teachers. Supervisors/directors had conducted from 0 to 20 formal observations using the evaluation system the previous year.

Areas of support services offered by these nine people at the time of the workshop included personnel, federal programs, exceptional children's program, elementary supervision, curriculum/instruction, vocational/secondary supervision, instructional materials/testing supervision, and science/math/computer coordination. Experience in the current positions ranged from one to ten years. Background experiences of the central office personnel included teaching at the elementary and secondary levels, principalships, general supervision, vocational teaching and supervision, media coordination, and assistant superintendent/superintendent responsibilities. Total education experience ranged from eight to 31 years.

The building level administrators included two high school principals with teaching staffs of 36 and 52, two middle school principals with teaching staffs of 26 and 37, and four elementary principals with teaching staffs of 15,
21, 36, and 37 members. The principals were responsible for observing and evaluating teachers using the state-mandated Teacher Performance Appraisal Instrument (TPAI) for all instructional personnel at their buildings for annual performance appraisal. Principals had conducted from 28 to 111 formal observations using the instrument the previous year. Years experience in the current positions ranged from one to 18 years. The principals had background experiences teaching various elementary and secondary subjects, vocational education, media, health and physical education, and early childhood. Total years in education ranged from 17 to 31 years.

Formal education and training for all participants consisted of a minimum of a master's degree, with nine having advanced degrees and two having doctorate degrees. Specialized training for administrators and supervisors is offered periodically by the North Carolina Department of Public Instruction (DPI), the Local Education Agency (LEA), and independent consulting firms. All participants had been trained in the North Carolina Effective Teaching Training Program and the North Carolina Teacher Performance Appraisal System. Nine people had training in the North Carolina Mentor/Support Team Training component. The two-day Department of Public Instruction Professional Development
Plan workshop had been attended by two people. Other professional training programs which had been attended by the participants included "Assertive Discipline" (5), "Clinical Supervision" (12), "TESA - Teacher Expectation and Student Achievement" (4), "COET - Classroom Organization for Effective Teaching" (7), and "Instructional Presentation - Madeline Hunter, SIP, STAR, or Six-step Lesson" (10).

RESPONSES TO INDIVIDUAL LESSONS

The possible ratings for the teaching behaviors observed in each tape are: Superior (6), Well Above Standard (5), Above Standard (4), At Standard (3), Below Standard (2), and Unsatisfactory (1). The five teaching behaviors (functions) rated are: (1) Management of Instructional Time, (2) Management of Student Behavior, (3) Instructional Presentation, (4) Instructional Monitoring of Student Performance, and (5) Instructional Feedback. After the videotaped lesson was observed using the Formative Observation Data Instrument (FODI) to make notes and the Formative Observation Data Analysis (FODA) to summarize the notes, a rating was assigned to each function. Participants rated the functions individually. Participants then met in small groups and arrived at a consensus rating for each
function after group discussion. These consensus team ratings are presented with the report for each tape. The data collected for each of the videotapes are presented in chronological order, as the tapes were viewed in the rating activity workshop. The videotaped lessons had not been normed previously with a single accurate rating for each function. The consensus process described above is the approved way to arrive at an accurate rating, according to the process adopted by the North Carolina Department of Public Instruction for use with the North Carolina teacher evaluation instrument.

DESCRIPTIONS OF VIDEOTAPED LESSONS AND RATINGS

TAPE 1: Description and Ratings

A primary reading class is conducted by a white female teacher in her early 30s. A group of seven children is led in a lesson to identify specific words and to sound out phonetic symbols. For two activities, the children and the teacher sit on the floor. Word cards are manipulated by the teacher and the students while on the floor. Students move to the front of the group and stand next to the teacher to respond to questions. Students move to different areas on
the floor and to individual desks for independent activities in writing and reading. Movement of the students is orderly. Other children in the class are monitored by a teacher assistant. The taped lesson lasts 33 minutes.

In the assignment of ratings by "All Participants" including principals and central office administrators, 14 evaluators assessed the lesson. (See Tables 1 and 2.) A mean of 3.57 and a median score of 3.5 were obtained. A standard deviation (SD) of 0.65 is indicative of a small variation in scores. Standard deviations of .70 and lower are considered to indicate close agreement using the six point rating scale of the North Carolina evaluation system. Higher standard deviations result in the assignment of an adjacent rating when using whole numbers as are used with the teacher evaluation instrument. Function 2, Management of Student Behavior, had a slightly higher variation with a standard deviation of 0.80, a mean of 2.79 and a median of 3. In the third function, Instructional Presentation, the one score of 5, Well Above Standard, is beyond the one-point variation that is acceptable for the NCTPAS, if we accept the median of 3 as a valid rating. Function 4, Monitoring
## TAPE 1: MEANS AND STANDARD DEVIATIONS FOR FUNCTION RATINGS

<table>
<thead>
<tr>
<th>FUNCTIONS</th>
<th>AP*</th>
<th>PR*</th>
<th>CO*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>1</td>
<td>3.57</td>
<td>.65</td>
<td>4.0</td>
</tr>
<tr>
<td>2</td>
<td>2.79</td>
<td>.80</td>
<td>3.40</td>
</tr>
<tr>
<td>3</td>
<td>2.64</td>
<td>.93</td>
<td>3.20</td>
</tr>
<tr>
<td>4</td>
<td>3.0</td>
<td>.78</td>
<td>3.40</td>
</tr>
<tr>
<td>5</td>
<td>3.64</td>
<td>.93</td>
<td>3.80</td>
</tr>
</tbody>
</table>

**Note.** AP = All Participants (N=14); PR = Principals (N=5); CO = Central Office Administrators (N=9).

FUNCTIONS = 1) Management of Instructional Time; 2) Management of Student Behavior; 3) Instructional Presentation; 4) Monitoring of Student Performance; 5) Instructional Feedback.
Table 2

INDIVIDUAL RATINGS FOR TAPE 1

<table>
<thead>
<tr>
<th>RATERS</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>P2</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>P3</td>
<td>5</td>
<td>3</td>
<td>5</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>P4</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>P5 *</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P6 *</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P7</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>P8 *</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P9 *</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>C2</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>C3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>C4</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>C5</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>C6</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>C7</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>C8</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>C9</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

Note.  P = Principals; C = Central Office Administrators.

FUNCTIONS = 1) Management of Instructional Time; 2) Management of Student Behavior; 3) Instructional Presentation; 4) Monitoring of Student Performance; 5) Instructional Feedback.

RATINGS: 6 = Superior; 5 = Well Above Standard; 4 = Above Standard; 3 = At Standard; 2 = Below Standard; 1 = Unsatisfactory.

* = Participant did not rate this lesson.
of Student Performance, had a mean of 3.0 and a median of 3. Closeness of the mean and the median are desirable in ratings from a group; however, the span in individual ratings from 2, Below Standard, to 5, Well Above Standard, is broader than acceptable. Function 5, Instructional Feedback, had ratings which ranged from 3, At Standard, to 6, Superior, a wider range than considered acceptable. The mean for Function 5 was 3.64 and the median was 3.5. The standard deviation (SD), the measurement of the variation of all ratings, consisted of a single standard deviation of 0.70 or below, a small deviation, and four standard deviations which were above 0.70. Only Function 1, Management of Instructional Time, had a standard deviation which would be considered a small variation (SD 0.65), indicating close agreement among raters. The other four functions had wider variations. Each of the three times that a rating exceeded the one-point tolerance from most of the other ratings, a principal had awarded the ratings.

The ratings recorded by the group of "Principals" had a mean of 4.0 and a median of 4 for Function 1. Function 2, Management of Student Behavior, projected a standard deviation of 0.55, indicative of close agreement. For Instructional Presentation, Function 3, principals’ ratings
calculated a standard deviation of 1.30, a high degree of variation. The range of four points from a 2, Below Standard, to a 5, Well Above Standard, was too wide to be considered acceptable for a single function rating. "Principals" rated Function 4, Instructional Monitoring of Student Performance, with a standard deviation of 0.89. The rating of 5, Well Above Standard, is more than one point higher than the other scores of 3, At Standard, the most common rating. Function 5, Instructional Feedback, had a standard deviation of 1.30, representing a wide variation. For this tape, two of the three times that a score was beyond the one-point tolerance, the scores were given by the same principal (P3). The third instance was the rating of another principal (P4). In only one of the five functions, "Principals" had a small variation (less than .70) of ratings as measured by the standard deviation.

The nine "Central Office Administrators" rated Tape 1 functions within a closer range than did "Principals." For Function 1, Management of Instructional Time, "Central Office Administrators" scored a standard deviation of 0.50, a small variation. The standard deviation for Function 2 was 0.73, slightly higher than the first function variation. Function 3 had a small variation of 0.50. The standard deviation of 0.67 for Function 4 was the third function to
have a small variation for the "Central Office Administrators" ratings. Instructional Feedback, Function 5 (SD 0.73) had greater variation. For this first videotaped lesson, "Central Office Administrators" were close in their ratings on three of the five functions, as measured by the standard deviation.

Tape 1 was rated by individuals only. No team ratings were conducted for this tape rating. This same tape was rated again after all seven tapes were rated, so that comparison of individual administrators' agreement in ratings could be made. It is labeled Tape 8 for the final tape ratings.

TAPE 2: Description and Ratings

A twelfth grade English class of 30 students is taught a lesson on outlining by a white male teacher in his early 40s. The first 18 minutes is total-group focus with the teacher identifying the parts of an outline with teacher-solicited input from the students who are sitting in rows of desks facing the teacher. During the last portion of the class, the students are divided into small groups of four or five students to practice outlining a specific topic assigned by the teacher. Students move their desks to form
the five small discussion circles. The students talk to one another in the small groups without prodding. The class lasts 33 minutes.

Tape 2 was viewed and rated by a total of 13 participants. (See Tables 3 and 4.) Results for "All Participants" reveal a mean of 4.69 and a median of 5 for Function 1. The standard deviation of 0.63 denote a small variation in the scores. In Function 2 the standard deviation is 0.64, also a small variation. The mean of 5.08 is close to the median of 5. The third function, Instructional Presentation, had the same mean and median as Function 2, but the standard deviation was slightly greater, at 0.76. Function 4, also with the median score of 5, had a mean of 5.15 and a standard deviation of 0.80. Even though this function had a higher mean, the variation across raters was greater also. Function 5 had the greatest variation as indicated by the standard deviation of 0.95. The average score of 4.31 and the median of 4 were lower ratings for this function than for other functions in Tape 2.

The variation of scores by "Principals" on the scores for the first function is the smallest variation for the principals. The standard deviation of 0.45 is indicative of
Table 3

TAPE 2: MEANS AND STANDARD DEVIATIONS FOR FUNCTION RATINGS

PARTICIPANT GROUPS

<table>
<thead>
<tr>
<th>FUNCTIONS</th>
<th>AP*</th>
<th>PR*</th>
<th>CO*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>1</td>
<td>4.69</td>
<td>.63</td>
<td>4.80</td>
</tr>
<tr>
<td>2</td>
<td>5.08</td>
<td>.64</td>
<td>5.20</td>
</tr>
<tr>
<td>3</td>
<td>5.08</td>
<td>.76</td>
<td>5.20</td>
</tr>
<tr>
<td>4</td>
<td>5.15</td>
<td>.80</td>
<td>5.40</td>
</tr>
<tr>
<td>5</td>
<td>4.31</td>
<td>.95</td>
<td>4.40</td>
</tr>
</tbody>
</table>

Note. AP = All Participants (N=13); PR = Principals (N=5); CO = Central Office Administrators (N=8).
FUNCTIONS = 1) Management of Instructional Time; 2) Management of Student Behavior; 3) Instructional Presentation; 4) Monitoring of Student Performance; 5) Instructional Feedback.
Table 4

INDIVIDUAL RATINGS FOR TAPE 2

FUNCTIONS

<table>
<thead>
<tr>
<th>RATERS</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>P2</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>P3</td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>P4</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>P5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P6</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>P7</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>P8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1</td>
<td>5</td>
<td>6</td>
<td>5</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>C2</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>C3</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>C4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>C5</td>
<td>6</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>C6</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>C7</td>
<td>4</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>C8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C9</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3</td>
</tr>
</tbody>
</table>

Note. P = Principals; C = Central Office Administrators.
FUNCTIONS = 1) Management of Instructional Time; 2) Management of Student Behavior; 3) Instructional Presentation; 4) Monitoring of Student Performance; 5) Instructional Feedback.
RATINGS: 6 = Superior; 5 = Well Above Standard; 4 = Above Standard; 3 = At Standard; 2 = Below Standard; 1 = Unsatisfactory.
* = Participant did not rate this lesson.
the agreement in scoring. In Function 2, Management of Student Behavior, the mean of 5.20 and median of 5 are close. The standard deviation was 0.45, also a small variation as calculated for the scores of the five principals. In Function 3, the standard deviation of 0.84 indicates greater variation than the standard deviation for "All Participants." Function 4 had a standard deviation of 0.89 and Function 5 a standard deviation of 0.55. The variation of the last function was less than the total group standard deviation. Three of the five functions had small standard deviations of ratings for "Principals."

"Central Office Administrators" had wider variation across their scores for the functions in Tape 2 than did the total group of participants. The eight administrators in this group had standard deviations ranging from 0.74 to 1.16, all higher than the recommended 0.70 or below. The greatest deviation was for Function 5, Instructional Feedback, and the least deviation was for Function 1, Management of Instructional Time.

After rating a taped lesson individually, participants joined together to form teams to reach consensus ratings for the five functions in each lesson. The teams were made up of balanced representation from the principals and the
central office administrators. Teams were formed by drawing names from the two groups, principals and central office administrators. The first tape to be rated by teams was Tape 2. Two participants formed a team and reached a consensus score for each function. For the third tape, three-member teams were formed. Four-member teams rated Tape 4. Tapes 5, 6, and 7 were rated by teams also. These rating activities were observed by a non-participant in the rating activity. The observations are reported with the report for each tape.

The seven teams that rated Tape 2, the lesson on "Outlining" for twelfth grade students, rated all functions At Standard or higher. For Functions 1, 2, 3, and 4, all ratings were Above Standard or higher (4, 5, or 6). For Instructional Feedback, Function 5, two teams scored At Standard (3), one team scored Above Standard (4), and four teams scored Well Above Standard (5). The mean rating for each function for the seven teams is presented with the mean rating by all participants' individual scores (Table 5).

Strengths identified by the teams included the following: Instructional Presentation; "all areas"; "summary of material presented was consistent"; students were on task and organized; "All areas excellent";
### Table 5

**TAPE 2: RATINGS BY TWO-MEMBER TEAMS**

**FUNCTIONS***

<table>
<thead>
<tr>
<th>TEAMS</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>B</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>C</td>
<td>5</td>
<td>6</td>
<td>5</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>D</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>E</td>
<td>6</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>F</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>G</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>3</td>
</tr>
</tbody>
</table>

**MEANS** 4.85 5.14 5.28 5.28 4.28

**AP MEANS** 4.69 5.08 5.08 5.15 4.31

*Note. FUNCTIONS = 1) Management of Instructional Time; 2) Management of Student Behavior; 3) Instructional Presentation; 4) Monitoring of Student Performance; 5) Instructional Feedback. AP MEANS = Mean scores derived from all participants' individual ratings for the tape.*
organized, procedures set; step-by-step instructions, examples, monitoring, varied activities, timed appropriately. Comments on areas that need improvement included: feedback to individual students; sustaining feedback; arrange flexible seating; "Encourage students to speak loudly when presenting"; used "OK" and "now" many times. Four teams listed Instructional Feedback as the most important function to work with the teacher to improve. Instructional Presentation, Managing Student Behavior, and Monitoring of Student Performance were listed one time each.

In reaching consensus, the two-member teams had changes in individual ratings 21 times out of a possible 75 scores. A change in rating was recorded each time an individual score was lower or higher than the final team rating for the function. These numbers indicate which direction the rater had to change to reach consensus agreement. A change was recorded as "Up" when the evaluator's original rating was lower than the team rating. For this tape, 18 ratings were raised. It was recorded as "Down" when the original rating was higher than the team rating. Only three ratings were lowered during consensus rating (Table 6). To change six times as many ratings "Up" versus "Down" is more extreme than the changes on other tapes that followed. During the follow-up interviews, participants were asked if they
Table 6

NUMBER OF TIMES AND THE DIRECTION THAT PARTICIPANTS CHANGED THEIR SCORES TO REACH GROUP CONSENSUS

<table>
<thead>
<tr>
<th>PARTICIPANTS</th>
<th>Tape 2</th>
<th>Tape 3</th>
<th>Tape 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Up</td>
<td>Dn</td>
<td>Tot</td>
</tr>
<tr>
<td>P1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>P2</td>
<td>3</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>P3</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>P4</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>P5</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>P6</td>
<td>4</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>P7</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>P8</td>
<td></td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>P9</td>
<td></td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>C1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C2</td>
<td>5</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>C3</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C4</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C5</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>C6</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C7</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>C8</td>
<td></td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>C9</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

Totals 18 3 21 19 9 28 10 13 23

Note. "Up" and "Dn" (Down) indicate whether the participants changed their scores to a higher rating (Up) or to a lower rating (Dn).
Participants did not rate the tapes which show no ratings.
Table 6
(Continued)

NUMBER OF TIMES AND THE DIRECTION THAT PARTICIPANTS
CHANGED THEIR SCORES TO REACH GROUP CONSENSUS

<table>
<thead>
<tr>
<th>PARTICIPANTS</th>
<th>Tape 5</th>
<th></th>
<th></th>
<th>Tape 6</th>
<th></th>
<th></th>
<th></th>
<th>Tape 7</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Up</td>
<td>Dn</td>
<td>Tot</td>
<td>Up</td>
<td>Dn</td>
<td>Tot</td>
<td></td>
<td>Up</td>
<td>Dn</td>
<td>Tot</td>
<td></td>
</tr>
<tr>
<td>P1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P4</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P5</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P6</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P7</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C3</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C4</td>
<td>5</td>
<td>0</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C5</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C6</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C8</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C9</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Totals</td>
<td>8</td>
<td>12</td>
<td>20</td>
<td>13</td>
<td>13</td>
<td>26</td>
<td>4</td>
<td>15</td>
<td>19</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. "Up" and "Dn" (Down) indicate whether the participant changed their scores to a higher rating (Up) or to a lower rating (Dn).
Participants did not rate the tapes which show no ratings.
changed their ratings. Comparison of their responses to the data on actual changes in ratings is made.

TAPE 3: Description and Ratings

A literature lesson on the story "Rikki Tikki Tavi" is taught to a class of 28 seventh grade students. The white male in his early 40s, utilizes a whole-group discussion technique, patterned with question-and-answer interchanges between teacher and students. The pupils volunteer answers and questions without prodding, remaining in their seats throughout the lesson. The teacher moves from the front of the room, walking to desks and down rows during the discussion. Each student and the teacher have a textbook which is used to follow the story as the discussion takes place. Student behavior is appropriate for the class. The 37-minute lesson is the first class discussion of the story.

The total group ratings for Tape 3 were more varied than on any other lesson ratings. (See Tables 7 and 8.) All functions had standard deviations at 0.76 or above. Two functions, 2 and 4, varied at the 1.02 level. In all functions, the mean and median were within .57 points of one another.
Table 7

TAPE 3: MEANS AND STANDARD DEVIATIONS FOR FUNCTION RATINGS

<table>
<thead>
<tr>
<th>PARTICIPANT GROUPS</th>
<th>AP*</th>
<th>PR*</th>
<th>CO*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>FUNCTIONS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>4.79</td>
<td>.97</td>
<td>4.50</td>
</tr>
<tr>
<td>2</td>
<td>4.57</td>
<td>1.02</td>
<td>4.50</td>
</tr>
<tr>
<td>3</td>
<td>5.14</td>
<td>.86</td>
<td>4.83</td>
</tr>
<tr>
<td>4</td>
<td>4.43</td>
<td>1.02</td>
<td>4.50</td>
</tr>
<tr>
<td>5</td>
<td>4.50</td>
<td>.76</td>
<td>4.67</td>
</tr>
</tbody>
</table>

Note. AP = All Participants (N=14); PR = Principals (N=6); CO = Central Office Administrators (N=8).
FUNCTIONS = 1) Management of Instructional Time; 2) Management of Student Behavior; 3) Instructional Presentation; 4) Monitoring of Student Performance; 5) Instructional Feedback.
Table 8

INDIVIDUAL RATINGS FOR TAPE 3

<table>
<thead>
<tr>
<th>RATERS</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>P2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P4</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>P5</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>P6</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>P7</td>
<td>5</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>P8</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>P9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1</td>
<td>6</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>C2</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>C3</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>C4</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>C5</td>
<td>4</td>
<td>3</td>
<td>5</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>C6</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>C7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C8</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>C9</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>4</td>
</tr>
</tbody>
</table>

Note. P = Principals; C = Central Office Administrators.
FUNCTIONS = 1) Management of Instructional Time;
2) Management of Student Behavior; 3) Instructional Presentation; 4) Monitoring of Student Performance;
5) Instructional Feedback.
RATINGS: 6 = Superior; 5 = Well Above Standard;
4 = Above Standard; 3 = At Standard; 2) Below Standard; 1 = Unsatisfactory.
* = Participant did not rate this lesson.
"Principals' ratings varied greatly in four of the five functions. Functions 1, 2, 3, and 4 had standard deviations greater than 1.0. On Function 5 there was closer agreement. The 0.52 standard deviation indicated little variation in the ratings among the principals. The means and medians for the functions were close. In Function 4, the mean and the median were both 4.5.

"Central Office Administrators" had a small variation in their ratings for Function 3. The smallest standard deviation for these administrators was for the same function (3) which had the second largest standard deviation for "Principals." The other four functions had standard deviations higher than 0.70, with the highest being 1.06 for Function 4. The mean and the median coincided at 5, Well Above Standard, for Function 1.

The three-member teams reached consensus on Tape 3, the seventh grade literature lesson. Five teams were formed for this activity. Two functions had ratings of At Standard (3) by one team (see Table 9). All other ratings were Above Standard (4), Well Above Standard (5), or Superior (6), by all teams. One team rated Superior for every function. The
Table 9

TAPE 3: RATINGS BY THREE-MEMBER TEAMS

<table>
<thead>
<tr>
<th>FUNCTIONS*</th>
<th>TEAMS</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>6</td>
<td>5</td>
<td>6</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>MEANS</td>
<td>5.00</td>
<td>4.80</td>
<td>5.20</td>
<td>4.60</td>
<td>4.70</td>
<td></td>
</tr>
</tbody>
</table>

AP MEANS* 4.79 4.57 5.14 4.43 4.50

Note: FUNCTIONS=1) Management of Instructional Time; 2) Management of Student Behavior; 3) Instructional Presentation; 4) Monitoring of Student Performance; 5) Instructional Feedback.
AP MEANS= Mean scores derived from all participants' individual ratings for the tape.
teams' mean rating for each function is higher in every case than the mean rating for individual scores of all participants.

The teams recorded strengths of the lesson as follows: feedback, level of communication, and examples; teacher mastery of subject area; teacher-student rapport; "Appropriate real life examples." Areas that need improvement were listed: monitoring pattern - move around room; "Draw more detail from student responses"; monitoring. Managing Instructional Time, Instructional Presentation, and Monitoring Instruction, were listed as function areas that needed work. A suggestion was given for the teacher to develop questioning techniques so that students would give more complete answers when responding. Also addressed was the teacher's excessive use of the term "Go ahead."

There were a total of 28 changes in individual scores to reach consensus on the possible 75 scores. Frequency of changes by an individual evaluator ranged from a high of four times to a low of zero times. There were only 9 instances of raters' agreeing to lower scores, and 19 instances of raters' changing to higher scores for consensus.
TAPE 4: Description and Ratings

A black male in his mid-30s teaches a seventh grade math class. Probability is the topic of the 33-minute lesson. Total-group instruction with the teacher standing at the front of the room of 27 students is utilized. The teacher turns his back to the class and works at the chalkboard at times. Although students remain in their desks, off-task behavior is evident for the entire lesson. Student-initiated comments prevail in the class. The teacher answers his own questions when students do not respond.

Ratings from "All Participants" were lower than for any other lessons. (See Tables 10 and 11.) The mean scores ranged from 1.27 for Function 2 to a high of 1.73 for Functions 1 and 3. The standard deviations indicate the best agreement for the ratings among the evaluators. Functions 1, 2, and 3 had standard deviations of 0.46, indicating very close agreement. Functions 4 and 5 had standard deviations of 0.49 and 0.63, both small variations. All 15 participants gave ratings of 1, Unsatisfactory, or 2, Below Standard, in all but one instance when a principal gave a 3, At Standard, on Function 5. The teacher's performance in each of the five function areas was scored.
Table 10

TAPE 4: MEANS AND STANDARD DEVIATIONS FOR FUNCTION RATINGS

PARTICIPANT GROUPS

<table>
<thead>
<tr>
<th>FUNCTIONS</th>
<th>AP*</th>
<th>PR*</th>
<th>CO*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>1</td>
<td>1.73</td>
<td>.46</td>
<td>1.57</td>
</tr>
<tr>
<td>2</td>
<td>1.13</td>
<td>.46</td>
<td>1.14</td>
</tr>
<tr>
<td>3</td>
<td>1.73</td>
<td>.46</td>
<td>1.57</td>
</tr>
<tr>
<td>4</td>
<td>1.33</td>
<td>.49</td>
<td>1.29</td>
</tr>
<tr>
<td>5</td>
<td>1.63</td>
<td>.63</td>
<td>1.43</td>
</tr>
</tbody>
</table>

Note. AP = All Participants (N=15); PR = Principals (N=7); CO = Central Office Administrators (N=8).
FUNCTIONS = 1) Management of Instructional Time; 2) Management of Student Behavior; 3) Instructional Presentation; 4) Monitoring of Student Performance; 5) Instructional Feedback.
### Table 11

**INDIVIDUAL RATINGS FOR TAPE 4**

#### FUNCTIONS

<table>
<thead>
<tr>
<th>RATERS</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>P2 *</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>P4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>P5</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>P6</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>P7</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>P8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>P9 *</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>C2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>C3</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>C4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>C5</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>C6</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>C7 *</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C8</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>C9</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

**Note.**  
*P = Principals; C = Central Office Administrators.*  
**FUNCTIONS** = 1) Management of Instructional Time; 2) Management of Student Behavior; 3) Instructional Presentation; 4) Monitoring of Student Performance; 5) Instructional Feedback.  
**RATINGS:** 6 = Superior; 5 = Well Above Standard; 4 = Above Standard; 3 = At Standard; 2 = Below Standard.  
* = Participant did not rate this lesson.
below an acceptable standard of teaching, with the one exception.

"Principals" ratings, based on the ratings from seven principals, had close variation levels on the first four functions. Only on Instructional Feedback did the standard deviation go above the 0.70 level (SD 0.79). The closest agreement was on Function 2, Management of Student Behavior, for a standard deviation of 0.38. Only one principal rated a 2, Below Standard, while all other principals rated a 1, Unsatisfactory. The function with the standard deviation of 0.79 had one score of a 3, At Standard, with all other ratings 2, Below Standard, or 1, Unsatisfactory.

"Central Office Administrators" agreed closely on all five functions. On Management of Instructional Time (1) and Instructional Presentation (3) the standard deviation was 0.35. In each function only one administrator gave a 1, Unsatisfactory, in comparison to a 2, Below Standard, given by all others. Function 5, which had the least agreement among principals, had a standard deviation of 0.46 for "Central Office Administrators." The other two functions had standard deviations of 0.52, well within the range of acceptable variation. The median scores for Functions 1 through 4 were the same for "All Participants,"
"Principals," and "Central Office Administrators." On Function 5 the "Principals" median score was one point lower than the other two groups' scores.

In viewing and assessing the seventh grade math lesson in Tape 4, the four-member teams agreed that the teaching performance was less than acceptable. All four of the teams rated every function as Below Standard (2) or Unsatisfactory (1). Mean ratings for the teams collectively were very close to the mean ratings for all individual participants (Table 12).

The only strength listed by the teams was "knowledge of the subject matter." Designating specific areas that need improvement was difficult for the teams. "All Five!" was written on one report. Other suggestions were: observe master teacher, have follow-up conference with him, and develop a professional development plan with the principal; monitor student behavior and address students' inappropriate behavior; make assignments clear; reinforce stated rules for answering questions; monitor student time-on-task; needs eye contact with students; knew subject but could not teach. The frustration felt by one observer was stated on the report, "The only strength I saw was that he showed up for class!"
### Table 12

**TAPE 4: RATINGS BY FOUR-MEMBER TEAMS**

<table>
<thead>
<tr>
<th>FUNCTIONS*</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>TEAMS</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>A</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>C</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>D</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>MEANS</td>
<td>1.50</td>
<td>1.13</td>
<td>1.75</td>
<td>1.25</td>
<td>1.75</td>
</tr>
<tr>
<td>AP MEANS*</td>
<td>1.73</td>
<td>1.13</td>
<td>1.73</td>
<td>1.33</td>
<td>1.63</td>
</tr>
</tbody>
</table>

*Note.* FUNCTIONS=1) Management of Instructional Time; 2) Management of Student Behavior; 3) Instructional Presentation; 4) Monitoring of Student Performance; 5) Instructional Feedback. AP MEANS= Mean scores derived from all participants' individual ratings for the tape.
All teams agreed that Instructional Presentation, Function 3, was the function which needed the most attention in order for the teacher to improve. Suggested strategies included: use side view rather than back to class; organize lesson plan including objective; improve diction; focus on total class; check individual performance of all students through careful monitoring; circulate in class; observe another teacher who is very strong in management skills; set up rules for class participation. Additional suggestions were made which dealt with similar problems.

To reach consensus, 23 changes in scores were made within the 75 possible ratings. Of the 23 changes, 10 were changed from lower to higher scores and 13 were changed from higher to lower ratings. The most changes any one rater made were four, the least was zero.

TAPE 5: Description and Ratings

The literature lesson about the story "Monsters Are Due" is taught to a class of 28 seventh grade students by a white male in his early 40s. The 35-minute lesson is conducted using a whole-class question-and-answer approach to review the story and the meaning of the literary plot, significance, and texture. Students sit in desks in rows
facing the front of the class and respond to the teacher by raising their hands before being recognized. There is constant dialogue between the teacher and students. The teacher stays at the front of the class for most of the presentation. The chalkboard is used for vocabulary words that students are assigned. Students are attentive to the teacher.

The fifteen participants who rated this lesson varied in their ratings on each function. (See Tables 13 and 14.) In Function 1, the standard deviation of 0.59 is the only one which indicates close agreement. The other four functions had standard deviations ranging from 0.85 to 1.06. The means and medians for all functions were close, but Function 5, *Instructional Feedback*, had the same score of 4, Above Standard, for each. In Functions 2 and 3, there was one rating in each which was outside the tolerance limit of one point of the mean. Function 4 had two ratings outside the limit.

The ratings for "Principals" revealed higher variability in the scores. Every function had a standard deviation higher than for the same function for "All Participants." Only one standard deviation was closer than the 0.70 limit: Function 1 had 0.64, indicating closer
Table 13

TAPE 5: MEANS AND STANDARD DEVIATIONS FOR FUNCTION RATINGS

PARTICIPANT GROUPS

<table>
<thead>
<tr>
<th>FUNCTIONS</th>
<th>AP*</th>
<th>PR*</th>
<th>CO*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>1</td>
<td>3.80</td>
<td>.59</td>
<td>4.13</td>
</tr>
<tr>
<td>2</td>
<td>3.93</td>
<td>.88</td>
<td>4.13</td>
</tr>
<tr>
<td>3</td>
<td>4.06</td>
<td>.88</td>
<td>4.25</td>
</tr>
<tr>
<td>4</td>
<td>3.40</td>
<td>1.06</td>
<td>3.75</td>
</tr>
<tr>
<td>5</td>
<td>4.0</td>
<td>.85</td>
<td>4.0</td>
</tr>
</tbody>
</table>

Note. AP = All Participants (N=15); PR = Principals (N=8); CO = Central Office Administrators (N=7).
FUNCTIONS = 1) Management of Instructional Time; 2) Management of Student Behavior; 3) Instructional Presentation; 4) Monitoring of Student Performance; 5) Instructional Feedback.
<table>
<thead>
<tr>
<th>RATERS</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>P2</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>P3</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>P4</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>P5</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>P6</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>P7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P8</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>P9</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>C1</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>C2</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>C3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>C4</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>C5</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>C6</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>C7</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>C8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. P = Principals; C = Central Office Administrators.  
FUNCTIONS = 1) Management of Instructional Time; 2) Management of Student Behavior; 3) Instructional Presentation; 4) Monitoring of Student Performance; 5) Instructional Feedback.  
* = Participant did not rate this lesson.
agreement. Function 4 had the widest variation, 1.28. The mean and median for Function 5 were the same score of 4, Above Standard, as were the mean and median for "All Participants" in that function. In other functions the means and medians were close but not the same. In the ratings for Management of Student Behavior (2) and Instructional Presentation (3) one principal rated higher that the other seven principals. In Instructional Monitoring (4), a different principal had the singleton Superior rating of 6. On that same function, a low score of 2, Below Standard, was recorded by an individual, accounting for a wide spread in ratings.

This tape of a seventh grade literature lesson, was rated by seven two-member teams. Function 1, Management of Instructional Time, was scored Above Standard (4) by five of the seven teams. Single scores of At Standard (3) and Well Above Standard (5) were reported for the other teams. Management of Student Behavior, Function 2, and Instructional Presentation, Function 3, received ratings of At Standard (3), Above Standard (4), and Well Above Standard (5). Function 4, Monitoring Instruction, was the only function which received a score of Below Standard (2). This function also received a score of Superior (6) by one team.
These two ratings represent diversity in judgments of the quality of the teaching performance. Other teams scored the function At Standard (3) and Above Standard (4). Instructional Feedback, Function 5, received scores ranging from At Standard (3) to Well Above Standard (5). The mean scores for each function for all individual participants were within .21 points of the corresponding mean scores for the teams (Table 15).

Time-on-task, teacher preparation, and the way students responded were mentioned as strengths by one team. The fact that the teacher knew his subject matter well was brought out by four teams. The teacher's explanations on background material were complimented by one team. Student responses were noted by three teams, as indicated in a discussion of the manner in which the teacher's questioning and probing techniques obtained answers from students. The excellent quality of teacher feedback for incorrect answers was noted by two teams. Outside examples "to enhance the presentation" and student-teacher rapport were listed as strengths. Management of Student Behavior and Monitoring of Student Performance were strong practices.

Areas that need improvement were the lesson technique and the method of instruction. Four teams listed the
Table 15

TAPE 5: RATINGS BY TWO-MEMBER TEAMS

<table>
<thead>
<tr>
<th>TEAMS</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>B</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>C</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>D</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>E</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>F</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>G</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>MEANS</td>
<td>4.0</td>
<td>3.85</td>
<td>3.85</td>
<td>3.28</td>
<td>3.85</td>
</tr>
</tbody>
</table>

AP MEANS* 3.80 3.93 4.06 3.40 4.0

Note. FUNCTIONS=1) Management of Instructional Time; 2) Management of Student Behavior; 3) Instructional Presentation; 4) Monitoring of Student Performance; 5) Instructional Feedback. AP MEANS= Mean scores derived from all participants' individual ratings for the tape.
dependence on a lecture method as too boring for an entire class period. The incorporation of varied teaching techniques such as dramatization was suggested. Monitoring the instruction of all students, especially those on the left side of the room, was recommended. The teacher needed to circulate among all the students, not just those on one side of the room.

Reaching consensus was not as difficult for some teams as for others. Out of a possible 75 ratings, 20 scores were changed to reach consensus. Scores for 12 items were lowered and 8 items were raised to reach agreement. Five participants did not make any changes in their ratings.

The interactions of the team members noted by the observers demonstrated eagerness of the raters to score the teaching performance fairly. Discussion between the two members included many examples and reasons for the individual ratings. When discussing Monitoring of Student Performance, Function 4, one rater stated, "I gave a 2 [Below Standard] because the teacher didn't circulate. She didn't get away from her desk." The other team member countered with two other reasons why he gave a 3, At Standard. After the explanation, the first rater easily agreed to raise her score to a 3. "I don't mind going up,"
she said. "You have given some good reasons and I can see your point." Interchanges such as this one took place in each of the seven groups.

In two instances, a team member agreed to a score to reach consensus, but expressed reluctance in doing so. "I'm comfortable with Above Standard (4) lesson, but not with Well Above. I believe it's a 4 but I'm willing to go to a 5 for consensus," the administrator asserted. On another function, the same person gave a score lower than the other team member, saying, "I gave him a 3, but could go with a 4." The second team member had given a 6, Superior, a difference of three points. After each rater gave reasons and read examples of behaviors from written notes, they both agreed to change. The final decision of a 5, Well Above Standard, was made after reading the descriptions of the scores on the rating scale. Once the decision to give a consensus rating of 5 was made, the two team members verbalized reasons for selecting the 5. The final comment made by one evaluator was, "Quality Job, but I rated At Standard; I see now he was higher after our discussion."

An administrator who was not so sure of the rating he had assigned, was convinced by the other team member to change from a 5, Well Above Standard, to a 4, Above
Standard. Both people gave their reasons and examples on which they based their individual scores. As the higher rater gave the five reasons for the score of 5, Well Above Standard, the lower rater began shaking his head affirmatively, then interrupted by explaining, "I may have just lost that part. Since you are explaining that way, I can go along with that." As he erased his number on his paper, the higher rater tried to stop him, saying, "Don't change yours. Let's go on to another function and come back if you aren't sure." The higher rater then continued by giving another reason for the 5, in a more insistent voice. The lower rater gave in, then began to insist that a 5 was appropriate. This team made the same change for another function. The one rater who scored one point higher than the other, gave reasons and the lower rater, same one as in the previous instance, changed his score saying, "I just don't shoot too high, I guess." At times of disagreement, both team members looked back at their notes, then one of them would read the rating descriptions from the Teacher Performance Appraisal Instrument. The observer noted that the rater who convinced the other team member to change his score, talked more and led the discussion by directing questions to the other member. Both raters had thorough notes on the teaching episode.
Two other teams used similar procedures for reaching consensus. Each member talked about their observations, gave specific examples from their notes, then stated individual scores. With one team, leadership was established by one member picking up the rating forms and completing the data, asking questions to the other member. When there was disagreement, the person who was completing the forms, gave numerous reasons for her ratings, and the other team member then agreed with her. There was never a difference of more than one point with these two raters. The dominant team member rated by the instrument, using specific code numbers when referring to practices. The other rater often addressed a different function when giving reasons for rating a particular function. An example was his reference to "in-depth answers, using outside examples" when discussing Management of Instructional Time. The more precise rater quickly pointed out that those practices were not relevant for Instructional Time, but could be used in the discussion of Feedback. The two members on the last team changed scores to reach consensus after giving reasons for their scores. Each posed questions for clarification as points were made by the other member.
TAPE 6: Description and Ratings

A small group of kindergarten students are taught a reading lesson by a white female in her late 20s. The 31-minute lesson focuses on the oral reading of the story and on questions about the plot and characters. The students sit in chairs in a circle with the teacher next to a chart stand on which a calendar sits. The teacher begins the story by asking the students to stand at their seats and demonstrate how they would act out various emotions. After the completion of the story, the teacher leads the students in a review of the calendar, asking them to identify month, day, and year. The students remain in their chairs except for two of them who are asked to come to the front to identify calendar dates. Students exhibit orderly behavior.

The fifteen evaluators in the "All Participants" group had close agreement on all functions. (See Tables 16 and 17.) The standard deviations ranged from 0.46 to 0.63. The closest agreement was reached on Functions 4 and 5, Instructional Monitoring and Instructional Feedback. The ratings were mostly a 1, Unsatisfactory, or a 2, Below Standard. A rating of 3, At Standard, was given on Function 2 and on Function 3 by central office individuals.
Table 16

Tape 6: Means and Standard Deviations for Function Ratings

Participant Groups

<table>
<thead>
<tr>
<th>FUNCTIONS</th>
<th>AP* Mean</th>
<th>SD</th>
<th>PR* Mean</th>
<th>SD</th>
<th>CO* Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.60</td>
<td>.63</td>
<td>1.25</td>
<td>.46</td>
<td>2.0</td>
<td>.58</td>
</tr>
<tr>
<td>2</td>
<td>1.73</td>
<td>.59</td>
<td>1.50</td>
<td>.53</td>
<td>2.0</td>
<td>.58</td>
</tr>
<tr>
<td>3</td>
<td>1.33</td>
<td>.51</td>
<td>1.25</td>
<td>.46</td>
<td>1.57</td>
<td>.53</td>
</tr>
<tr>
<td>4</td>
<td>1.73</td>
<td>.46</td>
<td>1.63</td>
<td>.52</td>
<td>1.86</td>
<td>.38</td>
</tr>
<tr>
<td>5</td>
<td>1.73</td>
<td>.46</td>
<td>1.63</td>
<td>.52</td>
<td>1.86</td>
<td>.38</td>
</tr>
</tbody>
</table>

Note. AP = All Participants (N=15); PR = Principals (N=8); CO = Central Office Administrators (N=7). FUNCTIONS = 1) Management of Instructional Time; 2) Management of Student Behavior; 3) Instructional Presentation; 4) Monitoring of Student Performance; 5) Instructional Feedback.
Table 17

INDIVIDUAL RATINGS FOR TAPE 6

FUNCTIONS

<table>
<thead>
<tr>
<th>RATERS</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>P2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>P3</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>P4</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>P5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>P6</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>P7 *</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>P9</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>C1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>C2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>C3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>C4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>C5</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>C6</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>C7</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>C8 *</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C9 *</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note.  P = Principal;  C = Central Office Administrators.

FUNCTIONS = 1) Management of Instructional Time; 2) Management of Student Behavior; 3) Instructional Presentation; 4) Monitoring of Student Performance; 5) Instructional Feedback.

RATINGS: 6 = Superior; 5 = Well Above Standard; 4 = Above Standard; 3 = At Standard; 2 = Below Standard; 1 = Unsatisfactory.

* = Participant did not rate this lesson.
The median score was 2, Below Standard, for all functions except Function 3, which was a 1, Unsatisfactory. The means ranged from 1.40 to 1.73.

"Principals" had close agreement on all five functions. For this group, this is the only lesson in which close agreement was achieved for every teaching behavior. The eight principals scored Unsatisfactory (1) or Below Standard (2) for every function. For Management of Instructional Time (1) and Instructional Presentation (3), 0.46 was the standard deviation. The highest standard deviation was 0.53. No principal had all ratings at the Unsatisfactory or Below Standard level for every function.

The standard deviations for all functions for the "Central Office Administrators" were below 0.70. The most divergent ratings were for Functions 1 and 2 with standard deviations of 0.58. The closest scores were for Functions 4 and 5 with variations of 0.38. These two functions exhibited the closest scores with only one of the seven ratings different from the others. In the 35 possible ratings by the seven administrators, only two were a 3, At Standard. All others were Below Standard or Unsatisfactory.
This sixth tape of a kindergarten lesson, was rated by four teams with four members on each team. All teams rated every function below satisfactory. Scores of 2, Below Standard, and 1, Unsatisfactory, were given in every case. Function 1, Management of Instructional Time, was given a score of 2, Below Standard, by every team. The second function, Management of Student Behavior, received scores of Below Standard (2) from two teams, and Unsatisfactory (1) from two teams. The same breakdown was present for Functions 3 and 4. Function 5, Instructional Feedback, received three ratings of Below Standard (2) and one rating of Unsatisfactory (1). Mean scores for the teams were close to the mean scores of all participants. All means were 2.0 or below (Table 18).

Strengths listed by teams were the clearly stated objective, positive feedback, and the "teacher's obvious love, patience, and warmth for the children." Comments were made in each instance that a strength was listed only because it would be difficult to find good practices in the lesson.

Areas that need improvement were organization and management, managing instructional time, managing student behavior, utilization of the classroom assistant, and "all
Table 18

TAPE 6: RATINGS BY FOUR-MEMBER TEAMS

<table>
<thead>
<tr>
<th>TEAMS</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>B</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>C</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>D</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>MEANS</td>
<td>2.0</td>
<td>1.50</td>
<td>1.50</td>
<td>1.50</td>
<td>1.75</td>
</tr>
</tbody>
</table>

AP MEANS* 1.60 1.73 1.33 1.73 1.73

Note. FUNCTIONS=1) Management of Instructional Time; 2) Management of Student Behavior; 3) Instructional Presentation; 4) Monitoring of Student Performance; 5) Instructional Feedback.

AP MEANS= Mean scores derived from all participants' individual ratings for the tape.
areas." Strategies to help the teacher included help with classroom management techniques, assertive discipline workshop, assistance with planning, observation of her own videotaped lesson, and watch another teacher teach a lesson.

In reaching consensus, the individual team members made changes in 26 of a possible 75 ratings. The changes from lower to higher scores accounted for one half of the total changes. The other 13 scores were changed from higher to lower ratings. Of the 15 evaluators, two had no changes.

The four-member teams demonstrated some changes in leadership from the two-member teams. Observers noted how the discussion began when the groups first sat down together in a team. With four people to reach consensus, there was more effort made to formalize the order of discussion. In the first four-member team, a principal who had been a follower in the two-member teams, took the lead by telling the others the procedures to use. "We need to go around and all speak, then get consensus," he began. The other team members spoke in turn, giving their individual ratings, then cited reasons and observations. With each function this procedure was followed, with the same principal calling out the final consensus rating. One rater gave higher ratings on two of the functions. After giving
reasons for her ratings, she said in an apologetic tone, "Mine is higher than theirs - I gave the benefit of the doubt. I've been there before." There was agreement on the areas of improvement for the teacher. Each participant contributed to the discussion, making an effort to clarify their ratings, giving reasons and examples of teaching behaviors. Comments such as the following were made by the raters: "With everything going on in the classroom, the presentation could not be good." "The teacher lost control." "Most of the feedback dealt with behavior. He used positive feedback, but it did not help." As one participant verbalized observations, others supported the speaker with nods, smiles, and words of acceptance such as "Yes," "You're right," "I agree."

One group of four had a discussion in which there was disagreement about the outcome of student learning as a result of the teacher's behavior. Three of the raters agreed that a 2, Below Standard, was an appropriate score for the first three functions. The fourth team member insisted that a 2 was too high a score. Reasons were cited by those who advocated a 2; additional reasons were cited by the person who rated a 1, Unsatisfactory. An energetic, sometimes heated discussion of the quality of the teacher's instruction and classroom practices ensued. Two members of
the team reached an impasse over the ratings. The team decided on the lower score of 1, Unsatisfactory, for the teacher after arguing and defending their differing views. All team members strongly supported their views, but continued listening to one another with smiles and some joking to lighten the intensity of the discussion. After the activity was over, the members referred to the good time they had trying to convince one another that their own score was correct. Reference was made to the fact that all agreed that the teacher was a poor teacher who should be rated less than standard, but the disagreement was over the degree of weakness. Whether she was "salvageable" and could be rated a 2, Below Standard, or was "hopeless" and deserved a 1, Unsatisfactory, was the point of contention.

A team with two principals and two central office administrators with experience at different grade levels shared their views on the teaching episodes. The people with the most experience at the elementary level expressed more agreement among themselves than did the high school-oriented raters. The high school administrator rated lower than the others. A principal kept the group on task and working toward consensus when the pace slowed, while a supervisor read the instrument and rating descriptions. One central office administrator continually pointed out the
potential of the teacher to improve, even with the problems in conducting the class. Trying to convince the raters of a score of Below Standard instead of Unsatisfactory, the administrator said, "I understand the disorganization, but I want to pull the potential out. There is a sign of light there!" After citing more reasons to support a particular view, the group finally reached consensus on each rating. Each person changed their individual ratings at some time to reach consensus.

Problems with observing and evaluating a kindergarten teacher when one's training, teaching experience, and administrative experience are with higher grade levels were discussed by the last team. No one on that team had teaching experience at the kindergarten level; however, the ratings for the team were in line with the other teams. A review of generic teaching practices and how the teacher exhibited them resulted from the concern with the kindergarten teacher. The teacher's relationship with the teacher assistant in the classroom was debated. The team decided that the teacher was scared to tell another adult what to do to assist with instruction. It was agreed that the function to highlight for improvement efforts was Instructional Presentation. Strategies to help the teacher in that area were listed.
"Euthanasia" is the topic of the ninth grade English class taught by a white female in her early 30s. The 20-minute lesson is taught using a total group method for the 26 students. An overhead projector is used to provide the only copy of a story that a student is asked to read aloud. Difficulty in reading the small print slows down the activity. Students sit in their desks in rows facing the front of the class. The teacher leads a discussion and questions the students from the front of the classroom. Student-initiated comments prevail during the lesson, often in the form of loud shouts. Off-task behavior is evident. The teacher asks questions and provides the answers herself in numerous cases. Incorrect information is given by the teacher in her explanation. During the last five minutes of the class, the students pull their desks together to form groups of three or four to work on a group assignment. The teacher moves from group to group, repeating instructions.

The ratings by "All Participants" were close for all functions in Tape 7. (See Tables 19 and 20.) Of the 14 ratings, all were a 2, Below Standard, for Function 2, Management of Student Behavior. That all evaluators rated the teacher 2, Below Standard,
Table 19

TAPE 7: MEANS AND STANDARD DEVIATIONS FOR FUNCTION RATINGS

PARTICIPANT GROUPS

<table>
<thead>
<tr>
<th>FUNCTIONS</th>
<th>AP* Mean</th>
<th>SD</th>
<th>PR* Mean</th>
<th>SD</th>
<th>CO* Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.50</td>
<td>.65</td>
<td>2.63</td>
<td>.74</td>
<td>2.33</td>
<td>.52</td>
</tr>
<tr>
<td>2</td>
<td>2.0</td>
<td>.00</td>
<td>2.0</td>
<td>.00</td>
<td>2.0</td>
<td>.00</td>
</tr>
<tr>
<td>3</td>
<td>2.35</td>
<td>.50</td>
<td>2.25</td>
<td>.46</td>
<td>2.50</td>
<td>.55</td>
</tr>
<tr>
<td>4</td>
<td>1.85</td>
<td>.36</td>
<td>1.88</td>
<td>.35</td>
<td>1.83</td>
<td>.41</td>
</tr>
<tr>
<td>5</td>
<td>2.64</td>
<td>.63</td>
<td>2.50</td>
<td>.53</td>
<td>2.83</td>
<td>.75</td>
</tr>
</tbody>
</table>

Note. AP = All Participants (N=14); PR = Principals (N=8); CO = Central Office Administrators (N=6).
FUNCTIONS = 1) Management of Instructional Time; 2) Management of Student Behavior; 3) Instructional Presentation; 4) Monitoring of Student Performance; 5) Instructional Feedback.
### Table 20

**INDIVIDUAL RATINGS FOR TAPE 7**

<table>
<thead>
<tr>
<th>RATERS</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>P2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>P3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>P4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>P5 *</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P6</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>P7</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>P8</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>P9</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>C1</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>C2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>C3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>C4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>C5</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>C6</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>C7 *</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C8 *</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C9 *</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. P = Principals; C = Central Office
Administrators.

**FUNCTIONS**

1) Management of Instructional Time;
2) Management of Student Behavior;
3) Instructional
4) Monitoring of Student Performance;
5) Instructional Feedback.

**RATINGS:** 6 = Superior; 5 = Well Above Standard;
4 = Above Standard; 3 = At Standard; 2 = Below
Standard; 1 = Unsatisfactory.

* = Participant did not rate this lesson.
showed agreement in the degree of weakness in this teacher's performance as evidenced by the standard deviation of 0.00.

Function 1, Management of Instructional Time, had a standard deviation of 0.65, with the variations for the other functions all being closer than this figure. In Function 1, Management of Instructional Time, six of the raters gave a rating for an acceptable score (3 or 4) and eight of the raters gave a 2, Below Standard. Five raters gave an acceptable score of 3, At Standard, for Function 3, Instructional Presentation, whereas nine gave a 2, Below Standard. Function 4, Instructional Monitoring, was scored at 2, Below Standard, or 1, Unsatisfactory, by all evaluators. In the last function, Instructional Feedback, six of the raters scored Below Standard (2), while eight raters scored satisfactory or higher (3 or 4). The division between the acceptable (3 or higher) score and the substandard (2 or 1) is the basis for problems in reaching consensus for this lesson.

"Principals" ratings were close in four of the five teaching behaviors. In Function 2, there was perfect agreement on the score of 2. Functions 3, 4, and 5 showed close, but not perfect, agreement. In Functions 3 and 5, there was a split between acceptable and substandard scores.
Function 4 was rated substandard by all eight raters. The split in Function 1, Management of Instructional Time, was evident in the standard deviation of 0.74. Four of the raters gave At Standard or Above Standard (3 or 4) and four gave Below Standard (2). The "Principals" variation of 0.74 for Function 1 was balanced by the closer agreement of "Central Office Administrators," resulting in less variation overall. Only two of the eight principals agreed that every function was Below Standard or Unsatisfactory.

"Central Office Administrators" had close agreement on all functions except Instructional Feedback. Management of Student Behavior showed perfect agreement in ratings, just as the "Principals" group had rated it. Function 4, Instructional Monitoring, was rated less than satisfactory by all six of the "Central Office Administrators." In Function 1 and 3, the administrators were split on their ratings of satisfactory (3) or unsatisfactory (2 or 1). Disagreement among "Central Office Administrators" for Function 5 was balanced by the "Principals" closer agreement on the same function, resulting in close agreement across "All Participants."

This ninth grade English lesson was viewed and rated by six-member teams. Two teams were formed for this activity.
The two teams were very close in their ratings. For Functions 1 through 4, ratings of 2, Below Standard, were given. For Function 5, *Instructional Feedback*, one team scored a 3, At Standard, the other team a 2, Below Standard. Mean ratings for the teams were lower than the mean ratings for the individual participants. Only one of the team averages was above 2.0; three of the average ratings for "All Participants" were above 2.0 (Table 21).

A single strength was recorded by the teams. The team that gave a score of 3, At Standard, for *Instructional Feedback*, noted it as a strength. Although individual raters had listed "presentation" and "management of student movement," these were not listed in the consensus assessment. The team that rated all Below Standard scores for the lesson had individual members give six different strengths, none of which were listed for the team consensus rating.

Areas that need improvement were more plentiful in the team ratings than were strengths. *Managing Student Behavior* and *Instructional Presentation* were selected as the most important function areas for improvement. Suggestions included: Establish rules for verbal behavior; Provide written activities to check all students; Move around
Table 21

TAPE 7: RATINGS BY SIX-MEMBER TEAMS

<table>
<thead>
<tr>
<th>FUNCTIONS*</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>TEAMS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>B</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>MEANS</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.0</td>
<td>2.5</td>
</tr>
<tr>
<td>AP MEANS*</td>
<td>2.5</td>
<td>2.0</td>
<td>2.35</td>
<td>1.85</td>
<td>2.64</td>
</tr>
</tbody>
</table>

Note. FUNCTIONS=1) Management of Instructional Time; 2) Management of Student Behavior; 3) Instructional Presentation; 4) Monitoring of Student Performance; 5) Instructional Feedback.
AP MEANS= Mean scores derived from all participants' individual ratings for the tape.
classroom; Provide variety in instruction methods. The development of a more concise lesson that does not include unnecessary material was noted as a suggested improvement.

Consensus was reached by the two teams with 19 changes in individual ratings. There were a total of 70 individual ratings. Only four scores were raised during the group discussions. Fifteen scores were lowered in the consensus activity. Nearly four times as many scores were lowered than were raised, a ratio which had not been recorded in any other tape.

Observers of the six-member teams reported that procedures were more formal and regulated in reaching consensus than in the smaller teams. With more people involved in the rating process, attempts were made to give everyone an opportunity to contribute ideas and still keep a reasonable time limit on the task. The first team was led by a central office administrator who took charge by designating which person should begin sharing observations and ratings. Other raters followed, going around the circle in turn clockwise. After each person had spoken and questions were asked by members who wanted clarification, the leader called for consensus. Each person gave at least two reasons for their ratings. Although the explanation by
one member for the rating for Function 3 Instructional Presentation was lengthy, the group had no disagreement on a consensus rating. The individual ratings were close for all functions. Three members of the group gave more input and elaboration about reasons than did the other three members.

The second team began its discussion with the posing of a question by a principal: "Who wants this teacher in your school?" Participants responded with smiles and answers of "Not me," and "I hope I can get someone better." In this team, a central office administrator took the role of leader initially, just as happened in the other six-member team. The leader gave her rating and stated four reasons. She then called on each member, moving around the table to the left. Each person looked at their notes, stated observations, and ended by citing a rating. This procedure was followed for each function. The leader clarified and re-stated points made by participants at least twice during each round of discussion. At a time when two members questioned the speaker about a comment, a third member asked "Are we going all the way around?" The observer noted irritation in the rater's voice and a tense facial expression as he attempted to keep the team members from interrupting before all members had spoken. A team member who questioned the direction the team was leaning for
consensus asked, "Are you going to let having materials ready be as important as time-on-task? Are you going to give her something she doesn't deserve?" The leader closed the discussion on that particular function saying, "No consensus. Let's go on down [to the next function]." They later came back to this function and reached consensus quickly.

For the next function, the member who had expressed irritation with the raters who spoke out of order, said "I want to start this one." He then gave his rating and reasons, followed by the other team members. This new order of speaking was followed to discuss the remainder of the functions. Throughout the discussion, raters looked from their notes to the person speaking, exhibiting facial expressions of support such as smiles and head nodding. Checks and notes were made on the written observations as other members talked about specific examples. One administrator brought up the question of the seating arrangement in the classroom and asked if anyone else noticed what he had noticed. He mentioned, "If you walk into this class and see all the Blacks sitting on one side and whites on the other side, all the boys up against the back wall, what do you think?" Three responses were given including "You could ask the teacher in the post-conference
if she had a seating chart." A principal suggested approaching the teacher and asking how she decided where each student would sit. Other comments reinforced the importance of an evaluator's attention to details in the classroom which may contribute to potential problems. In reaching consensus on the last function, the statement was made that "No one needs to be a 1." A rating of 2, Below Standard, was agreeable to the group.

TAPE 8: Description and Ratings

This videotape is the same one used for Tape 1. The same tape was used to compare the closeness in agreement of individual ratings at the beginning and end of the workshop activity. The primary reading class lasts 33 minutes.

The 14 participants who rated this lesson on the second viewing, had close agreement on Function 1, Management of Instructional Time, only. This function was the only one showing close agreement in the first viewing also. (See Tables 22 and 23.) In this last tape, the other functions had standard deviations above 0.70, going as high as 0.83. The variation levels for Functions 1, 3, 4, and 5 indicate closer agreement for Tape 8 than for Tape 1. In the last
### Table 22

**TAPE 8: MEANS AND STANDARD DEVIATIONS FOR FUNCTION RATINGS**

**PARTICIPANT GROUPS**

<table>
<thead>
<tr>
<th>FUNCTIONS</th>
<th>AP*</th>
<th>PR*</th>
<th>CO*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>1</td>
<td>3.14</td>
<td>.53</td>
<td>3.0</td>
</tr>
<tr>
<td>2</td>
<td>3.07</td>
<td>.83</td>
<td>3.0</td>
</tr>
<tr>
<td>3</td>
<td>3.07</td>
<td>.73</td>
<td>3.29</td>
</tr>
<tr>
<td>4</td>
<td>2.85</td>
<td>.77</td>
<td>2.71</td>
</tr>
<tr>
<td>5</td>
<td>3.64</td>
<td>.74</td>
<td>3.71</td>
</tr>
</tbody>
</table>

**Note.** AP = All Participants (N=14); PR = Principals (N=7); CO = Central Office Administrators (N=7).

FUNCTIONS:
1) Management of Instructional Time;
2) Management of Student Behavior;
3) Instructional Presentation;
4) Monitoring of Student Performance;
5) Instructional Feedback.
Table 23

INDIVIDUAL RATINGS FOR TAPE 8

<table>
<thead>
<tr>
<th>RATERS</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>P2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>P3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P4</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>P5</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>P6</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>P7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P8</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>P9</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>C1</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>C2</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>C3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>C4</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>C5</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>C6</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>C7</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>C8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. P = Principals; C = Central Office Administrators.
FUNCTIONS = 1) Management of Instructional Time; 2) Management of Student Behavior; 3) Instructional Presentation; 4) Monitoring of Student Performance; 5) Instructional Feedback.
RATINGS: 6 = Superior; 5 = Well Above Standard; 4 = Above Standard; 3 = At Standard; 2 = Below Standard; 1 = Unsatisfactory.
* = Participant did not rate this lesson.
viewing, one participant rated Below Standard (2) for 
*Management of Instructional Time* (Function 1). In Tape 1 
there had been no score less than At Standard (3) for that 
function. Function 2 on Tape 8 showed a 50% decrease in the 
number of Below Standard (2) ratings compared to Tape 1. 
For Functions 3 and 4 there were no scores of Well Above 
Standard (5) for Tape 8, a change from Tape 1. The scores 
became closer in Tape 8 for both of those functions, but by 
only a small degree for Function 4.

"Principals" had close agreement only on Function 1, 
*Management of Instructional Time*, with a standard deviation 
of 0.58. The mean of 3.0 was At Standard. Function 2 had a 
standard deviation of 1.0. The mean rating was the same as 
the mean for the first function. In the last three 
functions, the standard deviation was 0.76 for each of them. 
This score is just over the limit considered close 
agreement. The means for the "Principals" ranged from 2.71 
to 3.71 for these teaching behaviors. *Management of 
Student Behavior* had ratings which ranged from 2, Below 
Standard, to 5, Well Above Standard. The range for 
Functions 2, 3, and 4 were from 2, Below Standard, to 4, 
Above Standard. Disparities in the perceived quality of the 
teaching performance for these functions was evident. For
Instructional Feedback, Function 5, scores were all satisfactory or higher (3 to 5).

"Central Office Administrators" had less variation in ratings for three functions. Functions 1, 2, and 3 had standard deviations of 0.69 and below. The other two functions had variations that were slightly above the limit for close agreement. The mean ratings for Functions 1, 2, 4, and 5 were At Standard scores. Function 3, Instructional Presentation, had a mean of 2.86, just below the At Standard score of 3.0. In all functions except Management of Instructional Time, "Central Office Administrators" had ratings which ranged from 2, Below Standard, to 4, Above Standard. Function 2, Management of Student Behavior, had ratings of 3, At Standard, and 4, Above Standard. In contrast to the scores of the "Principals," the administrators from the central office kept their scores within a three-point range for every function.

SUMMARY OF RATINGS

The ratings for "All Participants" showed good agreement on three of the eight tapes. Tapes 4, 6, and 7 had average standard deviations of 0.53 or lower. In each
of these tapes, the variation levels for all individual functions were below 0.70. The average ratings for all functions in these three tapes were all below standard scores, below 3.0. The scores were in the Below Standard and Unsatisfactory (2 and 1) categories, indicative of poor teaching performance. In all eight tapes, Functions 1, Management of Instructional Time, and 2, Management of Student Behavior, showed the highest agreement, with average standard deviations of 0.64 and 0.65. In 20 of the 40 functions scored for the eight tapes, there was good agreement (<0.70 SD).

The least agreement for "All Participants" was on Tapes 3 and 5. Tapes 1, 2, and 8 also had standard deviations above 0.70. The mean ratings for these five tapes were in the At Standard (3) or Above Standard (4) range. Agreement on the ratings for the "All Participants" group was not reached on the satisfactory or better teachers (Ratings of 3, 4, 5, or 6). The degree of skill was not agreed on as readily as it was for the lower-rated teachers. The least agreement on a particular function was for Function 5, Instructional Feedback. Functions 3, 4, and 5 had close agreement in only three of the five tapes for the individual function ratings.
Ratings for "Principals" showed close agreement on four of the eight tapes. Tapes 2, 4, 6, and 7 had average standard deviations below 0.70. Tape 7 had the closest agreement with a 0.42. For Tapes 4, 6, and 7, the mean ratings for all functions were Below Standard or Unsatisfactory. Tape 2 had an average rating of Well Above Standard. Scores for individual functions in all tapes showed good agreement (<0.70 SD) in 19 of 40 functions. The functions with the best agreement of ratings overall were Management of Student Behavior (2) and Management of Instructional Time (1). Each of these functions had overall standard deviations under 0.70.

The most variation in ratings for "Principals" was for Tapes 1, 3, 5, and 8. All of these tapes had standard deviations above 0.70. The average ratings for these four tapes were in the At Standard or Above Standard range. The individual function with the least agreement was 3, Instructional Presentation. In five of the eight tapes there was not good agreement; in three of the eight tapes there was good agreement on this function. As in the "All Participants" ratings, the "Principals" had more divergent scores in the satisfactory and above range than in the below standard range. "Principals" agreed more on the poor
performance of teaching skills than on the satisfactory or better performance.

"Central Office Administrators" had close agreement on six of the eight tapes. Tape 4 had the best agreement. In the six tapes, 23 of the 30 individual functions met the limit for close agreement as demonstrated by standard deviations of less than 0.70. The overall ratings for these tapes ranged from unacceptable (two tapes) to acceptable (four tapes). This range does not indicate a problem with a particular level of teaching performance as was evident in the scores of "All Participants" and "Principals." The functions with highest agreement were Management of Instructional Time (1) and Instructional Presentation (3). In six of the eight tapes, Function 1 had a close standard deviation. Function 3 had close agreement in seven of the eight tapes. The mean ratings of these two functions were both standard and below. The range of scores was from Unsatisfactory to Well Above Standard.

The two tapes in which there was less agreement, Tapes 2 and 3, had standard deviations of 0.84. In no function on Tape 2 and in only one function on Tape 3 was there close agreement of ratings. Deviation went higher than 1.0 in two functions. The mean ratings in these two tapes were Above
Standard scores. The least agreement for an individual function in all eight tapes was in Function 5, Instructional Feedback. There was good agreement in only two tapes for this function. The mean ratings in all tapes for Function 5 ranged from Unsatisfactory to Above Standard, indicating no pattern for "bad" or "good" teaching.

In all classifications of ratings, Function 1, Management of Instructional Time, had highest agreement. "Principals" also had high agreement for Function 2, Management of Student Behavior, and "Central Office Administrators" had high agreement for Function 3, Instructional Presentation. Function 3 had low agreement for "Principals." For "All Participants," there was good agreement on the worst teachers with the lowest scores, and the least agreement on the acceptable or better teachers.

A comparison of each participant’s ratings from Tape 1 to Tape 8 shows the change in distance from the group mean for each function. For participants who did not rate all tapes, the calculation was based on the change from the first tape they rated to the last tape they rated. Table 24 shows the results for each participant. Of the eight participants who rated all 40 functions on the eight tapes, three moved their ratings closer to the group means for the
final observation in the workshop (C2, C3, C6). The one participant who viewed seven tapes did not have closer agreement on the last tape (P6). The participants who completed six, four, and three observations moved closer to the means for the last observations (P2, P8, P9, C8, C9). Of the participants who completed five observations, two changed their scores closer to the means (P3, P7), two did not get closer (P5, C7). Overall, 10 of the 18 participants (56%) had ratings closer to the mean scores for the last observation they did, than for their first observation and rating. Five of the "Principals" and five of the "Central Office Administrators" rated closer.

Comparisons of the changes in standard deviations and means for Tape 1 and Tape 8 by the three groups reveal closer agreement in the ratings for 10 of 15 functions (Table 25). The same teaching tape was used for Tape 1 and Tape 8. Comparing the ratings for the two tapes shows the change in agreement of ratings. For "All Participants," the standard deviation was lower in Tape 8 for Functions 1, 3, 4, and 5. Only on Function 1 was there an increase in the variation of scores for the total group of raters. The mean ratings were lower for Functions 1 and 4, higher for Functions 2 and 3, and the same for Function 5 in the comparison of scores for the last tape to the first tape.
### Table 24

**DIFFERENCES IN PARTICIPANTS’ RATING AND GROUP MEAN RATINGS FOR FIRST AND LAST OBSERVATIONS**

<table>
<thead>
<tr>
<th>P</th>
<th>N</th>
<th>TAPE</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>TOT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P1</td>
<td>8</td>
<td>T1</td>
<td>.25</td>
<td>.21</td>
<td>.36</td>
<td>0</td>
<td>.64</td>
<td>1.46</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T8</td>
<td>.86</td>
<td>.07</td>
<td>.93</td>
<td>.15</td>
<td>1.35</td>
<td>3.36</td>
</tr>
<tr>
<td>P2</td>
<td>6</td>
<td>T1</td>
<td>.75</td>
<td>1.21</td>
<td>.36</td>
<td>0</td>
<td>.64</td>
<td>2.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T8</td>
<td>.14</td>
<td>.07</td>
<td>.07</td>
<td>1.15</td>
<td>1.35</td>
<td>2.78</td>
</tr>
<tr>
<td>P3</td>
<td>5</td>
<td>T1</td>
<td>1.25</td>
<td>.21</td>
<td>2.36</td>
<td>0</td>
<td>2.36</td>
<td>6.18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T7</td>
<td>.5</td>
<td>0</td>
<td>.36</td>
<td>.15</td>
<td>.36</td>
<td>1.37</td>
</tr>
<tr>
<td>P4</td>
<td>8</td>
<td>T1</td>
<td>.25</td>
<td>.21</td>
<td>.36</td>
<td>2.0</td>
<td>.64</td>
<td>3.46</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T8</td>
<td>1.14</td>
<td>1.93</td>
<td>1.07</td>
<td>.85</td>
<td>.65</td>
<td>6.12</td>
</tr>
<tr>
<td>P5</td>
<td>5</td>
<td>T3</td>
<td>.21</td>
<td>.43</td>
<td>.86</td>
<td>.57</td>
<td>.5</td>
<td>2.57</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T8</td>
<td>.14</td>
<td>1.07</td>
<td>.07</td>
<td>.85</td>
<td>.65</td>
<td>2.78</td>
</tr>
<tr>
<td>P6</td>
<td>7</td>
<td>T2</td>
<td>.31</td>
<td>.08</td>
<td>.92</td>
<td>.15</td>
<td>.69</td>
<td>1.23</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T8</td>
<td>.14</td>
<td>.07</td>
<td>.93</td>
<td>.15</td>
<td>.65</td>
<td>1.94</td>
</tr>
<tr>
<td>P7</td>
<td>5</td>
<td>T1</td>
<td>.25</td>
<td>1.21</td>
<td>1.36</td>
<td>0</td>
<td>.64</td>
<td>3.46</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T7</td>
<td>.5</td>
<td>0</td>
<td>.36</td>
<td>.15</td>
<td>.64</td>
<td>1.65</td>
</tr>
<tr>
<td>P8</td>
<td>6</td>
<td>T3</td>
<td>1.21</td>
<td>.43</td>
<td>.86</td>
<td>1.57</td>
<td>.5</td>
<td>4.57</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T8</td>
<td>.14</td>
<td>1.07</td>
<td>.93</td>
<td>.15</td>
<td>.35</td>
<td>2.64</td>
</tr>
<tr>
<td>P9</td>
<td>4</td>
<td>T5</td>
<td>.07</td>
<td>.07</td>
<td>1.07</td>
<td>.4</td>
<td>0</td>
<td>1.61</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T8</td>
<td>.14</td>
<td>.07</td>
<td>.07</td>
<td>.85</td>
<td>.35</td>
<td>1.48</td>
</tr>
</tbody>
</table>

**Note.** FUNCTIONS = The five teaching functions on the NCTPAI.
P = Participants
N = The total number of videotaped lessons observed by the participant during the workshop.
TAPES = The first and last tapes in chronological sequence that were observed by the participant.
DIFFERENCES: Calculated by subtracting the participant’s rating from the group mean rating. Positive or negative values are not considered when finding the total.
### Table 24
(Continued)

DIFFERENCES IN PARTICIPANTS’ RATING AND GROUP MEAN RATINGS FOR FIRST AND LAST OBSERVATIONS

<table>
<thead>
<tr>
<th>P</th>
<th>N</th>
<th>TAPE</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>TOT</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>8</td>
<td>T1</td>
<td>.75</td>
<td>.79</td>
<td>.64</td>
<td>0</td>
<td>.64</td>
<td>2.82</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T8</td>
<td>.86</td>
<td>.93</td>
<td>.93</td>
<td>1.15</td>
<td>.35</td>
<td>4.22</td>
</tr>
<tr>
<td>C2</td>
<td>8</td>
<td>T1</td>
<td>.25</td>
<td>1.21</td>
<td>.64</td>
<td>1.0</td>
<td>1.36</td>
<td>4.46</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T8</td>
<td>.86</td>
<td>.07</td>
<td>1.07</td>
<td>.15</td>
<td>.35</td>
<td>2.5</td>
</tr>
<tr>
<td>C3</td>
<td>8</td>
<td>T1</td>
<td>.75</td>
<td>.79</td>
<td>.64</td>
<td>1.0</td>
<td>.64</td>
<td>3.82</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T8</td>
<td>.14</td>
<td>.07</td>
<td>1.07</td>
<td>.15</td>
<td>.35</td>
<td>1.78</td>
</tr>
<tr>
<td>C4</td>
<td>8</td>
<td>T1</td>
<td>.25</td>
<td>.21</td>
<td>.64</td>
<td>0</td>
<td>.36</td>
<td>1.46</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T8</td>
<td>.14</td>
<td>.07</td>
<td>.07</td>
<td>.85</td>
<td>.35</td>
<td>1.48</td>
</tr>
<tr>
<td>C5</td>
<td>8</td>
<td>T1</td>
<td>.25</td>
<td>.79</td>
<td>.36</td>
<td>0</td>
<td>.36</td>
<td>1.76</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T8</td>
<td>.14</td>
<td>1.07</td>
<td>.07</td>
<td>1.15</td>
<td>.35</td>
<td>2.78</td>
</tr>
<tr>
<td>C6</td>
<td>8</td>
<td>T1</td>
<td>.75</td>
<td>.21</td>
<td>.64</td>
<td>0</td>
<td>.36</td>
<td>1.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T8</td>
<td>.14</td>
<td>.93</td>
<td>.07</td>
<td>.15</td>
<td>.35</td>
<td>1.64</td>
</tr>
<tr>
<td>C7</td>
<td>5</td>
<td>T1</td>
<td>.75</td>
<td>.79</td>
<td>.36</td>
<td>0</td>
<td>.64</td>
<td>2.54</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T8</td>
<td>.14</td>
<td>.07</td>
<td>.07</td>
<td>.85</td>
<td>1.65</td>
<td>2.78</td>
</tr>
<tr>
<td>C8</td>
<td>3</td>
<td>T1</td>
<td>.75</td>
<td>.79</td>
<td>.64</td>
<td>1.0</td>
<td>.64</td>
<td>3.82</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T4</td>
<td>.29</td>
<td>.73</td>
<td>.27</td>
<td>.67</td>
<td>.4</td>
<td>2.36</td>
</tr>
<tr>
<td>C9</td>
<td>4</td>
<td>T1</td>
<td>.75</td>
<td>.79</td>
<td>.36</td>
<td>1.0</td>
<td>.64</td>
<td>3.57</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T4</td>
<td>.27</td>
<td>.27</td>
<td>.27</td>
<td>.33</td>
<td>.6</td>
<td>1.74</td>
</tr>
</tbody>
</table>

**Note.** FUNCTIONS = The five teaching functions on the NCTPAI.
P = Participants
N = The total number of videotaped lessons observed by the participant during the workshop.
TAPES = The first and last tapes in chronological sequence that were observed by the participant.
DIFFERENCES: Calculated by subtracting the participant’s rating from the group mean rating. Positive or negative values are not considered when finding the totals.
Table 25

COMPARISON OF STANDARD DEVIATIONS AND MEANS BETWEEN TAPE 1 AND TAPE 8

ALL PARTICIPANTS:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>T1</td>
<td>.65</td>
<td>3.57</td>
<td>.80</td>
<td>2.79</td>
<td>.93</td>
</tr>
<tr>
<td>T8</td>
<td>.53</td>
<td>3.14</td>
<td>.83</td>
<td>3.07</td>
<td>.73</td>
</tr>
</tbody>
</table>

Note: Tape 1: N = 14; Tape 8: N = 14.

PRINCIPALS:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>T1</td>
<td>.71</td>
<td>4.0</td>
<td>.55</td>
<td>3.40</td>
<td>1.3</td>
</tr>
<tr>
<td>T8</td>
<td>.58</td>
<td>3.0</td>
<td>1.0</td>
<td>3.0</td>
<td>.76</td>
</tr>
</tbody>
</table>

Note: Tape 1: N = 5; Tape 8: N = 7.

CENTRAL OFFICE ADMINISTRATORS:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>T1</td>
<td>.50</td>
<td>3.30</td>
<td>.73</td>
<td>2.44</td>
<td>.50</td>
</tr>
<tr>
<td>T8</td>
<td>.49</td>
<td>3.29</td>
<td>.69</td>
<td>3.14</td>
<td>.69</td>
</tr>
</tbody>
</table>

Note: Tape 1: N = 9; Tape 8: N = 7.
The "Principals" had closer standard deviations for Functions 1, 3, 4, and 5, in Tape 8 than in Tape 1. This pattern was the same as for "All Participants." Function 2 scores were wider apart in the last tape than the first. Although four functions had lower mean scores in Tape 8 than Tape 1, the differences were 1 point or less. The mean for Function 3 increased an insignificant amount.

"Central Office Administrators" as a separate group showed a different change from the "Principals." Two standard deviations decreased (Functions 1 and 2); three increased from Tape 1 to Tape 8 (Functions 3, 4, and 5). The means were higher for four functions in Tape 8 (Functions 2, 3, 4, and 5), but one increase was very slight (Function 5).

RATERS' PERCEPTIONS

An interview was conducted with participants two weeks following the final rating exercise in which Tape 8 was rated. The interviews were held with individuals using a set of 16 questions that were asked each person. As some of the participants responded, their answers encompassed other questions. The interview questions, in the order in which
they were posed, are in Appendix I. Responses to the questions will be presented grouped by topics.

QUESTION 1: WHAT IS THE PURPOSE OF THE NORTH CAROLINA TEACHER PERFORMANCE APPRAISAL SYSTEM OR OF ANY EVALUATION SYSTEM?

Every participant answered without hesitation that the purpose is improvement of teaching. Terms such as "help teachers to grow," "improve instruction," "to help teachers," and "growth" were used by raters. In the group discussions, it was stated over and over that helping teachers improve is the reason for evaluating teachers. In no instance did a participant cite the removal or demotion of a teacher as the purpose. With prodding from the interviewer, it was acknowledged that the system could be used for personnel decisions, but that would be after all resources had been exhausted to improve the teacher.

QUESTION 2: WHY DID YOU RATE THE WAY YOU DID?

For this question, the participants' responses were more varied than the answers to the first question. Some participants described their training as a basis for the ratings. The "instrument descriptions of the scores and
what I observed" was the basis for one evaluator. Another rater stated "I look at the scale, what is required by the lesson plan guide and by the job description. I look for specific things." The same rater continued with the comment that checking is done to see "if the teacher does an adequate job in areas that are addressed or identified. I ask myself 'Is the person meeting minimum expectations?' If the lesson is Below Standard and can be improved, I wouldn't go to Unsatisfactory."

Some described the process and the resources they use in rating. "What I wrote in the narrative, notes on my FODI, notes I took," "Observations, what I see," "What I'm looking for and what I see," were stated by three evaluators. One principal stated "Raw data, FODI, what I see. Things other than what is listed on the FODI - I use anything I see in that classroom. I count the number of times that the teacher said or did something. It's not just what is listed, it's a summary of all I saw in the classroom." The role of specific functions is addressed by the statements "I look at functions, pluses and minuses," and "I star or mark Feedback, and so forth, and go back through, type it up, then read through it."
Others described justifications they had imposed for themselves. It is a "value judgment," one states. Other responses included "I make a comparison to good teachers," "I look at how effective the lesson was, how well skills are taught, how much kids learn." A central office administrator claimed that he rates the teaching process as he sees it.

QUESTION 3: WHAT WOULD A TEACHER DO DIFFERENTLY TO RECEIVE A 5 INSTEAD OF A 4?

QUESTION 4: HOW SURE ARE YOU THAT THE SCORE SHOULD BE A 4 FOR THAT FUNCTION? WHY?

QUESTION 5: HOW DO YOU PERCEIVE THE SCORE OF 6 VERSUS 3?

These three questions investigate perceived differences in specific scores and the rater's confidence in the decision. Concern about the top three scores, Above Standard (4), Well Above Standard (5), and Superior (6), was stated by three interviewees. The problem lies with deciding "how much" above standard. A statement by one rater is indicative of the problem: "We don't have a good objective way to justify [the top three scores], unless there were some numerical scale or something." This same administrator continues that Above Standard would be
adequate, "the other two are just confused. The two top ones, I can't justify a difference. The only way to justify a difference in the top three is by comparing a person's performance with somebody else's performance and that's not what we're supposed to do." This was not the same rater who stated in a previous question that comparison to good teachers was the way to rate a person.

Reliance on previous experiences and observations to differentiate between a 4 and a 5 is expressed by a rater who said she would rate in relation to what she has seen or what she knows "ought to be. Above Standard is slightly better than average. This person put forth a little bit extra effort, but not as much as Superior." The one with the Superior "went all out - extra grouping, extra involvement, extra everything. You would not see it in every class." Another administrator with many years experience said that he looked at all of the indicators for each major function, then decided. "It's not an exact thing. It's difficult to decide. If Above Standard in more of these [indicators], then I may lean toward Above Standard. If doing mediocre on more... then I lean toward the one they have the most of. They won't be at the same level on every one of these major functions or even the indicators."
Deciding between a 4 and a 5 was explained in this way:
"I look at whether the teacher is just doing what's required or did exceptionally well, going above and beyond, doing well - that's a 5. Four's a little above standard; the teacher may be doing a few things." This administrator continued, "Trouble deciding? Only with a difference of 1 point; not between a 2 and a 4 - that's unsatisfactory versus satisfactory. I've given unsatisfactory if there was a definite thing that needed to be worked on and was documented." This same person talked about hesitancy to give high scores. "You're always becoming [better]. You can't be Superior. I give very few 6s."

A principal who liked the opportunity to observe with another person to check agreement, said that a 5 meant teaching all the time, "Into the lesson every time I walk in, not just beginning or ending." Another description was that a 5, Well Above Standard, would "not be frequent at all in comparison with Above Standard". He continues, "I haven't seen many 6s thus far. I don't recall having rated many 5s; not that I'm being overly critical. Going beyond that Above Standard is rare. You won't see many 6s." A 5, comments one principal, means "more and better - quality more than quantity, outstanding in most instances, not just once in a while. There is evidence of more than is required many
times. A 4 - sometimes more than the minimum but not always."

One observer indicated the willingness to negotiate between scores if necessary when deciding between a 4 and a 5. "I can see Below Standard or Standard. It would be easier for someone to talk me into an Above Standard and a Well Above Standard, than it would [between] Below Standard and At Standard. That’s the line. At Standard is acceptable. Below Standard is unacceptable." The distinction is further made, "At Standard has those things that should be on the lesson; if those are the only things that I see, that would make that lesson At Standard. If something added a great deal to the lesson, it can be negotiated between a 5 and a 4."

Some of the respondents commented on how sure they were about their rating being correct. One stated "I was positive, whether the teacher consistently demonstrated high level performance, I based my number on that." Another was less confident, explaining "We can never be absolutely sure or certain. I just base the score on what I saw. I can give evidence of the practices and the effect on the lesson." Two raters mentioned the practice of comparing observations with another observer to check for agreement.
"I had a supervisor come in," states a principal. "Amazing - We were both right on the money. That helps me. We've written down the same things. Amazing - That kinda tells me I'm not too far off." A second evaluator added, "I feel comfortable when I put it on paper - that much more comfortable when I have an opportunity to share my ratings with colleagues who observe the same person."

The necessity for observing a teacher more than one time to differentiate ratings of 4 and 5 is advocated in the statement "We're fooling ourselves if we judge on one class period. It's unfair to rate a teacher by being in the classroom one time. It's knowing that person every single day, every single class period, and also out of class." This administrator supports the idea that a central office person should not do a summative, because the evaluator "needs to be in that classroom five minutes here and five minutes there throughout the school year."

In describing the difference in a 6, Superior, versus a 3, At Standard, raters' perceptions fell in two categories: one gave re-statements and synonyms for the adjectives in the rating scale definitions; the other gave specific examples of the two ratings. Words used to describe a 3 were "average," "least you ought to be doing," "expected, a
minimum." A 6 is "superior, the best," "super teacher," "consistently outstanding," "very unique consistently — in every area." Specific examples given of a 3 included these perceptions: "Observing the minimal written or specified requirements. You're supposed to have certain components on the lesson plan — you did it," "Doing enough to get by with, not doing as well as they could do it," "Nothing exciting, but no great omissions. I would expect the teacher to work to improve areas." A 6 "keeps students enthused, interested, makes you want to take part. It means nothing else could be added to make it better — the teacher can help others." Student outcomes are considered by one evaluator when designating a 6 rating. "Test scores are gonna have to show it, that the child is really gripping what they're teaching, students must understand." This rater explains the need to use different test results criteria for certain teachers. "With teachers of LD [learning disabled] and exceptional children, you have to look at where they were and the growth made at the end of the year." A teacher who scores a 6 is one who does a good job repeatedly, a central office administrator states. "I can observe and pick out the numbers on the FODI and everything is there. Everything falls in place."
QUESTION 6: WHICH TEACHING FUNCTION DID YOU FIND THE MOST DIFFICULT TO RATE? EASIEST?

QUESTION 7: WHICH ONE OR TWO FUNCTION AREAS DO YOU THINK YOU ARE MORE KNOWLEDGEABLE IN FOR RATING PURPOSES?

During the workshop, much discussion took place to clarify and explain why a particular rating was selected. Participants shared their observations, checking for behaviors which may have been missed or interpreted differently by certain participants. In the discussions it was pointed out numerous times that certain functions may have been more difficult or easier for a rater to score. In the interview, the evaluators were asked which function area is most difficult to rate, which is the easiest, and which one the rater was most knowledgeable in. Participants sometimes named more than one function for each category. Of the responses given, more than half named Function 4, Monitoring of Student Performance, as the most difficult. One problem presented was the difficulty in hearing what is said and in seeing the interactions when a teacher goes to an individual student's desk. Further, the teacher can be seen moving about the room, but it is difficult to see exactly what is being done. The observer cannot usually see what the student has written at the desk, nor the body
language and facial expressions that the teacher sees.

Function 3, Instructional Presentation, was chosen as the most difficult by the next largest number of observers. Difficulty in determining if the lesson is relevant and if it is accomplishing the objective was cited. One administrator noted "if the kids keep their mouths shut, we'll never know" if the lesson was effective. Comments made by another person who selected Function 3 pertained to lack of familiarity with content areas. Functions 1, 2, and 5 were selected by two raters each. With Management of Student Behavior (2), it could be hard to distinguish between good behavior and a lack of creativity in the classroom. Creativity could be misinterpreted as lack of control. Instructional Feedback (5) can prove difficult because the observer in a single class cannot see all the things which may be done outside the class to check and react to the students' work.

Half of the participants selected Function 1, Management of Instructional Time, as the easiest to rate. Writing down times during the lesson, seeing if materials are ready, looking to see if students get on task quickly, are readily observed. As one evaluator stated, "The teacher either causes things to happen or they don't." Management of Student Behavior, Function 2, was next easiest to assess.
The ease of seeing how students behave and if the teacher has their attention were mentioned. A small number of evaluators cited Functions 4 and 5 as easiest. Contrary to the responses for the most difficult function, the comment was made that it was easy to see monitoring and to hear specific feedback. Instructional Feedback has less to look for, according to one interviewee. Function 3, Instructional Presentation, was not named by any evaluator as the easiest function.

In naming the function in which the participant is the most knowledgeable, Instructional Presentation, Function 3, was named by three-fourths of the administrators. This overwhelming majority of the raters suggested reasons such as the "basic training in instruction we have had," the affinity for "looking at different parts of the lesson and giving feedback," and "I have observed so many lessons, studied instruction, taught methods, that I can identify what is in the presentation and what is not." "Teaching experience," which likely was successful for these administrators before leaving the classroom, provides a "basis for this function." Function 2, Management of Student Behavior, was listed by the next largest number of people. Identifying appropriate and inappropriate behavior, training in Assertive Discipline and Classroom Organization
for Effective Teaching (COET), the ease of seeing the consequences of bad behavior, and knowing specific teaching behaviors which determine student behavior were cited as reasons for selection of this function. One person selected Function 4, Monitoring of Student Performance, and two selected Function 1, Management of Instructional Time, for the area of strongest knowledge. In four instances, raters named the same function as the easiest that they named as the one they had the most knowledge in. In only one case was the most knowledgeable function named as the most difficult.

QUESTION 8: DID YOU CHANGE YOUR RATINGS WHEN DISCUSSING THE LESSON?

QUESTION 10: IN REACHING A CONSENSUS WITH YOUR TEAM MEMBERS, DID YOU CHANGE YOUR MIND? WHY?

These two questions dealt with the willingness or reluctance to change scores during discussions of the lessons and during the team consensus rating activities which were a part of the teacher evaluation workshop attended by all the administrators in the school system. All participants except two stated that they did change their ratings at times during the discussions and consensus
ratings. The reasons cited for changing included "Compromise - not because I changed my ideas," "I went along with the group when I did not feel strongly about my rating," "Based on training and experience of other rater; for example, kindergarten level - I had no experience with that level. I changed after a satisfactory explanation." One rater said she changed up or down 1 point, never more than that, based on the discussion in the group. She stated, "I'd give in when the difference was 1 point, listen to the general group conversation. This happened once when they all said At Standard and I gave 1 point higher." Four other raters mentioned specific changes of 1 point. A principal said, "I changed for the reason given, not for who the person was. It was for the things they saw and I didn't notice." All of the participants who said they changed their ratings at times, listed as the main reason the justification or evidence given by the other team members.

The two participants who said they did not change their ratings gave examples they remembered from the activities. One rater described her own refusal to change in terms of "getting in a fight with" other team members who tried to convince her to change. She stated that she felt very sure of her position and nobody could show her anything to change. She said she had confidence that her rating was
right. She laughed as she remembered the good-natured argument the group had. The second rater who said the scores were never changed said the ratings were usually with the majority.

Data on actual changes in ratings are presented with the results of the team ratings in another part of this chapter. Table 6 lists each participant and the number of times scores were changed.

**QUESTION 9: DID YOU NOTICE A PATTERN IN YOUR RATINGS? WERE THEY USUALLY HIGHER OR LOWER THAN OTHER TEAM MEMBERS?**

Of the eleven participants who chose to appraise their rating patterns, four stated they rated the same as most others, and six stated they rated lower than most others. Only one person claimed to rate higher than other members. In actual ratings in the workshop, the individual's rating for each function was compared to the overall group mean for that function. The four raters who said in the interview that they rated the same as everyone else were correct. The other raters who felt they usually rated lower or higher than other evaluators were incorrect for the majority of their ratings. However, in examining only the ratings which were not the same as the group means, five of the rater's
self-appraisals were accurate. In the calculation of the ratings, every participant had a higher percentage of ratings that were the same as the group means, than of ratings which were lower or higher than the overall means. Of the ratings that were not the same, the individual ratings were higher than the group means for 6 of the participants (P3, P6, P7, C1, C2, C8), and lower than the group means for 10 of the participants (P1, P2, P4, P5, P8, C3, C4, C5, C7, C9). For two participants (P9, C6), the ratings were evenly divided between higher and lower than the group means. (See Table 26.)

A difference of one point higher (+) or lower (−) than a score is considered acceptable when using the North Carolina Teacher Performance Appraisal Instrument. Scores within 1 point of the mean, plus or minus, were calculated. For all participants, 93% (245) of the ratings which were not the same as the group means were within the plus or minus 1 point tolerance limit. For ten of the participants, every one of the scores that were not the same as the means were within this tolerance limit (P2, P5, P6, P7, P9, C3, C4, C6, C8, C9). Combining the ratings which were the same as the group means and the ratings which were within 1 point of the group means, reveals that 97% of all ratings were within the acceptable level of tolerance.
Table 26

COMPARISON OF INDIVIDUAL RATINGS TO GROUP MEANS

<table>
<thead>
<tr>
<th>RATER</th>
<th>SELF-APPRAISAL</th>
<th>N</th>
<th>SAME</th>
<th>+/- 1 PT.</th>
<th>TA</th>
<th>L</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>lower</td>
<td>40</td>
<td>16</td>
<td>21</td>
<td>37</td>
<td>17</td>
<td>7</td>
</tr>
<tr>
<td>P2</td>
<td>lower</td>
<td>30</td>
<td>17</td>
<td>13</td>
<td>30</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>P3</td>
<td>-</td>
<td>25</td>
<td>15</td>
<td>7</td>
<td>22</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>P4</td>
<td>lower</td>
<td>40</td>
<td>14</td>
<td>19</td>
<td>33</td>
<td>16</td>
<td>10</td>
</tr>
<tr>
<td>P5</td>
<td>-</td>
<td>25</td>
<td>10</td>
<td>15</td>
<td>25</td>
<td>13</td>
<td>2</td>
</tr>
<tr>
<td>P6</td>
<td>higher</td>
<td>35</td>
<td>22</td>
<td>13</td>
<td>35</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>P7</td>
<td>-</td>
<td>25</td>
<td>16</td>
<td>9</td>
<td>25</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>P8</td>
<td>same</td>
<td>30</td>
<td>19</td>
<td>10</td>
<td>29</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>P9</td>
<td>-</td>
<td>20</td>
<td>14</td>
<td>6</td>
<td>20</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>C1</td>
<td>-</td>
<td>40</td>
<td>20</td>
<td>19</td>
<td>39</td>
<td>8</td>
<td>12</td>
</tr>
<tr>
<td>C2</td>
<td>lower</td>
<td>40</td>
<td>22</td>
<td>17</td>
<td>39</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>C3</td>
<td>lower</td>
<td>40</td>
<td>24</td>
<td>16</td>
<td>40</td>
<td>13</td>
<td>3</td>
</tr>
<tr>
<td>C4</td>
<td>same</td>
<td>40</td>
<td>24</td>
<td>16</td>
<td>40</td>
<td>13</td>
<td>3</td>
</tr>
<tr>
<td>C5</td>
<td>same</td>
<td>40</td>
<td>24</td>
<td>14</td>
<td>38</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>C6</td>
<td>lower</td>
<td>40</td>
<td>28</td>
<td>12</td>
<td>40</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>C7</td>
<td>same</td>
<td>25</td>
<td>12</td>
<td>12</td>
<td>24</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>C8</td>
<td>-</td>
<td>15</td>
<td>3</td>
<td>12</td>
<td>15</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>C9</td>
<td>-</td>
<td>20</td>
<td>6</td>
<td>14</td>
<td>20</td>
<td>13</td>
<td>1</td>
</tr>
</tbody>
</table>

GROUP TOTALS: 570 306 245 551 174 104
PERCENT OF TOTAL: 100% 97%

Note. SELF-APPRAISAL = Rater's own opinion of rating pattern compared to other raters.
N = Total number of ratings.
SAME = Number of ratings which were the same as the group means.
+ or - 1 PT. = Ratings within 1 point higher or lower than the group means. This is the acceptable tolerance level for use in the NCTPAS.
TA = Total number of ratings which were the same or within 1 point of the group mean.
L = Number of ratings which were not the same as the group mean and which were lower than the mean.
H = Number of ratings which were not the same as the group mean and which were higher than the mean.
QUESTION 11: WITH WHICH TEAM WAS IT EASIER TO REACH A
CONSENSUS?

QUESTION 12: WHO WOULD YOU CHOOSE TO FORM A TEAM?

Questions about the makeup of the teams for reaching consensus asked for preferred number of team members and for specific names of people they would put on their team. Three times as many people (6) preferred two-member teams over four-member teams (2). Only one person had a preference for a six-member team, based on the interviewee’s thought that "the more the better" in seeking objectivity. Two evaluators had no opinion on an ideal number. Support given for a particular preference ranged from "the absence of controversy with two people" to "a wider range of views can be considered" with four people.

Administrators were asked to name specific people from the workshop who would be on their selected team, if given a choice. A variety of reasons were given for the selections. Of the 18 participants, 7 were named in varied combinations for the teams. One rater chose people with a cross section of experience levels: elementary, middle, and secondary. Another selection was made based on "knowledge about overall instruction." Three teams were selected with the priority that "they think as I do," "sees same types of things," and
have the "same philosophy, looks for the same things I do."
Objectivity, receptiveness, fairness, and clarity of reasons for their ratings were listed also.

QUESTION 14: WHO ARE THE TWO PEOPLE OUT OF THIS GROUP THAT YOU THINK ARE THE MOST KNOWLEDGEABLE IN THE TEACHER EVALUATION PROCESS?

QUESTION 15: WHO WOULD YOU LIKE TO HAVE RATE YOU?

In statements of opinion about the most knowledgeable two people in the group of workshop participants as far as the teacher evaluation process, one particular individual was selected by half of the respondents as one of their two choices. Two people named themselves as a choice. Of the 18 people in the group, 5 different people were named for this choice. In selecting the one person each administrator would like to have rate them, 90% named one of the same people they named as the most knowledgeable. One interviewee said it made little difference who did the rating. The same person who was named the most times as one of the most knowledgeable, was named the most times as the preferred evaluator.
QUESTION 13: IF YOU WERE DEVELOPING A RATING SCALE, HOW WOULD IT LOOK? HOW MANY POINTS OR CATEGORIES?

When given license to develop a rating scale to replace the six-point scale which is presently used with the North Carolina system, more selected a four-point scale than any other configuration. There were numerous labels given to the four levels of performance. Four people chose to leave off the two highest ratings of Well Above Standard and Superior. One person stated, "When I get above that [Above Standard], personalities start getting into it. If I were challenged, I could not handle it. I'm uncomfortable with those top two." Another rater said it was hard to differentiate between a 4 and a 5 so only two levels are needed above At Standard. Two suggested labeling the levels Below Standard, At Standard, Above Standard, Superior, providing one label for unacceptable scores and three for acceptable scores. One suggestion for a four-point scale was to eliminate Superior and Unsatisfactory. Superior is "untouchable" and one can always improve; Unsatisfactory is so negative, "you can't expect to get much improvement." The same administrator felt that when a teacher cannot leave the Below Standard classification, it is time to get rid of the teacher, for "Who wants that teacher teaching their child?"
A three-point scale using Below Standard, At Standard, and Above Standard was the choice of four raters. Reference was made to the importance of teaching as Acceptable or Unacceptable. Concern that we have allowed a lot of "shades of gray" was expressed by one advocate of the three-point scale. The six-point scale used now is satisfactory for six of the evaluators. Four of them stated they could not come up with anything better, so the present scale would be workable. Others had no preference for a particular scale. Both positive and negative aspects of different scales were mentioned by these people.

QUESTION 16: WAS THIS ACTIVITY WORTHWHILE? HOW WOULD YOU CHANGE IT IF WE DID IT AGAIN?

Participants gave positive responses to the first question and a number of suggestions for the second part of the question. The administrators who stated that the activity was worthwhile expressed concern with subjectivity of evaluators who do not have intermittent training in the process. One central office administrator said that although many people observe the same thing there "is not uniformity. Most with secondary experience went one way; middle and elementary might not see it the same way." College degrees, certifications, and having the evaluation
instrument in one's hand do not make one all powerful, she continued. Subjectivity can be reduced through discussion of the lesson and the teaching behaviors. "The philosophy of the evaluator affects the rating of teaching. Two people can look at something and see something different." The discussions in the workshop activity helped raters see "why we see it differently," she affirmed. It is important for the rater to know the philosophy of the school unit and what is expected of a teacher. What an evaluator personally thinks should not be the sole determining factor in rating a teacher. Regarding the possibility of a career ladder or merit pay plan based on teaching performance, this administrator insisted that group evaluation practice activities must continue so that raters can get closer in their evaluations.

Comments from others who supported the activity included: "It helped to see how far apart we were when we started, how close together we are, and what are our thoughts about the scores." "It opened our eyes as to how other people see things. Discussion helps clarify what we both saw and opened our eyes to something we didn't see." Some doubt in the value of the training for every person was expressed: "I don't know if it would change some individuals when they go back to their schools. I don't
know that they would be any different."
This person felt that "The more we [administrators] evaluate and discuss, the better off we are, or the better we are at doing our job." This same participant was concerned with subjectivity of evaluators. Comparing one teacher with another is something "you can't help but do. That's the way you meet standards of what is average. It's human." It was pointed out that teachers may get a better evaluation at one school than at another. Teachers transfer from school to school and do better or worse, based on who is observing them. The teacher does better because of what a principal who rates higher gave the teacher. This may lead the teacher to change for the better. The practice activity with the North Carolina instrument helps "narrow the focus to specifics; we're talking the same language....Discussing and defending your reasons is good for us."

Principals pointed out the significance of discussing reasons for ratings with others. Having this activity before doing any evaluations would have been more worthwhile to one principal. It should be done periodically, two more principals stated. Another building level administrator felt that "not as much of it is needed for veterans." A counter idea was alleged: "It forces [all principals] to look at instruction, not anything else. It is needed as a
"Talking with others is so helpful," another principal states. "I learned more about things to look at closer by talking to the other principals and supervisors."

Recommendations for changes in the activity were to add half a day or a day to look at the rating comparisons more closely; "Spending some time as a group making sure when we say a 1 we mean a 1," but not to the point of monotony; do not do it if evaluators are in "sync" with colleagues - ICP [Initially Certified Personnel] teams give the opportunity to check. Doing some mock lessons in which certain things are set up and seeing if raters can all identify what went wrong with the lesson was one suggestion. Changes in the scheduling of the activity were discussed by some: "Do it in the fall and spring"; "Do it once a year"; "Do it two days at a time - one day is not enough"; "Split the days." An administrator who listed five positive outcomes of the activity, included this comment for consideration: "Let's do some portions of the activity, but not all. You can always do something to death."
Characteristics of the principals and central office administrators were compared to characteristics of the teacher and the lesson in each of the eight tapes. Similarities in five specific areas were identified. Characteristics which were matched were: grade level training and experience (primary, elementary, middle school, high school); subject; race; gender; and age/experience. Information on the raters' backgrounds and traits were supplied on the Rater Information forms (see Appendix J) which were completed by the participants. Characteristics of the teachers and the videotaped lessons were identified by the researcher. Ages of the participants and of the videotaped teachers were estimated. Years of experience were given on the Rater Information forms to use as a basis for age of participants.

The characteristics match was done to check for bias in patterns of rating teaching performance. The evaluator's mean score for each tape was compared to the group means for each tape. Ratings were classified as higher, lower, or the same as the group means.
Consistency of ratings across tapes based on matching or non-matching characteristics of the evaluators and the teaching episodes, was not evident. Examination of ratings and of the characteristics matches revealed no patterns which could be used to predict scores. There were no indications of rating a teacher higher or lower based on similarities or differences in grade level, subject expertise, race, gender, and age/experience. Rater biases in the five areas did not appear in the data.

TEACHER PERCEPTIONS

A questionnaire was distributed to all teachers in the school system while the study was being conducted with administrators. Teachers were asked to answer the seven questions, including any comments they chose, without identifying themselves or their home schools. (See Appendix K.) Responses were received from 161 teachers. Of the teachers who responded, 60 were in the Initially Certified Program. In this program for beginning teachers, a minimum of three observations are required during the year.

According to the responses from the teachers, 96% felt that their observations the previous year had focused on the
important functions of the North Carolina Teacher Performance Appraisal Instrument. Eighty-one percent of the teachers thought that equal importance was given to the five observable functions. A smaller percentage, 61%, felt that the functions should be given equal importance, while 16% did not think they should. Comments were made in response to the question asking should the five observable functions be treated with equal importance. Teachers who answered "Yes" to the question made comments linking the five functions together. The dependence of one function on the others to have a successful lesson was acknowledged. Several comments expressed the feeling that all five functions should be used, but not necessarily every day; one function may be more evident than others in certain lessons. A teacher who was obviously not pleased with the number of times the principal observed replied, "I do not feel that the principal was in my classroom enough to judge these areas. I feel that he needs to see me teach several times in order to know my strengths and weaknesses."

Teachers who answered "No" to the same question made statements which emphasized certain functions: "Instructional presentation and instructional feedback are more important," "Monitoring and feedback should have the greatest importance, not as much on focus and closure,"
"especially student behavior." Flexibility is a concern for some: "depends on plans for the day," "not during a single class," "kind of activity dictates the importance of each function." One teacher focused on the objective of the observation in the following comment: "No, it seems to me that more emphasis was placed on trying to find faults rather than submitting what really was being done." Complete comments are available in Appendix K for the reader who would like to review all responses to this question.

Opinions were sought concerning observations by single or multiple observers. When asked if multiple observers are preferred, an affirmative answer was given by 52% of the teachers. Thirty-eight percent did not think it was important to have the same evaluator conduct all observations. When asked why or why not, teachers who answered in the affirmative stated reasons such as consistency, improvement can be seen by the same evaluator, progress or lack of progress is visible, and confusion may result from too many ideas. The instructor's style, determination whether the class is typical for that teacher, fairness and equality, and familiarity with pre-existing problems were concerns also. The teachers who responded that it was not important for the same evaluator to conduct all observations saw advantages to having "different ideas
for improvement," "focus on different problems," and "Better averaging of opinions." The reliability of the instrument with different people observing was mentioned. The sharing of the responsibility was seen as a strong point. One teacher wanted more people to be aware of teaching performance. Numerous other comments which were made are in Appendix K.

While 34% of the teachers thought there is consistency in evaluations when they are done by different administrators, 50% did not think there is consistency. The fact that half of the respondents did not think multiple observers would be consistent relates to the opinion that the same evaluator should conduct all observations. Seventy-one of the teachers reported that a single evaluator observed them; sixty-four teachers were observed by two people; and eighteen teachers were observed by three evaluators. Only four teachers reported that four or five different people observed them. Half of the teachers reported that three formal observations were conducted the previous school year. Fewer than three observations, the required number for the system that year, were conducted with 63 teachers.
CHAPTER FIVE

FINDINGS, CONCLUSIONS, RECOMMENDATIONS, AND IMPLICATIONS

This chapter presents a summary of the study of administrators' perceptions of the ratings used on the North Carolina Teacher Performance Appraisal Instrument. Following the summary of the study and comments by the researcher, findings and conclusions for each of the three research questions will be addressed. Specific recommendations and implications of the study for future work in teacher evaluation and administrators' roles conclude the chapter.

SUMMARY

Perceptions of administrators using the North Carolina Teacher Performance Appraisal Instrument to rate five observable teaching functions were examined in this study. The study was conducted in a small rural school system in northeastern North Carolina. School principals and central office administrators participated in a teacher evaluation workshop which focused on reaching agreement using the
six-point rating scale mandated for assessment of teachers in the state.

Two problems were addressed: the differences in public school administrators' perceptions of the ratings on the North Carolina Teacher Performance Appraisal Instrument and consistency in ratings for the first five teaching functions on the instrument. The purposes of the study were to investigate administrators' observations and reasons for rating as they did, and to examine the administrators' perceptions of the various ratings. The process of evaluating teaching behaviors and reaching agreement on the ratings using the North Carolina Teacher Performance Appraisal Instrument was also addressed in this inquiry. Attention was directed to three research questions: 1. Do administrators rate teaching performance consistently using the six-point rating scale of the North Carolina Teacher Performance Appraisal Instrument? 2. How do administrators perceive the ratings on the performance appraisal instrument? 3. Do administrators have confidence in their ratings of teaching performance?

Descriptive methods were used to explore administrators' perceptions and the rating process. A teacher evaluation workshop was conducted with building level principals and central office administrators in a small school system in North Carolina. The three-day workshop provided the
opportunity for administrators to view videotaped lessons of teachers of various grade levels and subjects, and to rate five observable teaching functions which are included on the North Carolina Teacher Performance Appraisal Instrument. The teaching behaviors rated were Management of Instructional Time, Management of Student Behavior, Instructional Presentation, Monitoring of Student Performance, and Instructional Feedback.

The workshop began with the viewing of a videotaped lesson presented to the group of participants. Each participant rated the five teaching behaviors, then the total group discussed their ratings and the basis for each rating. For each of the next three tapes, the lesson was viewed and rated individually, then discussed by small groups of administrators to obtain consensus ratings. Observers watched and recorded interactions among participants for the fifth, sixth, and seventh tapes as they followed the same procedures. The eighth tape, the same lesson as the first tape, was viewed and rated individually by the participants so that any changes in evaluators’ ratings from the beginning to the end of the workshop could be measured. Standard deviations and means were calculated for each teaching function for the lesson ratings for the purpose of measuring how close together the individuals rated.
Following the workshop, participants were interviewed individually to record their reactions to the workshop rating activities and to the teacher evaluation process. Teachers in the same school system completed a written questionnaire about their concerns and experiences with the teacher evaluation process. Characteristics of the administrators including grade level of experience, subject area, race, gender, and age/experience, were matched with each of the videotaped lessons to check for patterns of rater bias.

COMMENTS

The literature search revealed information which supports the continuing development and use of teacher evaluation to improve instruction in the nation's schools. Instructional leaders in effective schools observe teachers, give specific feedback to them, and lead in the development of professional improvement goals. Administrators who communicate high expectations for success for teachers and students provide teachers with direction and feedback on performance. A meaningful teacher evaluation system is present in schools which are effective.

Teacher evaluation today has taken on a more important role as the results are used to improve teaching and to make personnel decisions. Competence in using an evaluation
system for the purpose for which it was developed is a must. The literature supports the continuing development of administrators' skills in assessing teacher performance. Knowledge of what constitutes good teaching is necessary to use a system. Reliable, consistent ratings are emphasized.

Role conflict of administrators who evaluate and lead the instructional staff of a school has been noted throughout current research. As an instructional leader, a principal also must be a coach, displaying empathy while encouraging improvement on the job. As an evaluator, the administrator must be willing to observe and rate accurately in order to make personnel decisions.

Research on the North Carolina Teacher Performance Appraisal System revealed that adjustments and changes continue to be made in the process used to evaluate teachers. The state-mandated system is based on research and pilot studies which have taken place during the last ten years. As the prospect nears for a career development plan which will enable teachers to earn merit pay, emphasis is placed on the importance of reliable and consistent evaluations of the identified effective teaching behaviors. Progress continues in the direction of improving the process to attain a system which is palatable for teachers and administrators who are responsible for providing effective instruction for students.
FINDINGS AND CONCLUSIONS

In this section, findings and conclusions are presented for each of the research questions on which the study is based.

QUESTION 1: Do administrators rate teaching performance consistently using the six-point rating scale of the North Carolina Teacher Performance Appraisal Instrument?

Findings:

It was found that the administrators rated within an acceptable tolerance limit of plus or minus one point for the five observable teaching functions. There was closer agreement on the teaching behaviors which were rated below satisfactory than on the behaviors rated satisfactory and higher. As a group, the central office administrators had closer agreement in ratings than did the principals. Just over half (56%) of all participants moved their scores closer to the group mean ratings from the beginning of the workshop to the end of the workshop.

Conclusions:

Because of the tendency of evaluators to use the scores in the middle of the rating scale rather than the highest and lowest scores, consistency as defined using the plus or minus one point tolerance will continue to be evident. Accuracy in ratings is not assured because of consistency, a
potential problem in using this teacher evaluation process to make meaningful suggestions for improvement of teaching skills. Personnel decisions based primarily on the ratings on this instrument are subject to question.

QUESTION 2: How do administrators perceive the ratings on the performance instrument?

Findings:

Differences in the perceptions of what a particular rating means were evident. Administrators consistently stated definitions for ratings which were not unique to a singular rating. The same words and similar terminology were used to describe ratings which designated different levels of performance. When called on to tell why a certain rating was selected, administrators were unable to specify exact teaching behaviors that discriminated between ratings.

Discriminating between the ratings of At Standard, Above Standard, and Well Above Standard, was more difficult than deciding if a behavior was acceptable (At Standard or higher) or unacceptable (Below Standard, Unsatisfactory). Agreement of ratings was highest on Functions 1 and 2, Management of Instructional Time and Management of Student Behavior. The least agreement was for Function 5, Instructional Feedback.
Conclusions:

The absence of a model of good or perfect teaching required raters to interpret what they observe within their own limited experiences. The raters were unable to specify exact behaviors they were looking for when they observed the teachers.

The evaluators never became "learners." They never moved from the role of observer to the role of learner to evaluate whether or not the lesson was effective in teaching a student. They never said what made the teaching good, what made it not good. Specifics were avoided in describing teaching behaviors.

QUESTION 3: Do administrators have confidence in their ratings of teaching performance?

Findings:

Administrators willingly changed ratings to reach agreement. They had little difficulty reaching consensus on ratings when working in teams because of the eagerness to change their ratings to avoid conflict. Discussions centered around the reasons for particular ratings. Little effort was made to convince others to change to a particular rating when conflicting reasons were stated for the selection of a rating.
Conclusions:

The administrators did not use the rating scale for its purpose. There was a tendency to stay in the "safe zone" of ratings, and not use the extremes of 1 (Unsatisfactory) and 6 (Superior), in which justification for the ratings would be expected.

The evaluators did not exhibit self-confidence as raters. They were over-polite, taking on a cooptative character, acting as a member of the group, rather than taking a stand on differences of opinions in ratings. The administrators tried to please everyone and not hurt anyone’s feelings by evaluating harshly. Educational administrators tend to accept the opinions of others, possibly as a demonstration of their negotiation skills, a very important skill in school administration.

OTHER FINDINGS:

A serendipitous finding in this study was the value of the teacher evaluation workshop as a staff development activity for administrators. The participating administrators had been trained in the Teacher Performance Appraisal System, but felt that more training was needed to improve their skills. Practice in consensus rating of teacher performance proved valuable in building confidence in evaluation skills. The opportunity for administrators to
interact in a non-threatening setting was provided. Requests were made by the administrators for future staff development activities using the format and methods of the teacher evaluation workshop. Professional growth and camaraderie resulted from the workshop in which administrators could exchange ideas and share concerns with one another.

RECOMMENDATIONS

Specific recommendations are as follows:

- To emphasize the importance of performance appraisal as a means of ensuring good instruction at the school level, the evaluation of teachers should be weighted heavier in the Principal’s Performance Appraisal Instrument which is used to evaluate principals. At present teacher evaluation is rated as a small part of one of the twelve functions on the principal’s evaluation. Teacher evaluation should be a separate function on the instrument.

- More training is needed to teach evaluators to discriminate between the ratings which indicate above standard scores.

- Further development of model teaching behavior is needed to teach administrators what good teaching is.
Periodic workshops for renewal of evaluation skills and to build confidence in their own skills should be required for administrators who evaluate teachers.

Practice in oral and written communication to support rating decisions with specific terminology should be incorporated in the training for evaluators.

Funding for continuous training for expertise in teacher evaluation must become a regular part of the budget for school systems.

Extend opportunities for staff development in consensus rating. Develop and implement workshops in using consensus methods for rating teachers with TPAI.

IMPLICATIONS

This study examined the perceptions of administrators who used the North Carolina Teacher Performance Appraisal Instrument to evaluate teachers. The six ratings used to designate level of performance for each of five observable teaching behaviors were not interpreted the same way by the evaluators. Continuing to use the rating scale without further refinement of evaluators' discrimination skills will lead to a false sense of accuracy and confidence in final ratings.
Consistency of ratings was evident when examining the scores for each function, but reasons for the scores and observations which led to the scores were different. There was little discrimination exhibited in the above satisfactory scores. Using the four ratings of At Standard and above as now interpreted does not indicate mastery of the rating scale as intended to show different levels of expertise. Further study needs to be done to develop methods to teach evaluators to discriminate among the higher ratings. As a tool in assessing teaching performance for any of various reasons — merit pay, national certification, promotions, job advancement — the system must be used by administrators who exhibit expertise in observing and rating using all levels of the scale.

When principals' performance appraisals reflect their skills in teacher observation, evaluation, and assistance in the development of professional teaching skills, administrators will place more importance on the development and continuation of adequate assessment skills. Funding should be sought to reward administrators who exhibit skill and accuracy in evaluating teaching performance. Administrators who do not evaluate adequately using the chosen instrument, should be removed from the job.

Team consensus rating is a practice which should be used to alleviate problems of uncertainty in evaluator
scoring. Increased confidence in the final ratings, and ability to discuss more complete and accurate findings with the teacher are reasons to use team consensus ratings for evaluation of teaching behaviors. Although the consensus process was not a primary objective of this study, consensus rating was found to be a well-received activity which participating administrators liked. The opportunity to discuss reasons for ratings and interpretations of scores was a positive outcome of the teacher evaluation workshop.

There must be a concerted effort on the part of the North Carolina Department of Public Instruction and local school systems to provide additional training for administrators who are responsible for evaluating teachers. Teachers must be assured that evaluations will be consistent and fair regardless of the evaluator using the state performance appraisal system. Research leading to the development of training programs for evaluators is necessary to teach them to use the full scale of ratings with confidence.

Evaluators must be required to demonstrate competence in using the North Carolina Teacher Performance Appraisal Instrument before being allowed to evaluate teaching behaviors in local schools. Competence must include understanding and use of the total rating scale, mastery of specific oral and written feedback to the teacher, and
consistency in ratings. Demonstration of this competency every three to five years should be required for renewal of the license or certificate authorizing the educator as an evaluator of teaching behaviors in the classroom.

Further attempts to develop and fine tune the teacher appraisal instrument in the absence of additional extensive and rigorous training of evaluators is useless. No matter how perfect the measurement tool is, the expertise of the user continues to limit its effectiveness.
LIST OF REFERENCES


Edmonds, Ronald R. et al. Search for Effective Schools: The Identification and Analysis of City Schools That are Instructionally Effective for Poor Children, 1977.


North Carolina Department of Public Instruction [NCDPI].


§ 115C-326. Performance standards and criteria for professional employees; law suits arising out of this section.

(a) The State Board of Education, in consultation with local boards of education, shall develop uniform performance standards and criteria to be used in evaluating professional public school employees. It shall develop rules and regulations to recommend the use of these standards and criteria in the employee evaluation process. The performance standards and criteria shall be adopted by the Board by July 1, 1982, and may be modified in the discretion of the Board.

Local boards of education shall adopt rules and regulations by July 1, 1982, to provide for annual evaluation of all professional employees defined as teachers by G.S. 115C-325(a)(6). Local boards may also adopt rules and regulations requiring annual evaluation of other school employees not specifically covered in this section. Rules and regulations adopted by local boards shall utilize performance standards and criteria adopted by the State Board of Education pursuant to the first paragraph of this section; however, the standards and criteria used by local boards are not to be limited by those adopted by the State Board of Education.

(b) If any claim is made or any legal action is instituted against an employee of a local school administrative unit on account of an act done or an omission made in the course of the employee's duties in evaluating employees pursuant to this section, the local board of education, if the employee is held not liable, shall reimburse the employee for reasonable attorney's fees.

(c) The State Board of Education shall recommend to the General Assembly by December 1, 1986, a program to remedy deficiencies and difficulties revealed through the evaluation process required by this section and to develop new skills on the part of classroom teachers. (1979, 2nd Sess., c. 1137, s. 35; 1981, c. 423, s. 1; c. 559, s. 29.12; 1981 (Reg. Sess., 1982), c. 1282, s. 32.1; 1983 (Reg. Sess., 1984), c. 1103, s. 10.)
Appendix B

TEACHER PERFORMANCE APPRAISAL SYSTEM

INSTRUCTIONS

1. Based on the evidence from observation and discussion, the
evaluator is to rate the teacher's performance with respect to
the 8 major functions of teaching listed below.
2. The evaluator is encouraged to add pertinent comments at the
end of each major function.
3. The teacher is provided an opportunity to react to the
evaluator's ratings and comments.
4. The evaluator and the teacher must discuss the results of
the appraisal and any recommended action pertinent to it.
5. The teacher and the evaluator must sign the instrument in the
assigned spaces.
6. The instrument must be filed in the teacher's personnel
folder.
7. The rating scale will be as follows:

   Level of Performance

6. Superior

Performance within this function area is consistently outstanding. Teaching prac-
tices are demonstrated at the highest level of performance. Teacher continuously
seeks to expand scope of competencies and constantly undertakes additional, appro-
priate responsibilities.

5. Well Above Standard

Performance within this function area is frequently outstanding. Some teaching
practices are demonstrated at the highest level while others are at a consistently
high level. Teacher frequently seeks to expand scope of competencies and often
undertakes additional, appropriate responsibilities.

4. Above Standard

Performance within this function area is frequently high. Some teaching practices
are demonstrated at a high level while others are at a consistently adequate/acceptable
level. Teacher sometimes seeks to expand scope of competencies and occasionally
undertakes additional, appropriate responsibilities.

3. At Standard

Performance within this function area is consistently adequate/acceptable. Teaching
practices fully meet all performance expectations at an acceptable level. Teacher
maintains an adequate scope of competencies and performs additional responsibilities
as assigned.

2. Below Standard

Performance within this function area is sometimes inadequate/unacceptable and needs
improvement. Teacher requires supervision and assistance to maintain an adequate
scope of competencies, and sometimes fails to perform additional responsibilities
as assigned.

1 Unsatisfactory

Performance within this function area is consistently inadequate/unacceptable and
most practices require considerable improvement to fully meet minimum performance
expectations. Teacher requires close and frequent supervision in the performance
of all responsibilities.
1. Major Function: Management of Instructional Time

1.1 Teacher has materials, supplies and equipment ready at the start of the lesson or instructional activity.

1.2 Teacher gets the class started quickly.

1.3 Teacher gets students on task quickly at the beginning of each lesson or instructional activity.

1.4 Teacher maintains a high level of student time-on-task.

Comments

2. Major Function: Management of Student Behavior

2.1 Teacher has established a set of rules and procedures that govern the handling of routine administrative matters.

2.2 Teacher has established a set of rules and procedures that govern student verbal participation and talk during different types of activities—whole-class instruction, small group instruction, etc.

2.3 Teacher has established a set of rules and procedures that govern student movement in the classroom during different types of instructional activities.

2.4 Teacher frequently monitors the behavior of all students during whole-class, small group, and seat work activities and during transitions between instructional activities.
2.5 Teacher stops inappropriate behavior promptly and consistently, yet maintains the dignity of the student.

Comments

3. Major Function: Instructional Presentation

3.1 Teacher begins lesson or instructional activity with a review of previous material.

3.2 Teacher introduces the lesson or instructional activity and specifies learning objectives when appropriate.

3.3 Teacher speaks fluently and precisely.

3.4 Teacher presents the lesson or instructional activity using concepts and language understandable to the students.

3.5 Teacher provides relevant examples and demonstrations to illustrate concepts and skills.

3.6 Teacher assigns tasks that students handle with a high rate of success.

3.7 Teacher asks appropriate levels of questions that students handle with a high rate of success.

3.8 Teacher conducts lesson or instructional activity at a brisk pace, slowing presentations when necessary for student understanding but avoiding unnecessary slowdowns.

3.9 Teacher makes transitions between lessons and between instructional activities within lessons efficiently and smoothly.
3.10 Teacher makes sure that the assignment is clear.

3.11 Teacher summarizes the main point(s) of the lesson at the end of the lesson or instructional activity.

Comments ____________________________________________________________

______________________________________________________________________

______________________________________________________________________

______________________________________________________________________

4. Major Function: Instructional Monitoring of Student Performance

4.1 Teacher maintains clear, firm and reasonable work standards and due dates.

4.2 Teacher circulates during classwork to check all students' performance.

4.3 Teacher routinely uses oral, written, and other work products to check student progress.

4.4 Teacher poses questions clearly and one at a time.

Comments ____________________________________________________________

______________________________________________________________________

______________________________________________________________________

5. Major Function: Instructional Feedback

5.1 Teacher provides feedback on the correctness or incorrectness of in-class work to encourage student growth.
5.2 Teacher regularly provides prompt feedback on assigned out-of-class work.

5.3 Teacher affirms a correct oral response appropriately, and moves on.

5.4 Teacher provides sustaining feedback after an incorrect response or no response by probing, repeating the question, giving a clue, or allowing more time.

Comments

6. Major Function: Facilitating Instruction

6.1 Teacher has an instructional plan which is compatible with the school and system-wide curricular goals.

6.2 Teacher uses diagnostic information obtained from tests and other assessment procedures to develop and revise objectives and/or tasks.

6.3 Teacher maintains accurate records to document student performance.

6.4 Teacher has instructional plan that matches/aligns objectives, learning strategies, assessment and student needs at the appropriate level of difficulty.

6.5 Teacher uses available human and material resources to support the instructional program.
7. Major Function: Communicating Within The Educational Environment

7.1 Teacher treats all students in a fair and equitable manner.

7.2 Teacher interacts effectively with students, co-workers, parents, and community.

Comments

8. Major Function: Performing Non-Instructional Duties

8.1 Teacher carries out non-instructional duties as assigned and/or as need is perceived.

8.2 Teacher adheres to established laws, policies, rules, and regulations.

8.3 Teacher follows a plan for professional development and demonstrates evidence of growth.

Comments
### Appendix C

**FORMATIVE OBSERVATION DATA INSTRUMENT**

**Instruction:** Use this sheet to record anecdotal and sequentially these events which occur during observation. Be sure to code each instance of a TPAI practice as follows:

- appropriate use of practice
- strong or positive use of practice
- weak or negative use of practice

<table>
<thead>
<tr>
<th>Practice</th>
<th>Time</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. INSTRUCTIONAL TIME</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1 Materials ready</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.2 Class started quickly</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.3 Gets students on task</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.4 Maintains high time-on-task</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>2. STUDENT BEHAVIOR</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.1 Rules—Administrative Matters</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.2 Rules—Verbal Participation/Talk</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.3 Rules—Movement</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.4 Frequently monitors behavior</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.5 Stops inappropriate behavior</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>3. INSTRUCTIONAL PRESENTATION</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.1 Begins with review</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.2 Introduces lesson</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.3 Speaks fluently</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.4 Lesson understandable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.5 Provides relevant examples</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.6 High rate of success on tasks</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.7 Appropriate level of questions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.8 Brisk pace</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.9 Efficient, smooth transitions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.10 Assignment clear</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.11 Summarizes main points</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>4. INSTRUCTIONAL MONITORING</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.1 Maintains deadlines, standards</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.2 Circulates to check student performance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.3 Uses oral, written work products to check progress</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.4 Questions clearly and one at a time</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>5. INSTRUCTIONAL FEEDBACK</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.1 Feedback on in-class work</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.2 Prompt feedback on out-of-class work</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.3 Affirms correct answer quickly</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.4 Sustaining feedback on incorrect answers</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix D

FORMATIVE OBSERVATION DATA ANALYSIS

Based on your observations, address each of the following areas using statements which accurately reflect the quality of performance documented by your raw data.

<table>
<thead>
<tr>
<th>MANAGEMENT OF INSTRUCTIONAL TIME</th>
<th>MANAGEMENT OF STUDENT BEHAVIOR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>INSTRUCTIONAL PRESENTATION</th>
<th>INSTRUCTIONAL MONITORING</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lesson Design</td>
<td></td>
</tr>
<tr>
<td>Focus and Review</td>
<td></td>
</tr>
<tr>
<td>Lesson Objective</td>
<td></td>
</tr>
<tr>
<td>Teacher Input</td>
<td></td>
</tr>
<tr>
<td>Guided Practice</td>
<td></td>
</tr>
<tr>
<td>Independent Practice</td>
<td></td>
</tr>
<tr>
<td>Closure</td>
<td></td>
</tr>
<tr>
<td>Other Practices</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>INSTRUCTIONAL FEEDBACK</th>
<th></th>
</tr>
</thead>
</table>
FACILITATING INSTRUCTION (COMPLETE ONLY FOR INITIALLY CERTIFIED EMPLOYEE'S ANNUAL PERFORMANCE APPRAISAL)

COMMUNICATING WITHIN THE EDUCATIONAL ENVIRONMENT (COMPLETE ONLY FOR INITIALLY CERTIFIED EMPLOYEE'S ANNUAL PERFORMANCE APPRAISAL)

NON-INSTRUCTIONAL DUTIES (COMPLETE ONLY FOR INITIALLY CERTIFIED EMPLOYEE'S ANNUAL PERFORMANCE APPRAISAL)

From the classifications of raw data on this sheet, list the strengths observed in this lesson, and prioritize the areas needing improvements.

Strengths:

Areas That Need Improvement (Prioritize):

I have been furnished a copy of this analysis sheet. In addition, I have been given access to the raw data from which this analysis was prepared. I understand that the raw data will not be placed in my personnel file.

____ I have requested and been given a copy of the raw data.

____ I have not requested and do not wish a copy of the raw data.

Signed ___________________________ / Date

Teacher ___________________________ / Date

Principal ___________________________ / Date
## Professional Development Plan

<table>
<thead>
<tr>
<th>Certification</th>
<th>Initial</th>
<th>Continuing</th>
<th>Expiration Date of Certificate</th>
<th>School</th>
<th>Date of Professional Development Plan</th>
<th>Goal(s)</th>
</tr>
</thead>
</table>

### Major Functions

1. Management of Instructional Time
2. Management of Student Behavior
3. Instructional Presentation
4. Instructional Monitoring
5. Instructional Feedback
6. Facilitating Instruction
7. Interacting Within Educational Environment
8. Performing Non-Instructional Duties

<table>
<thead>
<tr>
<th>Practices and/or Strengths</th>
<th>Activities (Strategies)</th>
<th>Evidence of Completion</th>
<th>Resources</th>
<th>Target Date</th>
<th>Completion Date</th>
</tr>
</thead>
</table>

**ACCOMPLISHMENT:**

- Fully Accomplished
- Partially Accomplished
- Not Accomplished

**REASON**

<table>
<thead>
<tr>
<th>Date</th>
<th>SUPERVISOR’S COMMENTS</th>
</tr>
</thead>
</table>

**EMPLOYEE’S COMMENTS**

**SIGNATURES**

<table>
<thead>
<tr>
<th>Initial Conference</th>
<th>SUPERVISOR’S SIGNATURE</th>
<th>DATE</th>
<th>EMPLOYEE’S SIGNATURE</th>
<th>DATE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*If employee has initial certification, mentor and/or support team members should also sign.*
Appendix F

PRINCIPAL'S PERFORMANCE APPRAISAL INSTRUMENT
LEVELS OF PERFORMANCE

6 Superior
Performance within this function area is documented as being consistently outstanding. Indicators and supporting evidence demonstrate that, without exception, the principal/assistant principal is performing regularly and continuously at the highest level of expectancy.

5 Well Above Standard
Performance within this function area is frequently outstanding. Some indicators and supporting evidence are demonstrated at the highest level, while others are at a consistently high level. The principal/assistant principal frequently seeks to expand the scope of his/her competencies and often undertakes additional appropriate responsibilities.

4 Above Standard
Performance within this function area is frequently high. Some indicators and supporting evidence are demonstrated at a high level, while others are at a consistently adequate/acceptable level. The principal/assistant principal sometimes seeks to expand the scope of his/her competencies and occasionally undertakes additional, appropriate responsibilities.

3 At Standard
Performance within this function area is consistently adequate/acceptable. Indicators and supporting evidence fully meet all performance expectations at an adequate/acceptable level. The principal/assistant principal maintains an adequate scope of competencies and performs additional responsibilities as assigned.

2 Below Standard
Performance within this function area is sometimes inadequate/unacceptable and needs improvement. The principal/assistant principal requires supervision and assistance to maintain an adequate scope of competencies, and sometimes fails to perform additional responsibilities as assigned.

1 Unsatisfactory
Performance within this function area is consistently inadequate and/or unacceptable and most indicators and supporting evidence require considerable improvement to fully meet minimum performance expectations. The principal/assistant principal requires close and frequent supervision in the performance of all responsibilities.
PRINCIPAL'S PERFORMANCE APPRAISAL INSTRUMENT

Principal's Name ____________________________

School _____________________________________

INSTRUCTIONS:

1. Based on the evidence from observation and discussion, the evaluator is to rate the principal's performance with respect to the 12 major functions of principalship.

2. The principal is provided an opportunity to react to the evaluator's ratings and comments.

3. The evaluator and the principal must discuss the results of the appraisal and any recommended action pertinent to such.

4. The principal and the evaluator must sign the instrument in the assigned spaces.

5. The instrument must be filed in the principal's personnel folder.

Rating Scale
(Please Check)

HAJOR FUNCTIONS:

IA MAJOR FUNCTION: Plans a comprehensive instructional program for the school, including co- and extra-curricular aspects.

1. Identifies annual objectives that specify what the principal intends to accomplish in his/her school for the coming year.

2. Involves faculty, central office staff, curriculum specialists, parents, students, and other resource personnel in curriculum planning and program development appropriately.

3. Provides opportunities whereby students can have appropriate input into the educational program.
4. Establishes clearly-defined disciplinary procedures that have been communicated to and are understood by parents, students, staff, and the community.

5. Plans a comprehensive program for extra- and co-curricular non-instructional activities that support the instructional program, directly and/or indirectly.

6. Plans for appropriate utilization of community resources to enrich the school program.

7. Defines roles for staff members for completing tasks.

8. Arranges staff development programs that provide opportunities for professional growth.

Major Function: Implements a comprehensive instructional program for the school, including its co- and extra-curricular aspects.

1. Encourages and provides opportunities for faculty, central office staff, curriculum specialists, parents, students, and other resource personnel appropriately to participate in the school program.

2. Makes use of supervisor or administrative assistance to improve performance.

3. Provides in-service programs for personnel which enhance the quality of the instructional program.

4. Implements a student discipline program consistent with local, state, and federal laws and policies.

5. Approves and facilitates appropriate co- and extra-curricular activities that support and enhance the instructional program.
6. Uses community resources to enrich the school program.

**IC MAJOR FUNCTION:** Supervises the delivery of a comprehensive instructional program including co- and extra-curricular aspects.

1. Provides supervision to promote staff growth and the attainment of program goals.

2. Ensures that each teacher has developed or listed instructional objectives related to the subject matter for a given class.

3. Ensures that teachers utilize student data in planning lessons and courses of instruction.

4. Monitors a system for recording student performance and participation in the total school program.

5. Ensures that the discipline program is consistently enforced.

6. Monitors the use of the community resource program.

7. Delegates responsibility for completing tasks.

8. Follows-up to ensure that delegated responsibilities are adequately performed.

**IC MAJOR FUNCTION:** Evaluates the comprehensive school program including co- and extra-curricular aspects.

1. Evaluates the instructional program, using appropriate procedures and/or data to ensure attainment of instructional goals.

2. Evaluates instructional staff, using criteria and processes adopted by the board of education.
3. Evaluates community resource program, using appropriate procedures.

4. Evaluates discipline program, using appropriate procedures and/or data.

5. Evaluates extra-curricular program, using appropriate procedures, to ascertain goals attainment.

6. Evaluates co-curricular program, using appropriate procedures, to ascertain goals attainment.

2A MAJOR FUNCTION: Plans a comprehensive program for resource management.

1. Prepares school's budgetary requests.

2. Prepares a human resource program that reflects the goals and needs of the school's comprehensive programs.

3. Develops a procedure for identifying needed materials, supplies, and equipment.

4. Involves the staff in setting priorities concerning expenditures for instructional supplies.

5. Develops appropriate work schedules, class schedules, building use schedules, and other necessary schedules.

6. Develops a procedure for delivery of school's transportation services.

2B MAJOR FUNCTION: Implements the program for resource management.

1. Implements scheduled maintenance and safety inspection programs to assure proper maintenance of school plant and grounds.
2. Implements procedures for the appropriate inventorying and distribution of materials, supplies, and equipment.

3. Monitors expenditure of funds and assumes accountability for all monies.

4. Assures coordination of the school's financial operations with those of the school system.

5. Implements appropriate work schedules, class schedules, and building use schedules and other necessary schedules.

6. Conducts fire-drills and other emergency preparation procedures in accordance with state and local laws and policies.

7.Uses community resources that support the total school program.

8. Cooperates with the community in use of school facilities for community activities.

2C MAJOR FUNCTION: Supervises the program for resource management.

1. Delegates responsibility for completing tasks.

2. Follows up to ensure that delegated tasks are adequately performed.

3. Provides supervision to promote growth to attain goals of resource management program.

2D MAJOR FUNCTION: Evaluates the program for resource management and its implementation.

1. Evaluates non-instructional staff using appropriate instruments and procedures.
2. Evaluates attainment of goals of resource management.

3A MAJOR FUNCTION: Plans a comprehensive communications program.

1. Develops procedures for receiving information from individuals and groups within and outside the school.

2. Develops procedures for distributing routine information appropriately.

3. Develops procedures for distributing special information appropriately.

3B MAJOR FUNCTION: Implements a comprehensive communications program.

1. Disseminates classification, promotion, retention, suspension and expulsion policies, procedures, and criteria for students.

2. Communicates academic and behavioral expectations for students to appropriate groups.

3. Communicates clearly-defined disciplinary procedures to parents, students, staff, and the community.

4. Interprets and carries out the policies established by the local board, State Board of Education, North Carolina School Law, and federal law.

5. Complies with established lines of authority.

6. Completes records, reports, inventories, requisitions, and budgets.
7. Ensures that staff understands communication plan and procedures. 

8. Uses written communication skills effectively. 

9. Uses oral communication skills effectively. 

10. Uses conflict management and resolution skills to facilitate school operations. 

11. Interprets the school program, objectives, and policies to the community. 

3C MAJOR FUNCTION: Supervises implementation of the comprehensive communications program. 

1. Oversees all special school events that are designed to interpret the school program to the community. 

2. Monitors implementation and use of communication procedures. 

3. Provides information and support to newly-assigned staff and assists in their professional development. 

4. Provides opportunities whereby newly-assigned students can have appropriate input into the educational program. 

3D MAJOR FUNCTION: Evaluates the efficacy of the comprehensive communications program. 

1. Monitors and resolves any problems in the communications program. 

2. Evaluates specific communications strategies. 

3. Solicits input from others about the usefulness of the communications program.
4. Evaluates the program in terms of its goals.

Evaluator's Summary Comments

Principal's Reaction to Evaluation

Evaluator's Signature and Date  Principal's Signature and Date

"Signature indicates that the written evaluation has been seen and discussed."
Appendix G

"The Brandt Report"

A REVIEW WITH RECOMMENDATIONS OF THE
NORTH CAROLINA TEACHER PERFORMANCE
APPRAISAL INSTRUMENT

Submitted to:
THE EDUCATION SUBCOMMITTEE
OF THE JOINT LEGISLATIVE COMMISSION ON
GOVERNMENT OPERATIONS

North Carolina General Assembly
State Legislative Building
Raleigh, NC 27611

Submitted by:
Richard M. Brandt (Chair)
Daniel L. Duke
Russell L. French
Edward F. Iwanicki

May 21, 1988
EXECUTIVE SUMMARY

Our study and review of the TPAI leads us to one overriding conclusion. It is a quality instrument, one that is highly suited to its purposes. The recommendations we list below represent suggestions for improving it still further, especially to enhance its effectiveness as a basic tool for making career ladder decisions and for improving teaching. In no way do these recommendations imply fundamental deficiencies that would call for its elimination as the state teacher appraisal instrument. We consider our recommendations suggestions for fine-tuning an already good system of performance appraisal.

Likewise, our limited commentary on the PPAI and other specialty measures does not suggest they are inadequate, only that we had less to examine because of their recent development and lack of widespread use. On the surface, they look promising and, with the few exceptions we have noted, consistent with the TPAI in purpose and procedure.

We summarize our recommendations in the order in which they are presented in the text.

CRITERIA

1. Rearrange some functions and teaching practices to a total of twelve, ten of which relate directly to teaching and are observable in the classroom.

Specifically:
Reverse the order of functions 6 and 7.

Rename new function 6 "Communication and Interaction in the Classroom." Include function 6 as one of those to be observed in the classroom, and list relevant teaching practices as follows:

6.1 Teacher speaks fluently and precisely (currently 3.3)

6.2 Teacher presents the lesson or instructional activity using concepts and language understandable to the students (currently 3.4)

6.3 Teacher interacts effectively with students (currently part of 7.2)

6.4 Teacher treats all students in a fair and equitable manner (currently 7.1)

Divide current function 3 (Instructional Presentation) into five specific functions as follows:

- Initiating Instruction
- Motivating Students
- Managing Routines and Transitions
- Presenting Accurate and Appropriate Content
- Providing Closure

Teaching practices for these latter functions would be taken from current function 3 (excluding 3.3 and 3.4) and from such instruments as the Connecticut Competency Instrument.

Expand the literature search for effective
teaching practices to include coaching, modeling, cooperative and mastery learning models of teaching in addition to process-product research on direct instruction

- Encourage local districts to identify practices that reflect local needs for the two functions that are less observable in the classroom and relate less directly to instruction, i.e.:
  - Facilitating Instruction (currently F6)
  - Performing Non-Instructional Duties (currently F8)

FODI AND POST-OBSERVATION CONFERENCE

2. In addition to scripting, consider collecting other kinds of observational data on a pilot basis such as student participation and on-task behavior, teaching patterns, and other classroom action. While we do not recommend changing scripting procedures at this time, other kinds of data might well be needed in the future to assist in strengthening the objectivity and power of the TPAI to differentiate between good and superior teachers.

FODI

3. Provide teachers with copies of all descriptive data (FODIs and any other descriptive instruments used) in advance of the post-observation conference.
FODA
1. Provide space on the FODA form for the evaluatee to respond to observations and quality assessments made by the evaluator.

TRAINING
5. Provide evaluators with specific training in conferencing skills and using the FODI and FODA in post-observation conferences.

SCALE STEPS
5. Maintain the six-point rating scale.

RATING PROCEDURE
7. Follow a step-by-step procedure in making summative judgments:
   a. Is performance at least at standard?
   b. If so, is it at or somewhere above standard?
   c. If above, is it merely above standard (4) or well above (5) or superior (6)?

PDP
3. Revise the Professional Development Plan form somewhat to reflect more highly individualized, professional development beyond basic skill competency, especially for above standard teachers.

PORTFOLIO
9. Develop a form and procedure for recording and analyzing portfolio contents systematically in connection with the non-classroom functions. Such
forms and procedures should be tailored, however, to local district priorities.

**INTERVIEW**

10. Develop a structured interview format for collecting data from principals, special groups (e.g. school psychologists, counselors, etc.) and teachers in a systematic, consistent fashion about functions not readily observed in the classroom.

**CERTIFICATION OF EVALUATORS**

11. Require all evaluators to be certified for this role by passing a performance-based test of their competency to use the TPAI accurately and consistently. Evaluators should be required to demonstrate their competence on a regular basis against criterion measures. They should not be permitted to function as evaluators unless they are certified and remain certified as evaluators.

**CAREER LADDER DECISIONS**

12. Restrict the decision-making use of the TPAI to certification and career levels I and II decisions. Do not attempt to use it as the primary measure for career level III decisions.

**SUMMATIVE RATINGS BY CONSENSUS**

13. Require the joint participation of OEs, principals and other evaluators who are involved in the data collection process in the final summative judgment.
Training in achieving data-based consensus is also recommended.

**CONSECUTIVE OBSERVATIONS**

14. Experiment with observing teachers up to four consecutive days teaching the same class for possible inclusion of such a requirement for career level II applicants.

**JOINT OBSERVATIONS/ANALYSES**

15. Create panels of administrators and teachers to review and rate samples of FODIs and FODAs for comparability. Conduct joint observations, gather data, and share analyses to enhance the consistency of instrument use and interpretation.

**RELIABILITY STUDIES**

16. Conduct various studies of rating reliability at both the state and district levels to assess the consistency across raters and the stability of teacher patterns from one time to another.

**VALIDITY STUDIES**

17. Conduct validity studies within districts of (a) achievement gain scores of students taught by teachers rated at standard (#3) versus those taught by teachers rated 5 and 6; and (b) student gains in pilot districts versus those in comparable non-pilot districts.
UNIVERSITY RESEARCH ASSISTANCE

18. Request assistance from university faculty and provide graduate student financial support for the conduct of TPAI reliability and validity studies in the pilot effort.

WORD PROCESSORS

19. Provide evaluators with word processors and appropriate training as needed to assist them in writing FODIs and FODAs.

INCREASE OF OEs

20. Reduce the ratio of teachers to OEs to 48 to 1 in order to increase the amount of assistance available to teachers for remediating deficiencies and improving their instruction.

GROWTH BEYOND COMPETENCY

21. Focus career level III criteria on leadership functions and growth beyond competency in basic skills. Consider a more extensive range of criteria and measures.

PPAI/OTHER SPECIALTIES

22. Review the several instruments for principals and other specialty personnel for parallel coverage and structure. Improve the consistency in format and comprehensiveness wherever possible.

PRINCIPALS AS EVALUATORS

23. Assign greater weight to the performance of principals
as teacher evaluators than to other functions in the PPAI as one way to improve the TPAS.

EXTRA RESOURCES FOR TPAI REVISION

24. Provide extra assistance for the Department of Public Instruction to help implement the 23 recommendations above during summer 1988.
Appendix H

Teacher Evaluation Workshop Agenda and Script

TEACHER EVALUATION WORKSHOP AGENDA

OBSERVER TRAINING

WORKSHOP AGENDA:

SESSION 1
Introduction
Procedures
View Tape 1
Rate Individually
Discuss as Total Group
Collect Data

View Tape 2
Rate Individually
Form Two-member Teams
Discussion and Team Ratings
Collect Data

View Tape 3
Rate Individually
Form Three-member Teams
Discussion and Team Ratings
Collect Data

SESSION 2

View Tape 4
Rate Individually
Form Four-member Teams
Discussion and Team Ratings
Collect Data

CONSENSUS RATING ACTIVITIES

View Tape 5
Rate Individually
Form Two-member Teams
Discussion and Team Ratings
(Observer notes interactions)
Collect Data

View Tape 6
Form Four-member Teams
Discussion and Team Ratings
(Observer notes interactions)
Collect Data

SESSION 3

View Tape 7
Rate Individually
Form Six-member Teams
Discussion and Team Ratings
(Observer notes interactions)
Collect Data

View Tape 8
Rate Individually
Collect Data

CLOSURE

233
INSTRUCTIONS TO OBSERVERS: (To be said by coordinator)

You are going to be observing and recording the interchanges between team members as they rate teachers on the North Carolina Performance Appraisal System. Teams consist of administrators and supervisors. The teams will vary in size from two to six members. Individual raters have observed a videotape segment of teaching, made notes using the FODI and the FODA, and have rated each of the five teaching functions on a scale of one to six. The teams, which have been assigned randomly, will then work toward a consensus score for each of the five functions.

As you observe the team activity, you will record the interchanges using an audio cassette recorder which may be placed on the table. In addition to the audio recording, you will make written notes about what you see. Note especially the facial expressions and body movements which may indicate emotions or feelings which are not expressed verbally. You will look for signs of dominance, agreement, disagreement, submission, confusion, and support. Who is the most powerful? Who is the most persuasive? As team members decide on the team rating, you will note and record their reasons for assigning a particular numerical score. When team members score the same, the reasons for the score should be discussed, just as when the scores differ.

During the team discussion, you may also become a facilitator. If team members fail to reveal reasons for a particular score, you are to ask some questions to probe for the information. Examples of questions you may ask are: Why did you choose a three? Does your three differ from her (other team member) three? What did you see that told you it should be scored a three? You should participate only as much as necessary to get the raters to reveal and discuss their perceptions of the score to the other team members. Please be careful not to inject your biases or opinions. As a facilitator, you are encouraging the raters to discuss their perceptions, not yours.

PRACTICE:

Observers will view two people role-playing the team rating conference, using data from a previous observation. Following the role-play, observers will discuss and share the notes they took. Discussion will include determination of when the observer would probe and question the raters, and the questions which might be asked.
OVERVIEW OF PROCEDURES:  (To be said by coordinator)

Let me review the procedures with you:

Raters will view and rate tapes individually, then they will work in teams to reach consensus ratings.

You, the observers, view the team rating process. You will work with the team with your assigned number. (Number the observers one through six.)

When consensus has been reached and rating completed, the observer will clarify notes while the raters are viewing the next lesson.

Observers will turn in the audio cassette and the written notes.

APPRECIATION:  (To be said by coordinator)

Your participation in this project is very important. Your willingness to help will contribute to this project, for without the observations you make, we cannot get accurate data on the interchanges of team members. Thank you.
INTRODUCTION:

Overview of the workshop - Facilitator will say the following to the participants:

As a way to maintain and improve skills in using the NCTPAS for evaluating teaching performance, we will view teaching tapes and rate the first five functions on the Teacher Performance Appraisal Instrument. Our goal as evaluators is to reduce the diversity in scores on the six-point rating scale. The North Carolina Department of Public Instruction has set +1 or -1 point as acceptable for consistent ratings. The tolerance of +1 or -1 point is not a problem as long as raters understand what the particular rating means. The tolerance becomes significant when the scores cross the line between acceptable (a score of 3 or higher) and unacceptable (a score of 2 or below). Because of the interpretations of the score differences, it is important to find out the reasons that an evaluator selects a particular score. As we view and rate various teaching segments, we will be discussing your perceptions of the different numerical scores and why you selected the score. Please be aware that in real situations, you would not determine a rating for the teaching functions based on only one observation. For our purposes, we will rate each lesson.

During the procedures, you are encouraged to discuss your findings in your groups, talking about the evidences you observed and on which you based your scores. When you and team members agree on a rating, please talk about why you agree.
The procedures which will be used are as follows:

1) You will all view a teaching videotape, using the Formative Observation Data Instrument (FODI) and the Formative Observation Data Analysis (FODA) to record and summarize your observations.

2) You will rate the first five teaching functions using the NCTPAS six-point rating scale. This will be done individually.

3) Teams will be formed using random selection with at least one building-level principal and one central office supervisor/administrator on each team.

4) As a team, discuss your individual ratings.

5) Reach a consensus rating on each function for your team. As you discuss your findings, please do not change your individual ratings. Your team rating will reflect the team consensus.

On your tables you will find the forms for FODI, FODA, the Teacher Performance Appraisal Instrument, rating sheets, and extra paper.

BEGIN VIEWING AND RATING OF THE TAPELED LESSONS.

TAPE 1: (Facilitator reads instructions aloud.)

We will view this 30-minute tape, taking notes using the FODI and the FODA as you have all been trained. You will then rate each of the first five functions using the six-point scale.

Let’s take a look at the definitions of the ratings to refresh your memory. (Read aloud the definitions of the six ratings on the TPAI.) Adjust your seats so you can see the monitor, make sure you have the forms you need. Does anyone have any questions?
VIEW TAPE 1 [Approximately 30 minutes]

(Facilitator says)

Take a few minutes to clean up your notes, write your FODA, and rate the lesson on the first five functions. When you finish, we will discuss your findings. [Allow 15 minutes.]

Now that everyone has finished, let's discuss what you saw as the strengths of this teacher's lesson. Also, the areas that need improvement. (Facilitator leads discussion, viewers comment.)

As we look at each function area, I will record on this chart how many of you selected each rating. (Facilitator marks scores on chart as hands are raised.)

Please do not change your scores as we discuss. Your score is not wrong because some other rater's score is different. We will use this first videotape as a warm-up activity to give you practice in observing and rating the lesson before we work in small groups. We will discuss this lesson as a total group rather than as teams.

(Facilitator leads discussion, asking for input from participants. Questions to obtain scores will be asked.)

Look at Function 1. How many of you rated this a 6 - Superior? (Record number on chart.) a 5? a 4? a 3? a 2? a 1?

(When the numbers are on the chart, ask for volunteers to give reasons for selecting a 6; then a 5; then a 4, etc. Observers will be making notes on reasons given. Encourage interaction by asking for someone to share a different reason for the same score. To get to all functions, discussion may be limited to keep the pace moving. For Functions 2-5, follow the same process as for Function 1. Total discussion should be limited to 30 minutes.)

Collect all written data and ratings from participants at the end of the discussion.

BREAK
TAPE 2: (Facilitator reads aloud the instructions.)

For the second tape, we will utilize the same process for viewing the tape, using the FODI and the FODA, and rating each of the five functions on the six-point scale. After you have completed your individual ratings, you will be assigned to a two-member team to reach a consensus rating on each function.

VIEW TAPE 2

(Participants rate individually. Facilitator then gives instructions.)

You will be assigned to two-member teams. After your teams are assigned, spread out in the room so that your discussions will not interfere with others. Remember, do not change the ratings on your individual score sheets. Use the team score sheet for the consensus team ratings.

(Two-member teams are assigned by facilitator. Teams move to areas for discussion.)

TEAM DISCUSSION AND RATING

(When completed, the data will be collected and turned in.)

BREAK

TAPE 3: (Same process as followed for TAPE 2, using three-member teams as assigned by the facilitator.)

END SESSION 1
SESSION 2

TAPE 4: (Same process as followed for Tapes 2 and 3, using four-member teams as assigned by the facilitator.)

BREAK

(For the next three tapes, observers will record the interactions of the team members during the consensus-rating activity.)

TAPE 5: (Same process as followed for Tapes 2, 3, and 4, using two-member teams as assigned by the facilitator.)

During steps 4 and 5, an observer will be recording your interchanges as your team works. This observer will be noting your perceptions of the various scores and the interactions of team members in reaching consensus. The observer will not be rating the teaching functions, nor making personal judgments on the ratings you make.

(When discussions and consensus ratings are completed by the teams, data is collected and turned in.)

BREAK

TAPE 6: (Same process as followed for Tape 5, using four-member teams as assigned by the facilitator. Observers record interactions of the team members.)

(Collect data and turn in.)

END OF SESSION 2
SESSION 3

TAPE 7: (Same process as followed for Tapes 5 and 6, using six-member teams as assigned by the facilitator. Observers record interactions of the team members.)

(Collect data and turn in.)

BREAK

TAPE 8: (This is the same taped lesson as Tape 1. Participants view and rate the tape individually. No team consensus ratings are conducted. Instructions are given to the participants.)

VIEW AND RATE THE LESSON

(Collect data and turn in.)

CLOSURE AND APPRECIATION

(Facilitator expresses appreciation for the cooperation of the participants. A summary of the workshop is given, encouraging the participants to make comments on the procedures.)

Facilitator says to participants:

Within the next three weeks, I will talk with each of you individually to share with you some of the overall rating results and to discuss your individual reactions to the rating process and the Teacher Performance Appraisal process. Your time and efforts in this project are appreciated.

END OF WORKSHOP
INTERVIEW GUIDE

1. What is the purpose of the North Carolina Teacher Performance Appraisal System or of any evaluation system?

2. Why did you rate the way you did?

3. What would a teacher do differently to receive a 5 instead of a 4?

4. How sure are you that the score should be a 4 for that function? Why?

5. How do you perceive the score of 6 versus a 3?

6. Which teaching function did you find the most difficult to rate? Easiest?

7. Which one or two function areas do you think you are more knowledgeable in for rating purposes?

8. Did you change your ratings when discussing the lesson? Why?

9. Did you notice a pattern in your ratings? Were they usually higher or lower than other team members?

10. In reaching a consensus with your team members, did you change your mind? Why?

11. With which team was it easier to reach a consensus?

12. Who would you chose to form a team?

13. If you were developing a rating scale, how would it look? How many points or categories?

14. Who are the two people out of this group that you think are the most knowledgeable in the teacher evaluation process?

15. Who would you like to have rate you?

16. Was this activity worthwhile? How would you change it if we did it again?
Appendix J

RATER INFORMATION

NAME: ___________________________ PRESENT POSITION: __________________

NUMBER OF YEARS IN THIS JOB: ________

TOTAL NUMBER OF YEARS IN ANY FULL-TIME ADMINISTRATIVE OR SUPERVISORY POSITION: ________

PRIOR CLASSROOM TEACHING EXPERIENCE:

<table>
<thead>
<tr>
<th>Grade Level</th>
<th>Number of Years</th>
<th>Subject Area(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>K-3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9-12</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CHECK ANY TRAINING THAT YOU HAVE HAD DURING THE LAST FIVE YEARS:

____ N. C. Effective Teaching
____ N. C. Teacher Performance Appraisal System (TPAS)
____ TPAS Trainer Training
____ N. C. Mentor Training
____ Clinical Supervision
____ TESA (Teacher Expectation and Student Achievement)
____ Assertive Discipline
____ COET (Classroom Organization for Effective Teaching)
____ Six-Step Lesson (Hunter's Instructional Presentation, SIP, STAR, etc.)
____ PDP (N.C. Professional Development Plan training)

UNDERGRADUATE MAJOR IN COLLEGE: ____________________________

AREAS OF N. C. CERTIFICATION: ________________________________

NUMBER OF FORMAL OBSERVATIONS DONE USING THE N.C. TPAS DURING THE 1986-87 SCHOOL YEAR: ________
Appendix K
Teacher Questionnaire and Responses

TEACHER QUESTIONNAIRE

During the 1986-87 school year, Hertford County Schools adopted and used the North Carolina Performance Appraisal System for the evaluation of certified teachers. Your assistance in sharing your perceptions of the system is needed. Please answer the following questions. In order to assure that responses cannot be identified with any person or school, do not put your name on the form. Your input is appreciated.

1. Do you feel that the observations that were conducted by your building principal or designee last year focused on the important functions of the N. C. Teacher Performance Appraisal Instrument during each observation?

2. Were the five observable functions (Management of Instructional Time, Management of Student Behavior, Instructional Presentation, Monitoring of Student Performance, Instructional Feedback) treated with equal importance? Should they be?

Comments:

3. Is it important for the same evaluator to conduct all observations during the year? Why or why not?

4. Do you think there is consistency in evaluations when they are done by different administrators?

5. How many different people using the NCTPAS observed you in an actual teaching activity?

6. How many formal observations were conducted on you during the previous school year by any observers? (Formal observations are those for which you received oral and written feedback using the NCTPAS.)

7. Were you in the Initially Certified Program during the previous school year?
TEACHER QUESTIONNAIRE

During the 1986-87 school year, Hertford County Schools adopted and used the North Carolina Performance Appraisal System for the evaluation of certified teachers. Your assistance in sharing your perceptions of the system is needed. Please answer the following questions. In order to assure that responses cannot be identified with any person or school, do not put your name on the form. Your input is appreciated.

1. Do you feel that the observations that were conducted by your building principal or designee last year focused on the important functions of the N. C. Teacher Performance Appraisal Instrument during each observation?
   YES - 155   NO - 3   NO RESPONSE - 3

2. Were the five observable functions (Management of Instructional Time, Management of Student Behavior, Instructional Presentation, Monitoring of Student Performance, Instructional Feedback) treated with equal importance?
   YES - 131   NO - 6   OTHER - 5   NO RESPONSE - 19
   Should they be?
   YES - 99    NO - 25   OTHER - 13   NO RESPONSE - 24
   Comments:

3. Is it important for the same evaluator to conduct all observations during the year?
   YES - 84    NO - 61   OTHER - 8    NO RESPONSE - 8
   Why or why not?

4. Do you think there is consistency in evaluations when they are done by different administrators?
   YES - 55    NO - 80   OTHER - 15   NO RESPONSE - 11

5. How many different people using the NCTPAS observed you in an actual teaching activity?
   Zero- 1   One- 71   Two- 64   Three- 18   Four- 2
   Five- 2    No Response- 3

6. How many formal observations were conducted on you during the previous school year by any observers? (Formal observations are those for which you received oral and written feedback using the NCTPAS.)
   Zero- 1   One- 24   Two- 38   Three- 82   Four- 3
   Five or more- 4    No Response- 9

7. Were you in the Initially Certified Program during the previous school year?
   YES - 60    NO - 94    NO RESPONSE - 7
Question 2. Were the five observable functions treated with equal importance?

RESPONSE — YES

Comments:

Each one is dependent on the other for the teacher to be effective.

All important in education.

One is not any more important than the others. They are an integrated whole and should not be considered individually.

All equally important and the success of one depends highly on the others.

One builds on the other.

Depends upon the teacher. Management of time and instructional presentation would be different for resource, ECIA teachers.

They help keep the classroom going smoothly and with this, learning is taking place.

Most career teachers know exactly how to carry out a planned lesson and know exactly what to expect out of their students.

If they are designated criteria.

I do not feel that the principal was in my classroom enough to judge these areas. I feel that he needs to see me teach several times in order to know my strengths and weaknesses.

Because it is an equal combination of all these aspects that make an effective teacher.

All areas are interdependent - if any one is missing to any degree, the class can't be successful.

I feel more strongly toward instructional presentation and instructional monitoring.

The nature of the lesson dictates which function should have the most importance.

Sometimes, I don't think all the functions are necessary.

Should be considered in relation to the kind of activities...over long period of time should use all 5 functions. However, on any 1 particular day, the teacher may use one...function more than another one.
Question 2. Were the five observable functions treated with equal importance?

RESPONSE - NO

Comments:

Management of student behavior should be first...If this does not occur, none of the others can take place.

In different lessons, different aspects need more emphasis.

Some things are far more important...than others.

Some items should be looked at with flexibility in mind.

Depends on what and how a subject is being taught.

No, it seems to me that more emphasis was placed on trying to find faults rather than submitting what really was being done.

Depends on plans for the day.

Not during a single class.

Some days one will be of more importance than maybe 2 others.

Kind of activity dictates the importance of each function.

Instructional presentation and instructional feedback are more important.

Monitoring and feedback should have the greatest importance, not as much...on focus and closure....

Instructional presentation should be treated as the number one focus.

Instructional presentation should be rated the highest followed by management of time...observer should choose times...with consideration.

What is really important is that a clear-cut lesson be presented and that students understand and learn from the lesson....

Especially student behavior. An observer...doesn't see the real student.

One student's behavior should not be treated with the same importance as the others.
TEACHER QUESTIONNAIRE, continued

Question 3. Is it important for the same evaluator to conduct all observations during the year?

RESPONSE - YES

Why or why not?

Comments:

The evaluator is more consistent in his observations.
For consistency.
Eliminates personality conflicts.
Because the same evaluator will know if the teacher has improved.
Because that evaluator will better be able to see the consistencies and discrepancies in the teacher's performance as well as the lessons.
Consistency.
He or she can see improvement or progress better.
The same person will know more how the teacher teaches.
If they are familiar with the subject being taught.
To make suggestions or comments to be improved.
It would seem difficult to observe the 5 functions in one class period.
Better able to see strengths as well as weaknesses.
One evaluation builds on another!
He is more acquainted and familiar with the pre-existing problems...where a different observer may see a problem not knowing if it shows growth from the last time and evaluate severely.
Nobody can evaluate a teacher on one or two formal visits. There must be informal and formal by the same person to get an understanding of that teacher’s approach....
A person cannot go into a classroom once a year and be able to give a person a fair evaluation.
Different evaluators perceive different things in different ways.
TEACHER QUESTIONNAIRE, continued

Question 3. Is it important for the same evaluator to conduct all observations during the year?

RESPONSE - NO

Why or why not?

Comments:

More people need to be aware of our teaching. The principal, assistant principal, and supervisor should observe. It's too much for one person.

If the same items are to be observed anyone could be the observer.

Different observers SHOULD reach similar conclusions.

Different points of view may be helpful or useful!

Each evaluator should be trained to conduct the observation.

What one sees, another may not.

No difference if evaluating on same points.

There is more input if there is an area of discrepancy.

The reliability of the instrument is better...if several evaluators.

Different ideas for improvement.

I feel that you need to be informed of what their findings are each visit before the end of the year so that you are aware of it.

It may be more fair to share the responsibility and to make it more a committee evaluation.

Better "averaging" of opinions if several observers are involved.

If the purpose is feedback, then feedback from different evaluators would be more varied and constructive.

Only if they meet to combine observations and discuss positives and negatives. Two opinions are often more fair than one.
The vita has been removed from the scanned document