

379
N81d
No. 3275

A STUDY OF EVALUATOR CONSISTENCY/STABILITY
IN THE APPRAISAL OF TEACHER
PERFORMANCE

DISSERTATION

Presented to the Graduate Council of the
University of North Texas in Partial
Fulfillment of the Requirements

For the Degree of

DOCTOR OF EDUCATION

By

C. Ellen Stricklin Fuller, B.A., M.A.

Denton, Texas

December, 1990

Fuller, C. Ellen Stricklin, A Study of Evaluator Consistency/Stability in the Appraisal of Teacher Performance. Doctor of Education (Administrative Leadership), December, 1990, 128 pp., 4 tables, 21 illustrations, reference list, 97 titles.

This study considered the appraisal patterns of 29 teacher appraisers from a single school district over a two year period. It discussed the district's efforts to assist its appraisers in evaluating teachers. The study compared the frequency and range of the exceptional quality scores and overall performance scores assigned to teachers evaluated using the Texas Teacher Appraisal System in 1987-88 and 1988-89. It also compared the most and the least positive appraisers each year by individual schools and grade levels. The study considered the effectiveness of additional local training on the appraisers' consistency in evaluating teacher performance.

Additional local training involved these areas: viewing videotaped teaching sessions, debriefing groups after scoring observed teaching lessons, observing actual classroom with field teams, staff development on exceptional quality scoring and meaning, and discussing the meaning of teaching behaviors of the TTAS instrument with other district appraisers. The findings included:

1. The total spread or range among the 29 appraisers' exceptional quality and performance scores assigned to teachers was less in 1988-89 than 1987-88.
2. The elementary appraisers were more positive than the secondary ones in awarding exceptional quality points and the overall performance scores.

Based on the findings of this study the following implications may be drawn:

1. Without training by school districts for appraisers, in addition to the training mandated by the state, it should be anticipated that wide variations in scores will occur and that consistency in scores will not occur.
2. Efforts by school districts to provide training to appraisers in order to improve the consistency of appraisals will reduce the range of appraisal scores given by appraisers and in general will lower the scores given by the most "positive" appraisers and increase the scores given by the least "positive" appraisers.
3. The commitment of the district to improving consistency among its appraisers should enhance the perception of fairness of the district's evaluation system.
4. Local districts should develop ongoing appraiser training to instill the goals of consistency and fairness.

TABLE OF CONTENTS

	Page
LIST OF TABLES	v
LIST OF ILLUSTRATIONS	vi
 Chapter	
I. INTRODUCTION	1
Statement of Purpose	3
Purposes of Study	3
Research Questions	4
Background	5
Significance of Study	8
Definitions of Terms	9
Limitations	12
Summary	12
 II. RELATED RESEARCH	 14
The Texas Teacher Appraisal Instrument	14
Raters/Appraisers	16
Perceptions of Fairness and Purpose of Evaluation	20
The Training of Appraisers	23
Rater Consistency	28
Summary	38
 III. METHODS AND PROCEDURES	 39
Population	39
Permission for Conducting The Study	41
Procedures for Implementation of The Study	41
Description of District Training of	44
Procedures for Local Appraiser Training	46
Procedures for the Analysis of Data	52
Other Data Collected for Study	53
 IV. PRESENTATION OF FINDINGS	 55
Research Question One	55
Research Question Two	68
Research Question Three	72
Research Question Four	76
Research Question Five	77

Research Question Six	78
Other Data Findings	79
V. SUMMARY OF FINDINGS, IMPLICATIONS AND	
RECOMMENDATIONS	84
Summary of Findings	84
Implications and Suggestions for Further . .	
Research	89
Recommendations	90
APPENDICES	94
REFERENCES	120

LIST OF TABLES

1.	Appraiser scoring patterns	56
2.	Comparison of appraiser ranks in awarding E Q points from most to least positive	77
3.	Appraisers with extreme scoring patterns	79
4.	Education work experience of appraisers in 1988-89	82

LIST OF ILLUSTRATIONS

Illustration	Page
1. Texas teacher appraisal system	6
2. School A: Comparison of % of possible E Q points granted	58
3. School B: Comparison of % of possible E Q points granted	59
4. School C: Comparison of % of possible E Q points granted	60
5. School D: Comparison of % of possible E Q points granted	61
6. School E: Comparison of % of possible E Q points granted	62
7. School F: Comparison of % of possible E Q points granted	63
8. School G: Comparison of % of possible E Q points granted	64
9. School H: Comparison of % of possible E Q points granted	65
10. School I: Comparison of % of possible E Q points granted	66
11. School J: Comparison of % of possible E Q points granted	67
12. School K: Comparison of % of possible E Q points granted	68
13. School L: Comparison of % of possible E Q points granted	69
14. Percent of possible E Q points granted by school	70
15. Percent of possible E Q points granted	71
16. Comparison of mean performance scores by school	73

17.	Comparison of mean performance scores by grade level	74
18.	Percent of possible E Q points by appraiser	75
19.	Percent of teachers with masters degree	81
20.	E Q points granted compared to masters degrees held	81
21.	Field team practice schedule 1988-89	115

CHAPTER 1

INTRODUCTION

As the cry for educational reform in the 1980s spreads across our nation, the expectations for implementation naturally follow. As reform must find its way to the student, no area of education has been impacted more than building level personnel. Administrators perform as the instructional leaders in their buildings, and their leadership is crucial to any reform. The success of any educational process, however, relies upon instructional personnel. Teachers instruct by using teaching methodologies that their states or districts determine will effectively promote student achievement and student learning. Administrators evaluate teachers with appraisal instruments containing teaching behaviors which can enhance student learning.

As these reforms impact every school district within the state of Texas, reactions are naturally mixed. Teachers perceive this reform with varied levels of enthusiasm and acceptance. Building administrators also adapt to the imperatives with both mixed acceptance and understanding. The actual evaluation processes associated with school reform impacts instruction in the classrooms, administrators' roles, and often the teachers' pay checks. The effects of implementing reform send shock waves to some, wails

to others, and silence to still others. The implementation of reform comes in the classrooms of Texas schools as the teacher evaluation instrument and process known as the Texas Teacher Appraisal System (TTAS).

This study focuses on one of the main role players in this evaluation system, the building level administrator as evaluator or appraiser. These administrators perform the evaluations mandated by the Texas legislature.

The evaluation system resembles a velcro strip attached to classrooms, administrative offices, and pay checks of teachers. To this velcro strip of teacher evaluation clings many policies and practices. Some of these attachments connect through legal mandates while others attach themselves out of custom and personality. However, the fundamental variable in the teacher evaluation process, other than teacher, is the teacher's evaluator or appraiser. Greenfield (1987) notes that the largest single investment of the reform movement is in the procedures and finances used to evaluate teachers and the training of the evaluators. These trained appraisers use their experiences, personality, shared beliefs with the teachers, training to use the evaluation instrument, and their recognition of identified, effective teaching behaviors in functioning within this evaluation system. These appraisers must also, of course, utilize their professional judgements when implementing any evaluation system.

Because each administrator/evaluator brings a unique set of characteristics and professional judgements to the evaluation process, two questions result. Does each appraiser's uniqueness make it possible to achieve reasonable consistency when an appraiser's ratings are compared to the ratings of other appraisers within the same local district? Can a school district identify, and once having done so, modify an individual appraiser's consistency with fellow appraisers in rating teachers?

Statement of The Problem

The problem of this study is to describe and analyze the differences, if any, in the inter-rater or inter-observer reliability among 29 appraisers of a single school district based upon comparisons of the teacher appraisal records for 1987-88 and 1988-89.

Purposes of This Study

The purposes of this study include the following:

1. To provide information concerning the degree of inter-observer reliability in the evaluation of teacher performance that exists within the district itself, within its 12 schools, and within its elementary and secondary settings.
2. To compare the overall range and percentage of possible exceptional quality points assigned to teachers evaluated by this school district's appraisers in 1987-88 and 1988-89.

3. To compare the appraisers' rankings to determine the more and the least positive appraisers during this two-year period.

4. To determine the effectiveness of additional local training for increasing inter-observer reliability and consistency.

Research Questions

The following research questions are posed in order to achieve the purposes of this study.

1. Do appraisers within the same building who observe the same set of teachers evaluate these teachers' performances similarly?

2. Do elementary appraisers have similar or different patterns of evaluation from secondary appraisers?

3. Considering the exceptional quality scores between and within the two-year period, what is the range and percentage of scores among the appraisers?

4. Did the range and percentage of exceptional quality scores change from 1987-88 to 1988-89? Are the changes reflected in some, most, or all of the appraisers in the school district?

5. Does additional local appraiser training lead appraisers to evaluate more consistently in the area of exceptional quality points?

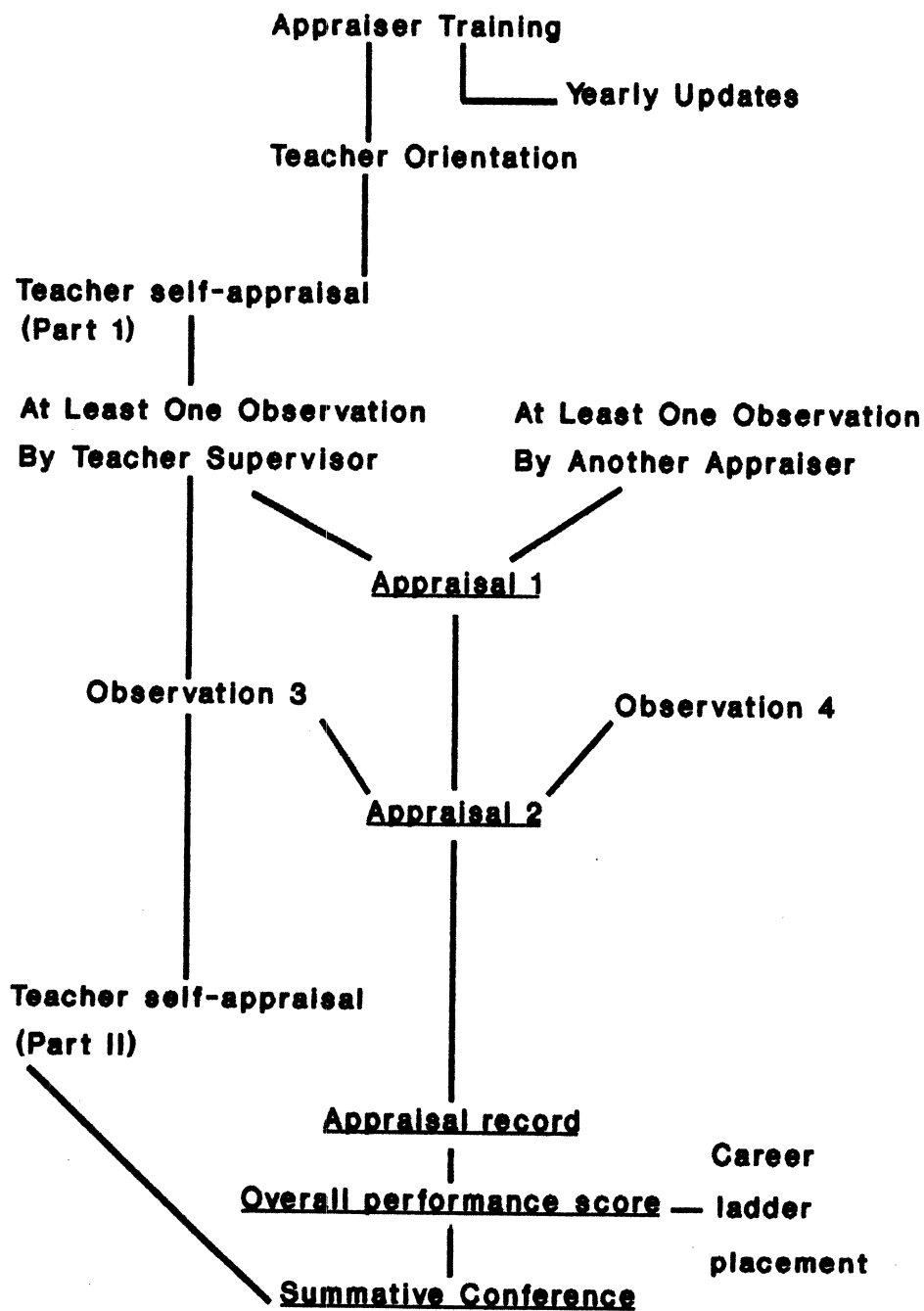
6. Do schools with a high percentage of teachers with master's degrees earn the greater number of exceptional quality scores?

Background

The Texas Legislature mandated the development and the implementation of an appraisal system for teachers in January 1985 as part of the omnibus educational reform package, House Bill 72. As shown in Figure 1, the state legislature intended for the TTAS to serve as a tool that would ultimately enhance the instruction within Texas classrooms and spur greater professional growth among teachers and administrators.

The state requires certified state trainers who must use an official state training manual developed by the Texas Education Agency (TEA) to instruct all persons who become TTAS appraisers in Texas. Requirements for appraisal certification include a 70% mastery of tests over the TTAS rules and procedures and a 70% scoring of videotaped teaching units when compared to scores of a state panel of experts. Mandated yearly updates for these appraisers to maintain their certification do not require further mastery testing or greater mastery of the scoring of other videotaped teaching units. Only one section within the training manual concerning exceptional quality mentions a possible appraiser problem of consistency.

Figure 1. Texas Teacher Appraisal System



Judging the exceptional quality dimension is necessarily subjective. One pitfall is that the appraiser's standard for exceptional quality may be too low. Historically, evaluators are generous when assigning scores. As a consequence, the appraiser must be convinced, based upon the available evidence, that the teaching behavior goes beyond the expectation and reflects unique qualities. (Texas Education Agency, 1986, p. 33)

The foundations or assumptions of the Texas appraisal system include a description of the view of teaching, the philosophy of generic teaching behaviors, the subjectiveness of quality performance, the expectations for professional growth, a commitment to educational goals, and the maintenance of accuracy of the appraisal process (Texas Education Agency, 1986b). None of the TTAS elements, however, establishes a methodology or suggests an infrastructure to monitor the inter-observer reliability or consistency of the appraisers. The apparent lack of concern for consistency and uniformity among the appraisers seems to be a deficiency in the TTAS. Over this two-year study, the Texas Education Agency has required several times that copies of local teachers' appraisal records be sent to the state. However, the results of analysis of these data have not been returned to the district. One assistant superintendent in the district in which this study was conducted communicated to the

district appraisers that the state expected each district to monitor its own appraiser consistency, "TEA is now requesting that school districts develop and implement plans to increase inter-rater reliability at the local level, and to focus on discrepancies among our own appraisal staff" (District memo, June 15, 1988). School districts who have utilized the TTAS instrument have data within their evaluation records to analyze the scoring of administrators who implement the career ladder system.

Significance of Study

This study presents an analysis of a single school district's efforts to implement the state mandated TTAS procedure. This analysis is significant for those involved: the teachers who are appraised, the administrators who act as the appraisers, and the central district administrators who implemented a career ladder system based on the TTAS instrument.

For the purposes of improvement and fundamental fairness to the teachers appraised, appraisers within a building and a district should evaluate consistently. The data from appraisals of all appraisers within a building and within a district should approximate a normal curve when graphed. Districts should chart the range of scoring among district appraisers to determine appraisers' inter-observer reliability. Districts should try to increase the inter-observer reliability through additional local training. This study

deals with some of these issues and the procedures of a district attempting to identify and to improve the inter-rater reliability of its appraisers.

Information obtained from this study will be used to offer direction and suggestions for districts in achieving consistency among district appraisers. The results will provide directions for future training and updating of appraisers. These deliberate local attempts to attain higher inter-observer consistency and reliability in scoring can show teacher groups their district's concern about fairness in the appraisal process. A district with a fair and consistent process of evaluation in place would do much to defuse future arguments concerning inconsistent ratings or evaluations.

Definition of Terms

For the purposes of this study, the following definitions are provided.

The appraisal is the evaluation process of TTAS based on state trained and certified appraisers classroom observations of teachers. The TTAS specifically includes two appraisals during each school year. These times are usually divided by the two semesters designated by the school calendar adopted by the local school board. Both the teacher supervisor and the "other" appraiser observe once each during a single appraisal. Teachers on higher levels of the

career-ladder system receive only one appraisal by each appraiser each year.

An appraisal record is the final written record of the TTAS. It combines the scores of the teacher supervisor and the "other" appraiser for the two appraisal periods and contains the overall performance score.

An appraiser is one of two individuals certified by the state in appraisal training assigned to evaluate the performance of a teacher. One appraiser is designated as the teacher supervisor, and the second is designated as the teacher's "other" appraiser. The teacher supervisor must hold administrator or supervisor certification. The "other" appraiser must have a valid teaching certificate, have taught at least two years in a pre-kindergarten, kindergarten, elementary, or secondary classroom, and be approved by the local board of trustees. Both appraisers must have received uniform training and be certified by the TEA as appraisers. Other terms used in the research of this paper to identify an appraiser are rater and evaluator.

Career ladder is the merit pay system for Texas teachers determined in part by the scoring of the TTAS. Teachers who receive high appraisal scores and professional growth through college courses or workshops move up the four-tiered career ladder. This ascent results in additional dollars to these professional teachers.

Consistency for the purposes of this study means the scoring of the appraisal system as fellow appraisers within a definite range determined by the local district. The range of 40 to 60 percentage of possible exceptional quality points is the consistent range in this study.

Exceptional quality is a measure of performance which indicates that a teacher has demonstrated a teaching behavior in a superior way and that the performance has contributed to a high level of success (student achievement). The appraiser determines this score by observing the extent of active student involvement in the lesson.

An evaluation record is a written record of TTAS completed by the "other" appraiser after a formal observation of a teacher and by the teacher supervisor at the end of each appraisal period.

An indicator is an individual item on the TTAS instrument which represents a teaching behavior.

Inter-rater/observer reliability is the level of consistency between appraisers, raters, and the scoring of an evaluation instrument usually at the same time for the same performance. This study, however, focuses on the consistency among appraisers from year to year, for different teaching performances, and at different appraisal times.

An observation record is the written record of TTAS completed by the teacher supervisor and the "other" appraiser after a formal observation of a teacher.

The overall performance score is the combination of the teacher supervisor's and the "other" appraiser's rating of a teacher using the TTAS instrument. This score determines a teacher's placement or maintenance on the career ladder each year.

Professional judgement is the subjective interpretation of classroom observation information by a person who has been trained or has developed expert knowledge in the analysis and recognition of effective teaching behaviors. Administrator's mandated uniform instructional leadership and TTAS training justify their use of professional judgements in assigning TTAS scores for individual teachers.

The TTAS, Texas Teacher Appraisal System, is the teacher evaluation system mandated by the Texas legislature and developed by the Texas Education Agency and the State Board of Education in Texas.

Limitations

Five limitations impact upon this study.

1. Information and experiences acquired apart from the district training could affect the appraiser's scoring.
2. One can not precisely determine the specific impact of the local training.
3. Two different appraisers observing the same designated teachers on different days may account for some of the difference in consistency.

4. Other districts can not fully generalize from this study to their particular locales.

5. The definition of consistency is inherently an arbitrary decision determined by the local school district.

Summary

The study of a single school district's inter-rater reliability includes four purposes and six research questions. Chapter 2 will provide a review of the related literature. Chapter 3 will present the methods and procedures of the study. Chapter 4 will include the findings of the study. Chapter 5 will provide implications and recommendations derived from the study.

CHAPTER 2

RELATED RESEARCH

This study focused on appraisers within the context of the teacher evaluation process. Five major areas of research are pertinent to the purposes of this study: the Texas Teacher Appraisal System instrument, the raters, the training of appraisers, the perceptions concerning fairness and purposes of raters, and rater reliability or consistency.

The Texas Teacher Appraisal System Instrument

House Bill 72 established the imperative for educational reform in Texas. One of the reform imperatives mandated the drafting of an appraisal system for teachers. The State Board of Education of Texas adopted the Texas Teacher Appraisal System and implemented it during the 1986-87 school year (Texas Education Agency, 1986). Personnel of the Texas Education Agency developed training procedures and manuals for the training of a statewide base of trainers. These trainers instructed every appraiser within the state for 40 hours. The trainers used the same procedures, timeline, and training manual. The Texas Education Agency organized a Teacher Orientation Manual to provide a resource for teachers for their roles within the evaluation system and to

inform them of the background, assumptions, overview, research, and elements of the Texas Teacher Appraisal System.

The timeline for the development of a Texas appraisal system for teachers began in January 1985. The staff of the Texas Education Agency (1986) managed the following steps in the development of the evaluation system:

- . review of literature on teaching.
- . survey of other statewide appraisal systems.
- . random survey of 156 school districts in Texas to analyze their teacher evaluation systems.
- . job-relatedness survey given to a sample of 30,000 Texas teachers
- . authorization of expert review of draft materials.
- . pilot study in October 1986, of six school districts: Slaton, Santa Rosa, Seguin, Port Arthur, Grandfalls-Royalty, and New Boston.
- . public hearing by State Board of Education on February 6 and 8, 1986.
- . revisions of the appraisal process.

The purposes and assumptions of this Texas evaluation system testify to the state's commitment to educational reform. The purposes express the need to improve instruction and to implement a merit pay system for teachers. The six assumptions listed in Chapter 1 form the basis of the evaluation system. They include generic teaching behaviors, the subjectiveness of quality performance, and the mainte-

nance of accuracy of the appraisal process (Texas Education Agency, 1986b). Figure 1 illustrates the elements and sequence of the TTAS instrument. The State Board of Education makes revisions to the TTAS in the scoring chart, in encouraging a pre and post conference, and in requiring the scheduling of observations with a three day notice to the teacher. The Texas Education Agency communicates these changes to the local districts through communication with the superintendent or through yearly training updates by the regional service center for all certified appraisers. These changes reflect the cyclical process of evaluation in which the inputs of teachers and administrators are considered in on-going implementation of the appraisal procedures.

Raters/Appraisers

Appraisal systems require the use of human raters for the scoring of the appraisal instrument. Frick and Semmel (1978) analyze the uniqueness of human raters. They argue that rarely would the observational skills of raters be identical or equivalent. Therefore, consistency for a district will not mean that all its appraisers will evaluate in the same way.

Popham (1985) notes that appraisers tend to compare a ratee's performance with their own in a classroom. Appraisers expect behaviors in the teachers they rate to be similar to the behaviors they had as teachers. Therefore, better teachers will expect better teaching behaviors while average

or less effective teachers will expect less of the teachers they evaluate. As these teachers become administrators in the future, their expectations of teaching behaviors will affect their appraisals of teachers.

The use of raters with different experiences and levels of expectations for teachers results in varied and inconsistent implementation and application of an appraisal system. Authors (Bridges, 1974; Brody, 1977) of studies describing the effects of human raters on appraisal systems advise those in charge of the management of such systems to be cautious of lack of consistency and fairness of evaluation results or scoring.

Evaluation involves not only the processing of varied personal dispositions unique to each appraiser but also observational techniques and information as well. Kaplan (1973) found that evaluation judgements are inconsistent because of the incorporation of the appraiser's personal dispositions and observational techniques. He further describes these evaluation judgements as "dispositions averaged with information" (Kaplan, 1973, p. 60). This means that an appraisal consists of the appraiser's interpretation of observational information colored by the appraiser's personal views and opinions. Braskamp, Brandenburg, and Ory (1984) add the element of interpretations as well when they argue that "evaluation is more than description" (p. 29). The raters interpret, perceive, and

determine the priority these descriptions. The teacher's efforts or teaching behaviors are evaluated by more personal criteria of the appraiser. Improving an appraiser's informational gathering may improve observational data, but can not remove their human traits of perception and interpretation.

The research of Murphy, Gannett, Herr, and Chen (1986) and Murphy and Balzer (1989) indicates that the rater's later memory for specific behavior is systematically affected by a general impression or personal schema of the ratee. Raters tend to remember what is consistent with their general impressions. They also organize data to be remembered by generalizing it into patterns. This generalization assists raters in remembering observational data. These patterns reflect the ones known to the raters from experience. Smither, Barry, and Reilly (1989) also suggest that raters use a strategy to focus their data which involves forming a comparative impression of ratees. Cardy and Kehoe (1984) call these global categories.

The conclusion to another study (Murphy et al., 1986) suggests that raters often show biases for ratees whom they have evaluated previously. Raters tend to remember those previous behaviors that are consistent with present performances. If the same raters are used during past and present appraisals, the scoring results are similar.

Bridges (1974) offers three reasons why teacher evaluations are too positive. These include appraisers lacking academic preparation to handle face-to-face negative information, difficulty of evaluating professionals who possess competencies markedly different from their own, and using leniency in evaluating as a means of establishing authority over subordinates.

Cardy and Dobbins (1986) studied the liking dimension or affective area of an appraiser for the ratee. These data reveal that the liking dimensions for a ratee can not be separated from the performance dimensions of the ratee. The rater's liking or disliking for a ratee directly influences the appraisal of that ratee.

Additional research by Cardy and Kehoe (1984) concerns raters with high levels of selective attention ability being able to differentiate more accurately among various dimensions of ratees' behaviors and performances. Appraisers can recognize detail in observation if they have a higher selection ability.

Researchers have examined the illusory correlation between job satisfaction and job performance (Smither, Collins, & Buda, 1989). The illusory correlation included a general impression of a ratee developed or based on another impression or event. For example, a rater may generalize a positive view of a ratee who is viewed as loyal and clean. One of their conclusions is that when ratees present the

impression of job satisfaction, their chances of a positive appraisal are more likely than either a negative or a neutral appraisal by the rater. Therefore, communication concerning job satisfaction during a pre-observation conference could affect an appraisal or evaluation.

Littlefield (1985) discusses the elements of the level of rater and observee contact. High levels of association between rater and ratee result in more positive ratings while low levels of contact associate with less positive ratings.

Pritchett (1986) studied the effects of close personal contact of administrators who were also appraisers with other practicing administrators. Such personal contact between and among appraisers results in more consistent scoring by the appraisers.

Perceptions of Fairness and Purpose of Evaluation

The ratees expect fairness in and a mission or purpose for their evaluations. When an appraisal system typifies fairness and clear purpose, the ratees will exhibit more acceptance of the system. The ratees' demand for fairness with their appraisal procedures extend to the fair implementation of the evaluation system by the school districts and their individual raters.

Ratees need to view their evaluations as part of a fair process. Glassman and Paulin (1982) argue that positive receptivity to evaluations is related to "the trust and

confidence which teachers have in the expertise of their evaluators" (p. 159). The training to assess received by appraisers needs to be communicated to the teachers initially in the evaluation process. Bradley (1983) indicates that the first step in establishing a sense of fairness in teacher appraisal requires principals to inform their staff of their training for use of evaluation instruments and explaining the procedures to follow. Staff development on the appraisal system enhances this fairness.

Edwards (1983) presents the idea that a careful system of analyzing the rater's role and the method for the use of the appraisal system augments the rater's perception of fairness. Greenberg (1986) describes seven components which may enhance a ratee's perception of the appraisal system's fairness. Four of these include the use of diaries, rater familiarity with ratees, consistent application of standards, and communication between rater and ratee before and after performances by ratees. Splitt (1985) agrees with Greenberg that consistency of raters increases the overall perception of fairness on the part of ratees. Consistency of appraisers needs to be from one observation to another and among the different appraisers.

Even at the university level the professors who are appraised state that one of the reasons of faculty dissatisfaction with their annual performance evaluations is incon-

sistency among the raters (Ormrod, 1986). This inconsistency held true for both student and colleague appraisals.

Another perception of the raters which affects their views of appraisals is the stated purpose or purposes of the evaluation system. Murphy, Balzer, Kellam, and Armstrong (1984) suggest that the purpose of evaluation does affect the manner in which the raters process behavioral information. The purposes of the overall evaluation process play important roles in the variability of ratings. Studies indicate that evaluation of teachers for purposes of instructional improvement and job retention are more consistent than ratings with the additional purpose of merit pay (Zedeck & Cascio, 1982). These combinations of purposes pose conflicts or incompatible goals for the raters (Banks & Murphy, 1985; Williams, DeNisi, Blencoe, & Cafferty, 1981). For example appraisal information for the purpose of job retention focuses comparisons between individuals and more global features while appraisals for instructional improvement target within-individual comparisons and particular teacher behaviors. One study (Cleveland, Murphy, & Williams, 1989) concludes that the implementation of performance appraisal systems has often presented the raters with conflicting demands and purposes. One purpose suggested for the improvement of instruction expects appraiser observation and supervision of teachers. Another purpose for teacher appraisals is merit pay.

The rater's concentration on things the ratee does wrong do not enhance positive purposes of evaluation to improve instruction (LeBrun, 1986). The rater becomes the expert rather than the teacher in the classroom or even a collaborative team of experts. The teacher's perception of the evaluation system becomes less positive as the teacher's role becomes less involved in the system.

The Training of Appraisers

Raters or appraisers obviously need training in understanding the implementation and the scoring of the evaluation instrument. Greenfield (1987) notes that the largest investment of educational reform lies in the evaluation of teachers and the training of evaluators. He suggests that districts commit to updating the skills of their administrative raters. The trainers of these raters usually determine the amount of time devoted to this training as well as the quality of information presented to the raters.

Several researchers (Brown, 1968; Harris, 1986; Ivancevich, 1979) have studied the effects of training on raters' consistency and accuracy. Studies (Brown, 1968; Ivancevich, 1979) indicate that formal rater training does have some impact on the reduction of halo error, a general impression from a particular event or impression. Harris (1986) provides a sequential training plan for raters which involves on-site classroom observations. He requires practical training of appraisers within actual classrooms. Trainers

can monitor the appraisers consistency in implementing the evaluation instrument and provide feedback to the appraisers concerning their individual observation techniques.

Other researchers (Wise, Darling-Hammond, McLaughlin, & Bernstein, 1984) suggest that instruction and practice in giving negative feedback be given to the ratee. They noticed that raters were reluctant to discuss negative evaluation results with ratees. The rater's communication system which involves listening and relating serious deficiencies of the rater needs continual review, assessment, and practice (Sullivan & Walker, 1981). This focus on communication skills was added to training of appraisers to improve their conferences with ratees on these issues.

Andrews (1985) states other elements of need for remediation and practice for raters. These elements include training in educational psychology to assess classroom climate better, evaluation methodology for the particular rating instrument, recognition of quality performance and varied deficiencies of performances, and documentation of observation data.

Other researchers (Alfonso, Firth, & Neville, 1984; Stow & Frudden, 1985) suggest training in the area of analyzing teacher behaviors and technical training in using an observation system. Since the TTAS is based on a set of established teacher behaviors, the focus of this type of training directly applies to the raters.

Training goals need to include increasing the awareness of the sources of the individual rater's bias and to immerse the rater in the evaluation process totally (Clive, 1984). Training needs to have individual components for gauging individual rater's problems and future development (McConnell, 1971).

Effective assessment programs used a variety of media to keep interest and to facilitate training (McConnell, 1971). He, also, utilized assessment centers for individual appraisers to be free from distractions and to become totally immersed in the understanding and implementation of the evaluation instrument.

Several projects involving performance evaluation encourage training in panels or groups. Panel judgements result in consistent, reliable ratings even when the panel members receive little training and do not have clear standards for judgements (Peterson, 1988). The verbal interaction and compromise among raters seems to provide a basis for more consistency. Friedman (1986) designed steps to develop objective appraisals, and he offered the provision that raters need an on-going training in finalizing their ratings. This on-going training provided periodic refresher training and time to clarify any appraiser questions.

Strategies of training that attempt to focus the raters to consider the same valuative standards for use as references for judgement of ratee performance. This adherence to

a standard results in more accurate and consistent ratings. Several authors refer to the frame-of-reference training as FOR training (Hauenstein & Foti, 1989; McIntyre, Smith, & Hassett, 1984; Pulakos, 1984, 1986). These studies assume that raters have different frames of references from one another, from the instrument, and from the training viewpoint. Through training in the valuative standards, these same raters become more consistent with each other in the use of an instrument for appraisal.

Another vital component of training should develop observational skills which focus on the descriptive rather than the valuative data (Borman, 1979; McGreal, 1983; Pulakos, 1984). Training needs to facilitate the use of the rating instrument to include laboratory sessions with films, videotapes, and written transcripts of simulated or real classroom settings (Harris, 1986). Also, on-site classroom observations need to be done by the training groups at least four or five times to develop minimal rating skill (Harris, 1986). Training should allow discussion of multidimensional performances, value of fair evaluations or ratings, and practices in categorizing effective and ineffective ratee behaviors (Hauenstein & Foti, 1989).

Other training goals include not only methods to increase raters' technical knowledge and skills but also methods to motivate raters to use skills (Bernadin & Pence, 1980). When raters begin to exhibit inconsistency with

other raters, they should attend retraining programs and participate in practice exercises (McIntyre et al., 1984). Such retraining should focus on discussion of rating areas that demonstrate the most inconsistent scoring (Hauenstein & Foti, 1989). Other authors encourage refresher training for raters after any idle periods or before any new period of observations (Atkinson & Murphy, 1987; Ferguson & Enger, 1985; Medley, Coker, & Soar, 1984). Zuick (1986) recommends retraining of raters whose ratings are not consistent with fellow raters. Training provides a similar orientation to the rating process and should incorporate real situations. Training should provide practice sessions for the rater to observe, record, and evaluate with other raters.

Research concerning the extent and the continuity of training of the appraisers has provided varied results. Gray (1982) discusses the purposes of on-going training. He suggests such continual training develops evaluators' skills and improves written documentation. Atkinson and Murray (1987) suggest that systematic practice sessions for all raters are helpful in improving inter-rater reliability.

Research organized to study the amount of time needed for training had varied results (Gaugler & Thornton, 1989; Lemley, 1986). Gaugler and Thornton proposed times ranging from .5 to 15 days to provide on-going training. Training needs to focus on methods to limit data processing demand on the raters. A longitudinal study of the effects of rater

training by Ivancevich (1979) concludes that training effects are reduced over time and that there is a carryover from rater training to decreasing halo and leniency error. He suggests that refresher training sessions need to be planned periodically to maintain rater training effects. Bernadin's (1978) research indicates that such training effects are dissipated after 13 weeks.

In conclusion, training of appraisers does offer chances for a school district to develop better observation skills locally. The organization of such a commitment implies much time and designation of expected outcomes and standards.

Rater Consistency

The role of the administrator not only as an appraiser but as a reliable inter-rater of teaching performance challenges researchers. Weare and Woodall (1985) developed an annotated bibliography of 32 articles concerning inter-rater reliability. Other areas of performance from theater to dentistry have recognized the role of the observer as an expert critic of performance.

Murphy and Balzer (1989) suggest that studies to determine rating accuracy using rater error may not be valid to determine rater consistency. Raters can error in different parts of an evaluation instrument and still assign the same score to the ratee. More current research by Bernadin and Pence (1980) hypothesizes that training designed to

reduce rater error may only substitute a more positive response with a less positive or central tendency response. They explain that training may increase consistency among raters; however, their accuracy may not be improved. Raters may shift their scoring patterns to reduce the stress from other peer appraisers. Raters may shift scoring more or less positively to appear more consistent with other appraisers for training purposes. Smither et al. (1989) support this conclusion. Although appraisers and ratees share biases, common ideas and goals, the resulting appraisal may be inaccurate; yet, a group of raters who consistently rate with error would have a high inter-rater reliability.

Several studies (Bernadin & Pence, 1980; Borman, 1975; Constable & Andrich, 1984; Doyle, 1983; Graves, 1982; Wells, 1982) of appraisers have implied that improvement in inter-rater reliability and consistency can occur. However, no method promises total consistency or continual improvement. For an item or score to have inter-rater reliability, several appraisers may score consistently or in the same way after considering one observation or one item on an instrument. Inter-rater consistency, as used in this paper, represents a broader dimension. This term refers to the close or narrow range of scores by all appraisers as compared to their fellow appraisers when using the same instrument. For example, an instrument may include 50 items for rating teachers. Data determine that raters score 40 items

the same or closely while the other 10 items have a greater range of scoring. The 40 items scored closely represent inter-rater reliability; whereas, the 10 items with greater range do not. An example of inter-rater consistency is when appraiser A scores the total instrument similarly to appraisers B-Z. If appraiser A's scores reveal a dissimilar score from the other appraisers, his score indicates inconsistency and a deviate score from the majority of the other appraisers (Vance, Winne, & Wright, 1983).

How, if ever, can this inconsistency change and even improve? The use of a structured evaluation instrument during the training of raters can improve the inter-rater consistency and stability (Doyle, 1983; Graves, 1982; Wells, 1982). Raters tend to give higher ratings to ratees with similar life styles and interpersonal needs (Bullock & David, 1985). Another study (Lewis, 1973) concerned the factors of the compatible views on rules, individual freedom, and policies during the principal-teacher ratings. Lewis further argues that rater and ratee should be matched by their views of teaching and methodologies to improve inter-rater reliability.

In contrast, the Association of Teacher Educators (1988) offers that raters need to come from outside the teacher's building or outside the district to obtain more objective and less biased evaluation. The authors argue

that this personal distance enhances rater honesty and accuracy.

Another idea presented is the necessity for two or three classroom visits (Association of Teacher Educators, 1988; Yap & Capie, 1985). One observation time can not provide the opportunities to determine the complexity and full view of a teacher's skill. The authors in a study of 25 teachers found that a single day observation did not provide sufficiently reliable data to differentiate teaching patterns or behaviors among these teachers.

Another study of raters (Littlefield & Troendly, 1987) suggests that using rating instruments that follow the raters' cognitive processes may result in more reproducible scoring. Otherwise, some, if not all, of the raters' thinking styles could conflict with the evaluation procedures and philosophy. The rater's data collection patterns and recall of observational items would be in conflict with the instrument and its implementation.

Darling-Hammond, Wise, and Pease (1983) note that when raters are asked to retrain or to be more consistent with other raters, trainers should immerse these raters in methods of adapting to change and implementing change. The raters can develop better rating skills as they learn how to adapt or alter their methods of rating.

The use of diaries or written records by raters assists in recalling distinctions in ratee performances (Bernadin & Walter, 1977; DeNisi, Robbins, & Cafferty, 1989). Borman

(1975) has indicated, that after training, raters can identify ratee's individual strengths and weaknesses better than before such training. Hall, Posner, and Harder (1989) compared what raters actually do when rating with what scholars say is best for performance appraisal. They identified this comparison as the practice-theory gap. Their study analyzes representatives of the systems of 36 hightechnology firms. The authors found seven key actions for improving performance appraisal:

1. Encourage managers to conduct frequent work counseling sessions.
2. Explicitly reward managers for developing their subordinates.
3. Provide process skills training that is directly related to performance appraisal.
4. Link the performance appraisal process to the company's human resource planning.
5. Focus appraisals more on future behavior than on past performance.
6. Establish explicit performance goals in the appraisal process.
7. Build an evaluative feedback loop into the performance appraisal system.

Attempting to define characteristics of an effective employee or ratee may improve ratings consistency (Feldman, 1981, p. 144). Feldman labels these effective employees as

"valued employee prototypes" who possess behaviors described within the evaluation instrument. He describes the human behavior of the rater as subjected to moods and varied situations and limited in the amount of time devoted to evaluations in real school settings. Furthermore, he introduces the idea that the organization or district has the responsibility of providing a positive atmosphere for both the rater and the ratee to have positive results.

McGreal (1983) suggests that the goal for effective training programs should be "the notion of developing in supervisors the ability to narrow their focus during observation and skill in collecting descriptive rather than evaluative data" (p. 96). He notes that in public school 80% of the classroom supervision is done by line administrators because their jobs demand it. Therefore, this forced rater-ratee relationship needs to be nurtured and supported at the district level. McGreal's book provides several suggestions for pre and post conference strategies and reasons to provide the time to conference with career teachers differently than beginning teachers. He presented methods of collecting observational descriptive data through videotapes and of preparing actual observation records during on-going review training sessions.

Constable and Andrich (1984) note that observers agree more with each other following discussion of previous sets of ratings. These observers recognize areas of common

ground and discuss areas of disagreement. Such discussions allow the development of more common ground among these appraisers.

Crain (1990) completed a study through the University of North Texas focusing on establishing inter-rater reliability of the TTAS instrument at the indicator level during the training of 700 appraisers. He reviewed reliability at the indicator level, one of 55 identified teaching behaviors to observe within the instrument. Observations by the appraisers consist of a single video teaching lesson. He found 43 of the 55 indicators to have high inter-rater reliability and 49 to have reliability; therefore, he concluded that the TTAS instrument as a whole was highly reliable.

Another study involving inter-rater reliability and the TTAS is by Dobbs (1989), who used a quasi-experimental design to investigate the consistency of appraisers awarding exceptional quality points.

Other fields such as dentistry, medicine, and journalism which use observers have developed training centers and programs to enhance more consistent ratings. Observers who do not perform well in the field are sent back to these centers for retraining (Napier & Latham, 1986).

The range or distance between raters' scores has posed research concerns. Often restricting this range defines a method to improve inter-rater reliability (Atkinson & Mur-

ray, 1987). Supervisor, as opposed to peer or self, ratings provide a more restricted range (Tsui, 1983). Most research indicates that a supervisor or expert does a more consistent appraisal of performance than does a peer or oneself (Cathoun, 1988). The Austin, Texas, school district developed a system to allow for range differences by converting all the appraiser scores to z-scores; then the district adjusted these scores in relation to the mean (Ligon & Ellis, 1986). This method provided a pattern of consistent scores among the appraisers until the mandate of the TTAS instrument.

Studies by Medley and Mitzel (1963) and Cooper (1981a) concern the halo effect of appraisers. Halo effect is the generalizing of a ratee's performance based on an earlier impression or event. Borman (1975) states that the halo effect "is perhaps the most pervasive rater error" (p. 556). Brown (1968) used Guilford's definition when he described the halo effect and argued that training could decrease this effect: "A particular rater tends to rate a particular ratee similarly on all traits" (p. 195). Becker and Cardy (1986) argue in their research conclusions concerning halo and other rater errors that raters need to understand more fully the processes of appraisal. Ivancevich (1979) argues that "at least for the present sample, there appears to be a carryover from rater training to reduce halo and leniency error to actual ratings in an organizational setting" (p. 507). Murphy and Balzer (1986) suggest that a lengthy time

allowed between the actual observation and the written appraisal may produce halo effects or other general impressions.

Rater inconsistency instead of changes in the ratee performance results in the different teacher ratings over time (Landy & Farr, 1983; Nowakowski, 1983). Bulzer (1988) presents information that other tasks and factors interfere with a rater's appraisal. Interferences noted include impressions, memory, personality, social background, motivation, and the organization's purposes for appraisals. He suggests use of a diary to minimize the raters' storage and retrieval processes.

Several studies describe advantages of raters using diaries to enhance their recall of observations within classrooms (Bernardin & Walter, 1977; Guion, 1965). Others suggest that the use of diaries aids inexperienced raters in becoming more consistent with the experts and in developing a system of written documentation (Taylor, Parker, & Ford, 1959). Additional research highlights the concept that raters use diaries to record data consistent with their biases or preconceived schema of the ratee (De Nisi et al., 1989).

Buck and Tiene (1989) found no significance between gender or attractiveness and the perception of competence. Their study used pictures and written philosophies of teach-

ers before observations. Then raters observed and rated teachers.

Several researchers (Murphy & Balzer, 1986; Smither et al., 1989a) have studied the use of video in rating performance. Ratings immediately after watching a videotaped performance result in more consistent ratings than ones rated after a time lapse. However, even when raters viewed a video a second time, their ratings did not significantly change. Other research (Barnes, 1987) indicates that the reliability of raters is 6% higher with the use of a video than with a real field experience. He concludes that when appraisers are placed in real situations, more perceptions and personal issues result. Other research indicates that a real classroom observation poses more complex issues for the rater than does a video observation which obviously focuses on one picture at a time (Medley et al., 1984). Still others suggest that raters use videotaped observation to view and to discuss in groups to minimize training and later rating errors (Latham, Wexley, & Purcell, 1975). Medley et al. (1984) describe the limitation of video classroom observation in this way: "Observing a classroom on a TV monitor is much more like watching it through a keyhole than actually visiting the classroom" (p. 126).

Summary

The related research detailed concepts relating to the appraisal process. Research on real field appraisals is

limited. The state of Texas in 1989 requested the TTAS appraisal data on randomly selected appraisers throughout the local districts. Results from collection of this information has not been shared with the local districts at this time.

CHAPTER 3

METHODS AND PROCEDURES

The Texas independent school district in this study used whole group instruction and individual practice techniques in an attempt to improve consistency among its appraisers within the district. The instruction and the techniques provided by this training to the appraisers at the local level was in addition to the state mandated training necessary for their certification. This chapter describes the procedures for conducting the study and the local efforts to train its appraisers in order to improve their inter-rater appraisal consistency over a two-year period.

Population

The population of this study was 29 building-level administrators within a single, Texas school district. The school district consisted of 9,000 students and 580 teachers. Of these teachers, 246 hold master's degrees or have graduate hours above the master's. In 1987-88, 236 teachers were on career ladder Level II and 133 were on Level III. In 1988-89, 232 teachers were on career ladder Level II, and 155 were on Level III.

The school district's administrators served as teacher appraisers during the 1987-88 and 1988-89 school years.

Since the Texas legislature requires the use of the Texas Teacher Appraisal System instrument, all these administrators have received state training and certification as TTAS appraisers. Since the TTAS instrument requires two appraisers for the teachers, the district appraisers functioned both as teacher supervisors and "other" appraisers. The local school board designated 14 of the appraisers as teacher supervisors and 15 as "other" appraisers for the purposes of implementing the TTAS instrument.

The teacher supervisor evaluates the teacher after a classroom observation, keeps the documentation for each teacher, and is responsible for combining each teacher's yearly evaluations into one appraisal record. The teacher supervisor's evaluation score counts as 60% of the teacher's overall performance score or total evaluation. The "other" appraiser also evaluates the teachers and contributes the remaining 40% to the overall evaluation score.

Three secondary principals, eight assistant secondary principals, nine elementary principals, and nine elementary assistant principals from 12 schools provided appraisal data. All but eight received training together for a week in the initial statewide TTAS training. All received the update training provided by the state in 1987, 1988, and in June 1989. Of these administrators, 24 hold master's degrees while five have their doctorates or graduate hours above the masters. The classroom experience of these ap-

praisers totals 275 years. Their teaching experience combined with their educational administrative experience covers 520 years. The gender composition of this population consists of 11 female and 18 male administrators.

Permission for Conducting The Study

The writer contacted one of the assistant superintendents within the Texas school district studied in order to receive permission for using information concerning the district's TTAS appraisers. The assistant superintendent met with the writer for discussion on several occasions and gave written permission for use of the district's TTAS data. This permission contained the limitation that the study could not identify the district or any district employee by name and that the confidentiality of those individuals included in the data be maintained throughout the two years.

The writer then sent the original permission letter to the Office of Research and Academic Grants, Institutional Review Board at the University of North Texas. The Institutional Review Board's chairman sent a written approval for the research and provided an exemption from further review by the university.

Procedures for Implementation of The Study

The Texas legislature mandated the State Board of Education to develop an instrument to be used for teacher appraisal. Texas Education Agency personnel developed the TTAS system. Personnel from regional and state agencies

trained all appraisers within the state to use the TTAS system. After classroom observation, each appraiser completed an observation record which documented the teacher's performance. The district established an official calendar for the two required appraisal periods. This district determined its appraisal periods during 1987-88 as September 16 through January 15 as the first and January 19 through May 13 as the second and during 1988-89 as September 15 through January 18 as the first and January 23 through May 13 as the second.

The personnel office staff organized lists of the administrators who functioned as supervisors and the "other" appraisers. These lists not only informed the teachers of their appraisers but also delineated which appraisers' observation records counted as the supervisor's with a weight of 60% and with the "other" appraiser's with a weight of 40%. Next, the district provided training to all appraisers so they could record and enter their appraisals on computers.

Appraisers, according to the school board policy, scheduled all teacher observations and held conferences with each teacher concerning the written observation record. The appraisers discussed and analyzed the scoring process for the observation record with teachers to develop a more consistent view of each item on the TTAS instrument.

At the end of every appraisal period, the teacher supervisors, usually the building principals, converted their written observation records to evaluation records. Then, they combined their records with the evaluation records of the "other" appraisers within the building into a written appraisal record. This procedure was repeated at the end of the second appraisal period. Finally, the teacher supervisors collected the computer disks containing data for each appraisal period and transferred them to yet another disk for the final appraisal record with an overall performance score for the year. The personnel office compiled these appraisal record disks from each building at the central office. A special project director organized the data from these building disks by school, district, individual appraiser, and grade-level.

The writer was given the compiled appraisal data for 1987-88 and 1988-89. She assigned a number to each appraiser and a letter to each school for identification of data and record keeping purposes. With the cooperation of the personnel office, the writer collected other data for analysis including career ladder placement of teachers, teaching and administrative experience of the appraisers, and college degrees held by both teachers and appraisers. Charts and tables according to school, district, and individual appraisers display these data and the TTAS data in various comparisons.

Description of District Training of Appraisers

District staff development presenters required the 29 appraisers to consider the importance of their inter-rater reliability beginning in the fall of 1987. The appraisers' training included viewing teaching videos, scoring the TTAS instrument individually, debriefing the scored observation record item-by-item in groups, watching videotapes for teaching behaviors that demonstrate exceptional quality, and discussing the items that the appraisers considered difficult to score or the items they inconsistently scored as a group.

In, 1988, after this awareness phase had concluded, the appraisers, as a group watched videotapes and practiced debriefing. Their debriefing discussions were limited to the exceptional quality areas scored. The appraisers decided that instead of using videotapes throughout the year as in 1987-88, they would use actual classroom situations to work on their individual inter-observer reliability or consistency in exceptional quality scoring. An assistant superintendent organized the appraisers into six teams. These teams included a mixture of elementary and secondary appraisers. A designated team leader would schedule a classroom observation for the entire group. Each team of appraisers would observe together, score the TTAS instrument individually, and record their collective scores per item on

one sheet. Then the leader would send the individual and collective written results to the central office.

Later in the year for training purposes only these teams would schedule and observe teachers who volunteered to allowed them in their classrooms. These practice team observations were done during a specified week during each of the months: September, November, January, and February. During each of these designated weeks of observation, the membership of the six teams would change for the next observation. The designated team leader would contact the teacher, remind the other appraisal team members of the time and place of the classroom observation, and return the scoring information to the central office. The number and incidence of exceptional quality points assigned by the individual appraiser and the consistency among these team members, as recorded on the collective record of each team, provided the information for the appraisers to identify the items they would debrief within their individual teams.

The district used other treatments to reinforce appraisal consistency and to make the appraisers aware of their commitment to have their scores more consistent. The appraisers held a conference with every teacher observed during the year in order to explain and to defend the scoring process. The "other" appraisers in one secondary building rotated the teachers whom they appraised each year. All appraisers participated as substitute teachers for four days

during the year to better understand the teaching behaviors within the TTAS instrument. All appraisers observed for an entire block of 45 to 55 minutes in each classroom. All of these classroom observations for TTAS were scheduled by both the teacher supervisors and the "other" appraisers in writing at least three days before the actual observation. The district organized a task force in 1987 to prepare guidelines for the interpretation of exceptional quality performance. The committee completed its work in January 1987 and revised the guidelines in August 1987.

The variety of the local district training provided different methods for assisting appraisers in becoming aware of their areas of inconsistency in scoring. Furthermore, the local training efforts provided means of increasing the contact appraisers had with each other. Elementary joined secondary appraisers on various school campuses in observation teams for evaluating and discussing their appraisal results.

Procedures for Local Appraiser Training

Throughout this two year study the local district organized and presented additional training and treatments to these 29 appraisers. The training focus was to provide practice with developing observational skills. The district used videotapes of teachers within and outside the district. Efforts were made to balance the grade-levels of the video-

tapes and the team observations. All appraisers were encouraged to bring videotapes of sample teaching from their own schools to place in the pool for observation during the year. Appraisers were encouraged to state and to defend their interpretation and opinions concerning the teaching behaviors which they observed. The discussions among the appraisers and the analysis of results of the whole group were intended to improve their inter-rater consistency. The assistant superintendent reminded the appraisers of this intention at closure each time and indicated that he hoped carryover into the actual classroom observations would result.

To begin the 1987-88 appraisal year, the 29 district appraisers reviewed a revised document prepared by a district task force concerning the district's procedures for the interpretations of exceptional quality (See Appendix A for complete procedures). This task force consisted of teachers and administrators from elementary and secondary schools. This document presented the appraisers with written questions to consider before their awarding exceptional quality points for each TTAS indicator or item. The revision reflected changes by the State Board of Education during the summer of 1987; however, the changes did not significantly alter the content or intent of the original document prepared in January 1987. The foreword of this "exceptional quality" document stated its purpose:

This document was developed as a supplement to the official appraisal instrument and is not intended to serve as the sole basis for the interpretation of indicators. When used in concert with the official instrument, we believe that the process will be greatly enhanced and that all parties to the system will benefit from a higher degree of accuracy and consistency. (see Appendix A for complete document)

The district appraisers received the guidelines for the 1987-88 appraisal calendar as adopted by the local school board. (see Appendix B for complete guidelines) These guidelines designated Appraisal Period 1 to be September 16, 1987 through January 15, 1988 and Appraisal Period 2 to be from January 19, 1988 through May 13, 1988. The assistant superintendent assigned the building administrators as either teacher supervisors or "other" appraisers for the 580 teachers.

The guidelines required all observations of teachers to be scheduled in writing with a minimum of three days notice. Following the observations, appraisers held conferences with all teachers. The appraisers discussed with the teachers the indicators on the TTAS which were in need of improvement as well as indicators observed to be presented effectively.

Each month the 29 appraisers met to view teaching videotapes obtained by the assistant superintendent. The

appraisers scored the tapes individually; the assistant superintendent recorded the results of each item on a chalk board. One of the administrators who was also a state trainer assisted in the recording and discussion of data. All present discussed the results and defended their opinions. Appraisers sat randomly for each videotaped viewing. All took notes and wrote documentation notes for the denial of credit for an indicator on the TTAS instrument or for rewarding of exceptional credit points. These procedures and treatments continued throughout the school year.

The building-level appraisers began the 1988-89 school year by studying a statistical summary for each appraiser's and school's TTAS data for the preceding appraisal periods. The assistant superintendent expressed the need for greater inter-rater consistency.

Staff development on exceptional quality followed this data distribution. The appraisers listened to speakers, received the document *Exceptional Quality: A Second Step* (Crain, Worley, & Dunn, 1987), discussed the meaning of exceptional quality, and practiced writing documentation for awarding exceptional quality points. One of the staff development speakers (Crain, 1987) described six assumptions to consider before awarding exceptional quality points:

1. The decision to award exceptional quality will ultimately be a subjective professional judgement.

2. The base point for making the decision about exceptional quality is the standard expectation.
3. The standard expectation is the standard of effectiveness.
4. The decision to award exceptional quality rests primarily on observed student behavior.
5. The real test of the exceptional quality decision is the ability to write documentation.
 - a. What student behavior was observed that suggests exceptional quality?
 - b. What teacher behavior may have led to the exceptional quality student behavior?
6. The decision to award exceptional quality is now scored at the criterion level. (Crain et al., 1987).

The appraisers in 1988-89 decided not to watch and score videotaped teaching lessons as in 1987-88. The district decided to extend the field practice idea required by the state at the opening of each school year for appraiser certification. All appraisers observed in field teams at different schools (see Appendix C for complete field team schedule). This actual in classroom and in district practice was intended to enhance inter-rater consistency within the local district. The more frequently the appraisers observed together and discussed their results, the closer their scoring should become. Each month each appraiser met

with a different multi-grade level evaluation team to observe a teacher and score and discuss the TTAS indicators and exceptional quality points. These elementary and secondary appraisers sent their evaluations to the central office for compilation. Each team observation lasted 45 minutes while each team discussion usually lasted one to two hours with additional time as needed to prepare the written results. All appraisers participated on five different campuses and were distributed among six different teams throughout the 1988-89 school year.

The appraisers also received their 1988-89 TTAS guidelines which were the same as the 1987-88 guidelines except for the dates and the stipulation that higher career ladder teachers were to receive only one appraisal. The teachers on levels 2, 3, or 4 received only one appraisal "uniformly extended throughout both scheduled appraisal periods with one observation completed in the fall and one in the spring" (see Appendix D for complete guidelines).

The fall of 1988 began for these 29 appraisers with a visual presentation of charts and graphs of the district, school, and individual statistics concerning the 1987-88 appraisals of all district teachers. Each appraiser received a packet which included information on the rankings of all the appraisers. These rankings focused on the appraisers' frequency of assigning exceptional quality to the teachers whom they appraised. The appraisers were ranked

from the most to the least positive appraiser in assigning these scores. The assistant superintendent explained the range of differences among individual appraisers and 12 schools, especially highlighting the exceptional quality data. Graphic forms displayed the differences and similarities among the appraisers.

Appraisers self-analyzed their individual scores and compared them to the other 28 appraisers, their school's scores in relation to the other schools, and both their individual and school's scores in relation to district average scores. Appraisers studied the items on which they scored less consistently than other appraisers. They considered these differences when working on their field teams and while appraising their designated teachers for the 1988-89 year.

Procedures for the Analysis of Data

An analysis of exceptional quality appraisal data from the independent school district in 1988-89 compared with the 1987-88 was made in order to determine if there is a difference between of the scores for these two appraisal years. This study uses only those administrators with appraisal scores for both year periods in the statistics. Each appraiser kept the same identification number both years while each school retained the same letter. The writer organized the scores of these appraisers in an ordinal pattern and has assigned each a ranking from the most to the least positive

for each year. The analysis investigates the patterns and stability of these appraisers' rankings for each year and between the two years. In addition, the study compares the schools and the grade-levels according to the exceptional quality scores and rankings of these appraisers. Tables and charts in Chapter 4 illustrate the comparisons of scores between these two years. The identification numbers and letters of the schools and appraisers assisted in tracking their scores as individuals and as part of a school appraisal team.

Other Data Collected for Study

Through the cooperation of the district's personnel office, this writer examined other data for this study. Information concerning building level administrators consisted of their experience levels in both teaching and administration and their experience with the local district. Their educational level, gender, and grade assignment were also noted.

The information concerning the teachers to be appraised consisted of their level of education, teaching experience, and the number of teachers on each level of the career ladder. School information included the number of teachers and administrators per campus, the size of the student population, and the percentage of teachers with master's degrees.

The collection of the above data enabled comparisons of the 29 appraisers in the study to be made. The next chapter presents the findings from these comparisons.

CHAPTER 4

PRESENTATION OF FINDINGS

The chapter presents an analysis of the appraisal data for the two years of this study. One of the purposes of this study is to provide and to compare information concerning the degree of inter-observer reliability among the appraisers and schools within the independent school district. The other purposes include comparing the range and ranking of appraisers' scores from most to least positive and determining the effectiveness of additional local training for increasing inter-observer consistency. This chapter presents the analysis of each research question posed in Chapter 1.

Research Question One

Research question one states, "Do appraisers within the same building who observe the same set of teachers evaluate these teachers' performances similarly?" The writer graphed the exceptional quality appraisal data to show the range from most to least positive and percentage of possible exceptional quality scores granted to teachers each year. Table 1 presents the scoring data for each appraiser included in the study. Figures 2 through 13 display the exceptional quality scores of the appraisers within each building which is identified by a designated school letter.

Table 1

Appraiser Scoring Patterns

APPRAISER	SCHOOL	# OF TEACHERS AY 1988	AY 1989	MAXIMUM AY 1988	EQ POINTS AY 1988	EQ POINTS AY 1989	GRANTED AY 1988	EQ POINTS AY 1989	PERCENT AY 1988	GRANTED AY 1989
1	E	32	35	864	945	945	838	614	97%	65%
2	E	32	35	864	945	945	812	614	94%	65%
3	A	33	37	891	999	999	659	589	74%	59%
4	A	33	37	891	999	999	659	549	74%	55%
5	I	35	35	945	945	945	652	539	69%	57%
6	G	37	42	999	1,134	945	659	601	66%	53%
7	C	34	35	918	945	945	588	595	64%	63%
8	J	64	61	1,728	1,647	1,647	1,071	906	62%	55%
9	C	34	35	918	945	945	532	444	58%	47%
10	J	32	31	864	837	837	501	452	58%	54%
11	J	37	42	999	1,134	945	559	612	56%	54%
12	G	35	35	945	945	945	501	529	53%	56%
13	I	38	37	1,026	999	999	523	480	51%	48%
14	H	40	39	1,080	1,053	999	540	369	50%	35%
15	L	40	37	1,080	999	999	450	519	49%	52%
16	B	34	37	918	999	999	503	490	49%	49%
17	H	38	37	1,026	810	810	415	316	48%	39%
18	J	32	30	864	999	999	441	500	48%	50%
19	B	34	37	918	999	999	454	616	42%	57%
20	F	40	40	1,080	1,080	1,080	454	583	42%	54%
21	F	40	40	1,080	1,080	1,080	432	411	40%	39%
22	K	40	39	1,080	1,053	810	316	275	39%	34%
23	K	30	30	810	810	810	389	442	36%	42%
24	L	40	39	1,080	1,053	1,053	378	579	35%	55%
25	L	40	39	1,080	1,053	1,053	369	658	35%	58%
26	D	39	42	1,053	1,134	1,134	337	590	32%	52%
27	D	39	42	1,053	1,134	1,134	251	316	32%	39%
28	K	29	30	1,783	810	810	494	713	31%	44%
29	K	59	60	1,593	1,620	1,620	292	390	27%	37%
	L	40	39	1,080	1,053	1,053				
DISTRICT TOTALS		1,090	1,117	29,430	30,159	30,159	15,069	15,291	51.2%	50.7%

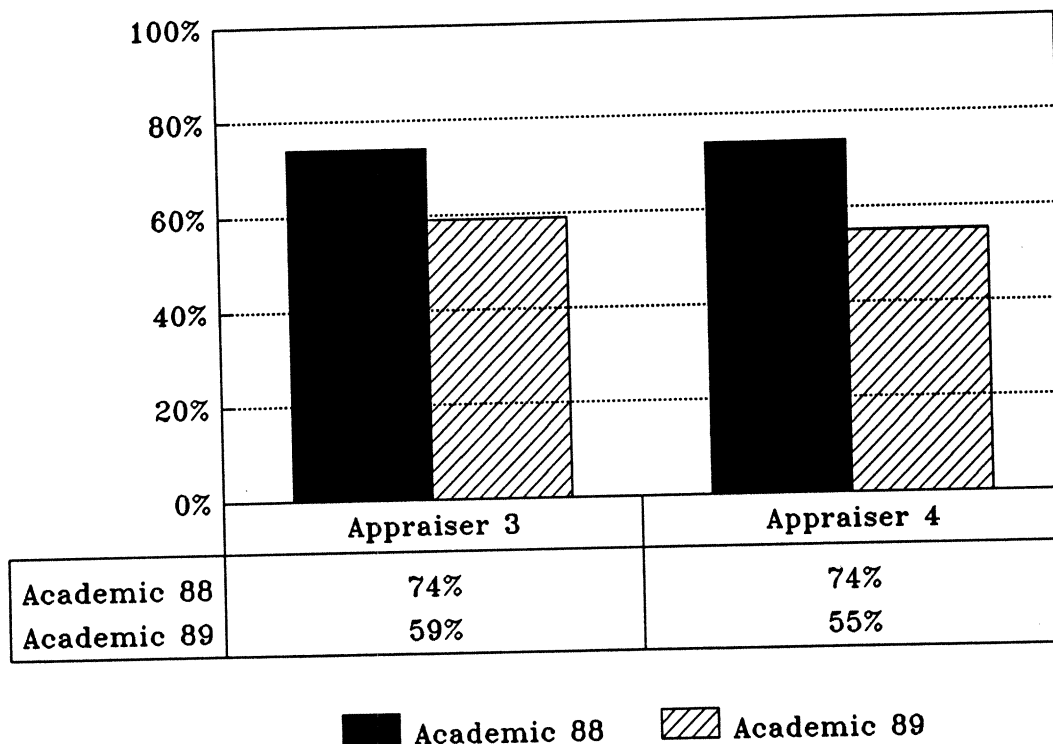
Schools A through I are elementary schools. Schools J, K, and L are secondary schools. The appraisers are numbered from one to 29 according to their rank from most to least positive in awarding exceptional quality points in 1987-88 (Table 1). Figures 2 through 13 present the results for 1987-88 and 1988-89 and indicates the numerical differences in scoring. For the purposes of this analysis, the exceptional quality used in these figures represents the number of times each appraiser granted exceptional quality points taken as a percentage of the total possibilities of so granting. This eliminated the disparity among appraisers of the number of teachers evaluated each year.

Figure 2 describes School A and appraisers 3 and 4. Both appraisers scored similarly over both years and reduced their awarding of points by 15 and 19 percentage points in 1988-89.

Figure 3 illustrates School B and appraisers 15 and 18. These appraisers scored consistently within two percentage points of each other in 1987-88 and 1988-89. These appraisers scored more consistently than any of their fellow appraisers.

Figure 4 illustrates School C and appraisers 7 and 9. Although these appraisers scored fewer exceptional quality points in 1988-89, they scored differently from each other by 6 and 16 percentage points in 1987-88 and in 1988-89, respectively.

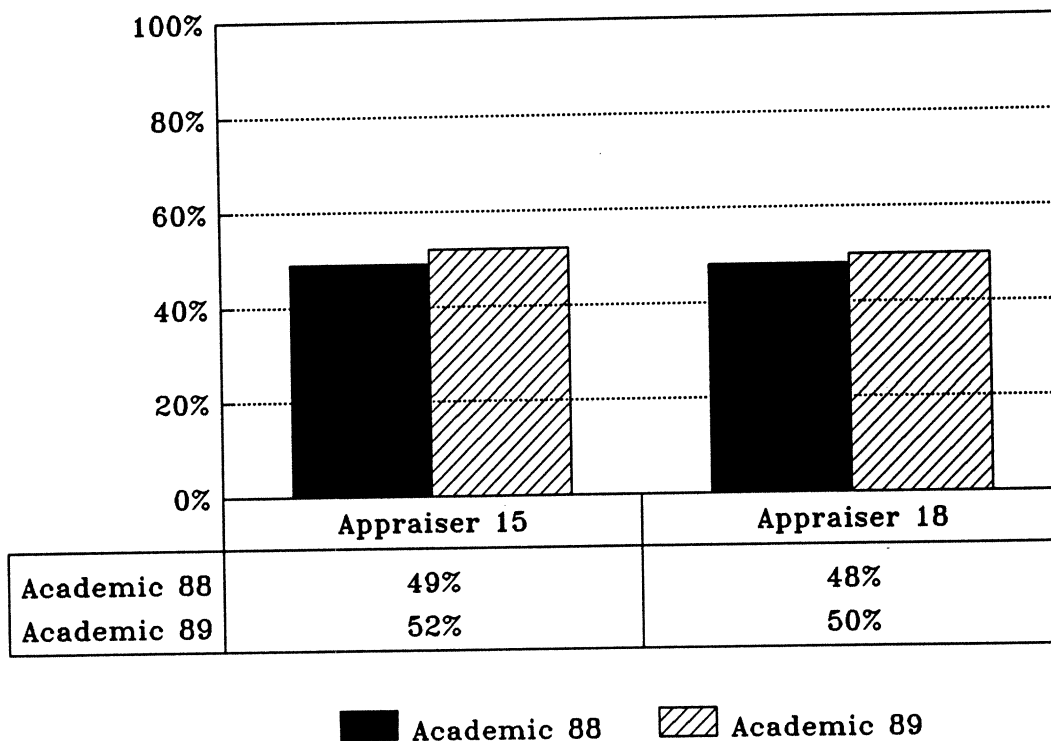
Figure 2. School A: Comparison of % of possible E Q granted



School D's scores include appraisers 25 and 26 (Figure 5). These appraisers scored within six percentage points of each other both years. They increased their awarding of exceptional quality by 23 and 29 percentage points in the second year, respectively. This school was the lowest of the nine elementary schools in awarding exceptional quality points in 1987-88.

Figure 6 illustrates the scoring results of School E and appraisers 1 and 2. These appraisers reduced their scores the most in 1988-89 from 1987-88. Appraiser 1 reduced scoring by 35 percentage points while Appraiser 2

Figure 3. School B: Comparison of % of possible E Q granted



reduced scoring by 29 percentage points. These appraisers scored consistently within three percentage points of each other for both years of this study.

Figure 7 charts School F and appraisers 19 and 20. These appraisers increased their scoring by 15 and 12 percentage points respectively in 1988-89. However, these appraisers scored within three percentage points of each other in 1988-89 and scored the same number in 1987-88.

Figure 4. School C: Comparison of % of possible E Q granted

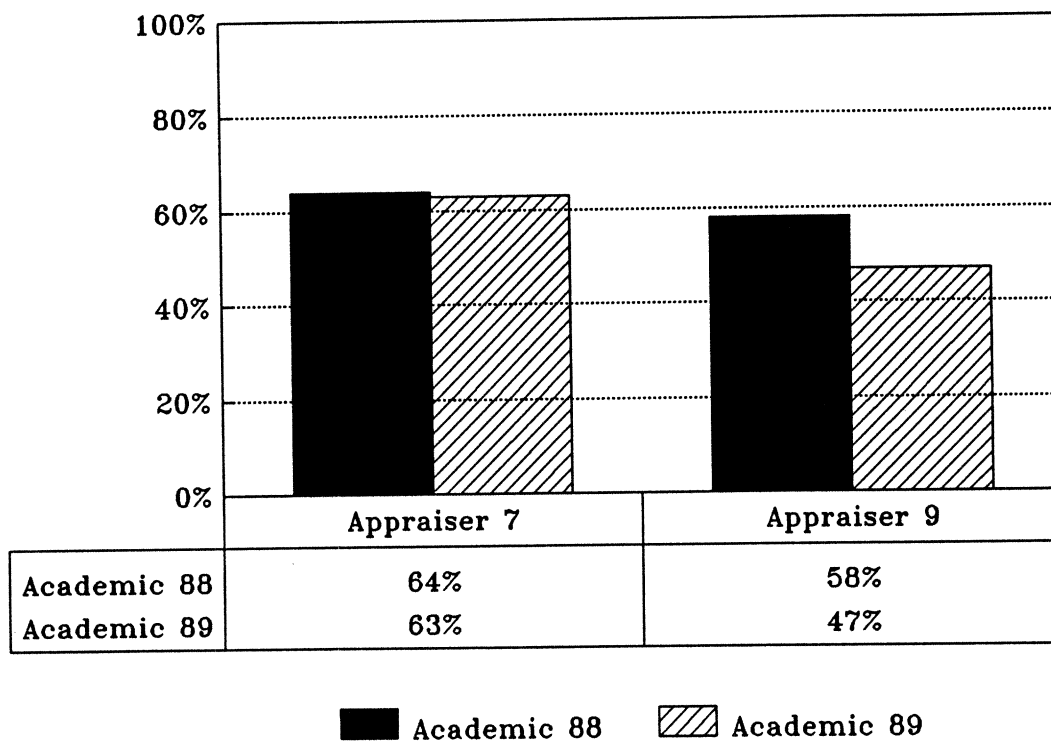


Figure 8 plots the score of School G and appraisers 6 and 11. These appraisers reduced their scores by 15 and two points in 1988-89 from 1987-88. Appraiser 6's change in scoring resulted in their scores being within one percentage point of each other in 1988-89.

Figure 9 reflects little change in School H and appraisers 13 and 16. These appraisers scored similarly each year. They were only two percentage points apart and scored in the middle range of appraisers each year.

Figure 5. School D: Comparison of % of possible E Q granted

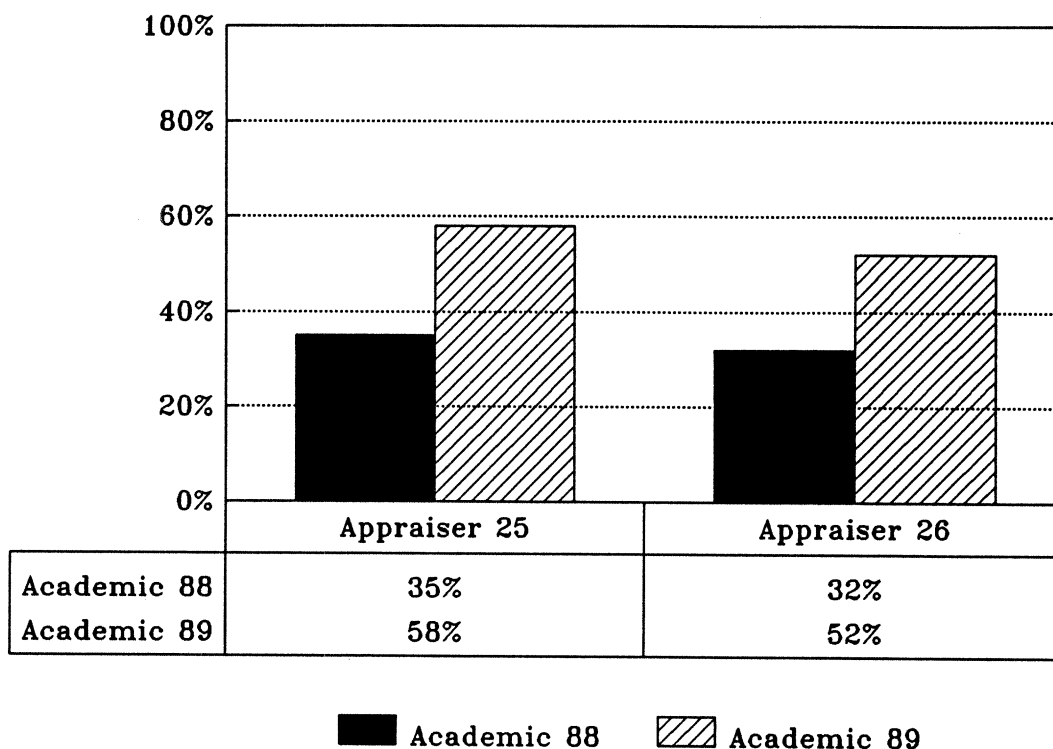
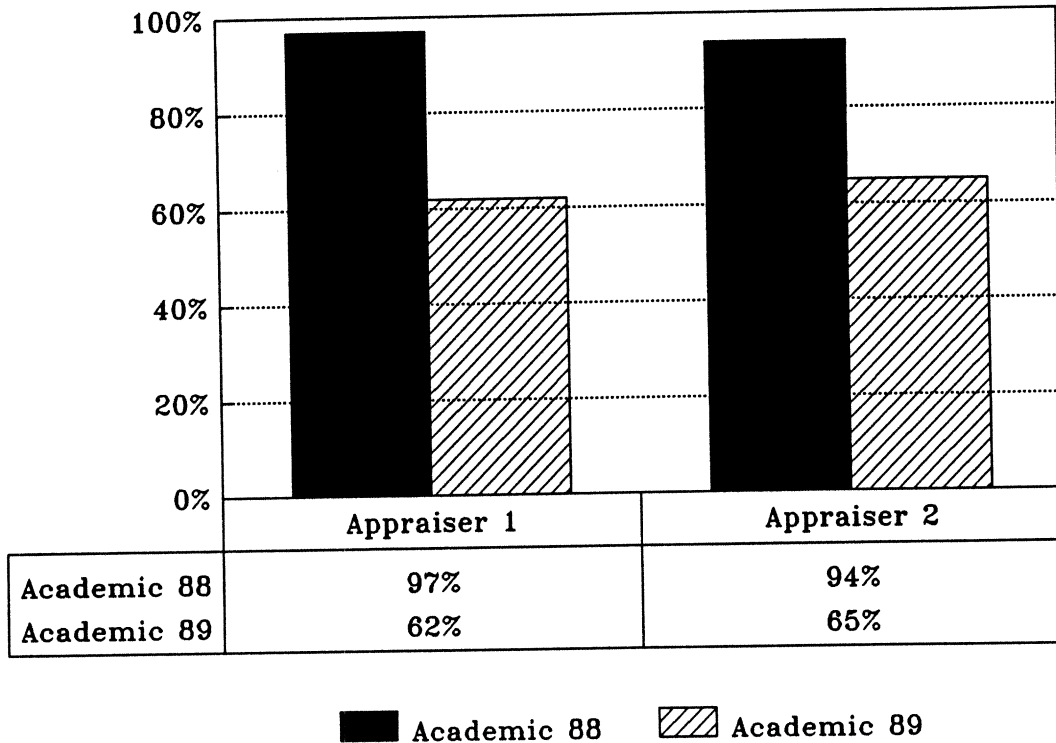


Figure 10 describes School I and appraisers 5 and 12. Appraiser 5 reduced scoring by 12 percentage points and Appraiser 12 increased scoring by three points. In 1988-89 these appraisers scored within one percentage point of each other.

As seen in Figure 11, School J with appraisers 8, 10, and 17 reduced scoring in 1988-89. Appraisers 8 and 10 scored within one percentage point in 1988-89; however, appraiser 17 differed with them by 16 percentage points. Figure 12 illustrates the scoring of School K with apprais-

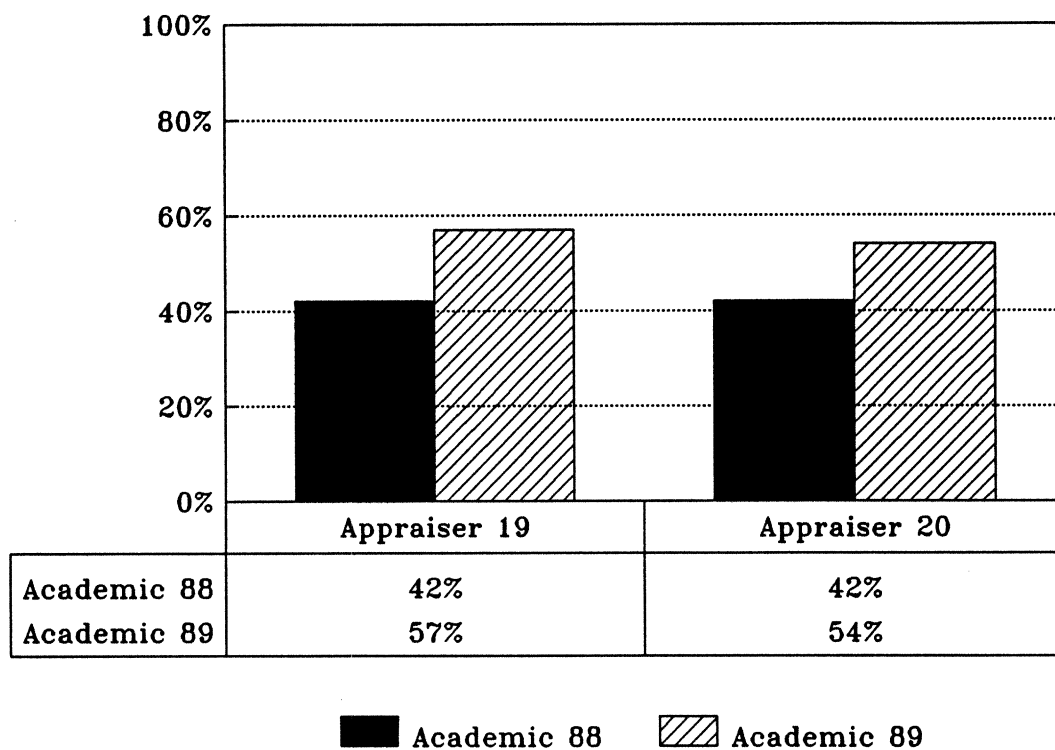
Figure 6. School E: Comparison of % of possible E Q granted



ers 22, 27, and 28. This school changed the designated teachers of the "other" appraisers 22 and 27 by rotating their lists of teachers each year. Their scores are similar in consideration of this rotation system. The teacher supervisor increased the awarding of scores by 13 percentage points.

Figure 13 describes School L with appraisers 14, 21, 23, 24, and 29. One appraiser who was not employed for both years of the study is omitted from the data. Four of the five appraisers increased their scores in 1988-89. Appraiser

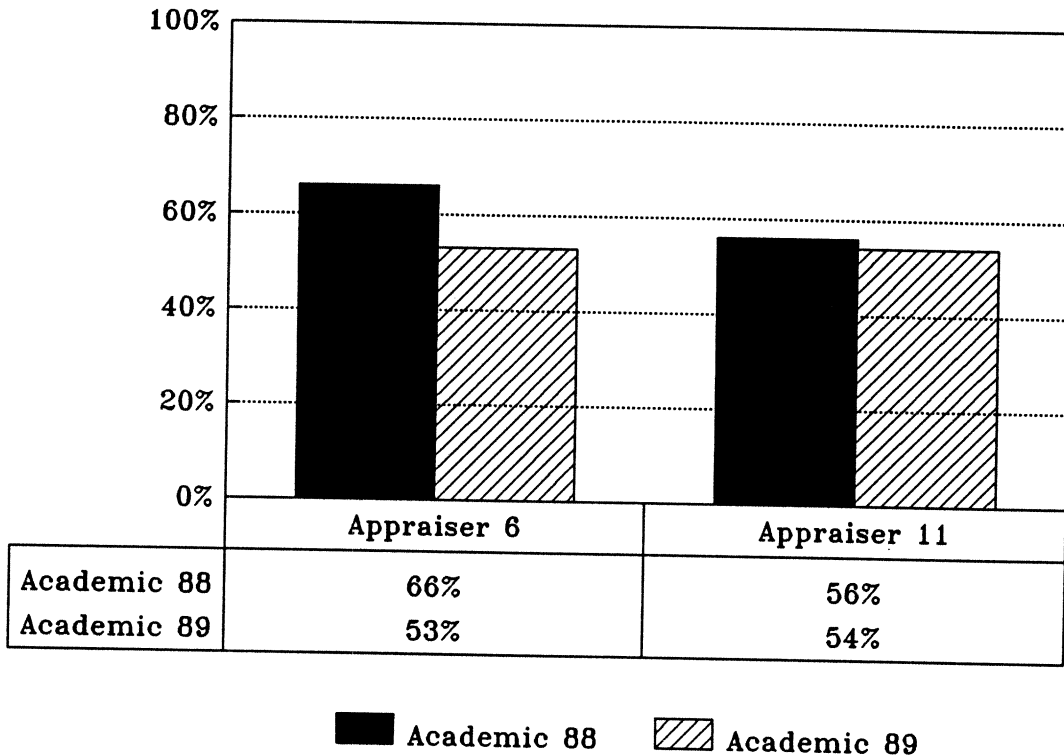
Figure 7. School F: Comparison of % of possible E Q granted



21 reduced scoring by only one percentage point. The increased scores ranged from 6 to 20 percentage points from 1987-88 to 1988-89.

Figures 2 through 13 indicate the percentage of the possible exceptional quality points granted among the 29 appraisers. Nine of the appraisers varied three percentage points or less in their awarding of exceptional quality points between 1987-88 and 1988-89. These nine appraisers represented seven schools. The scores of appraisers from four schools varied 10 or more points in 1987-88 and in

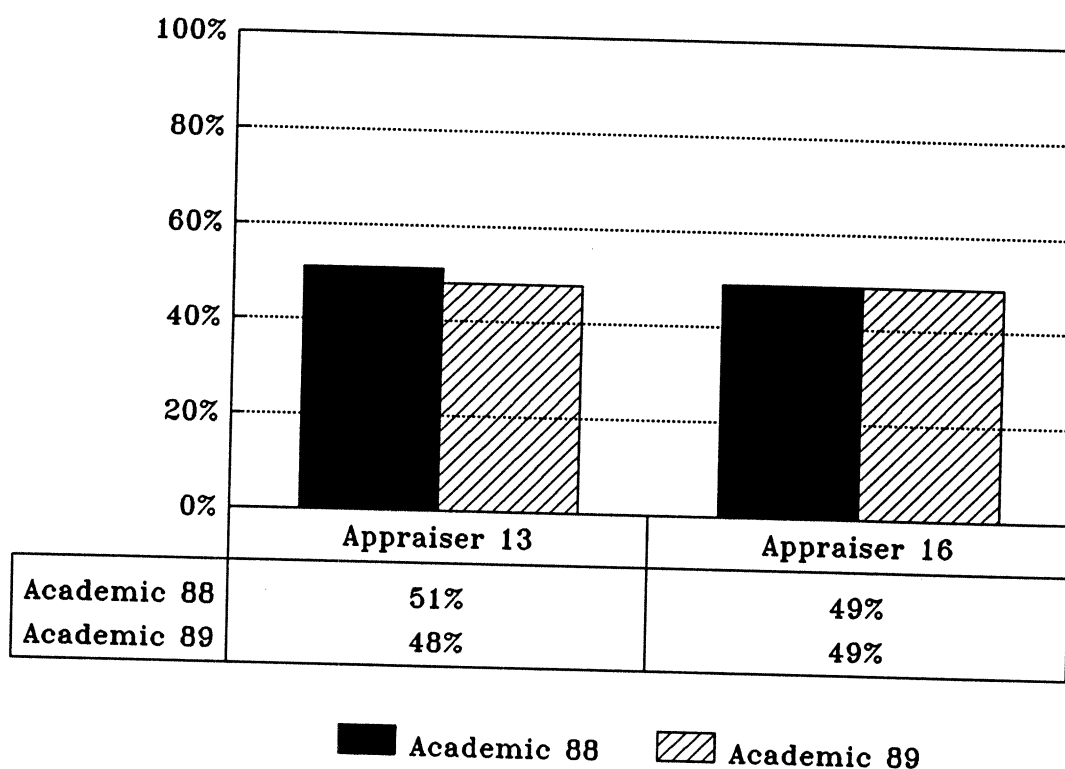
Figure 8. School G: Comparison of % of possible E Q granted



1988-89. The difference between the 1987-88 and 1988-89 scores for elementary appraisers indicated the greatest reduction in exceptional points granted. Elementary schools B, D, E, F, and H differed by three or less percentage points from 1987-88 to 1988-89. Of the remaining four elementary schools, Schools G and I had inconsistent scores both years; however, their scores were more consistent the second year by 9 and 15 percentage points respectively.

In the secondary schools L and J the appraisers evaluated the same teachers within their respective buildings

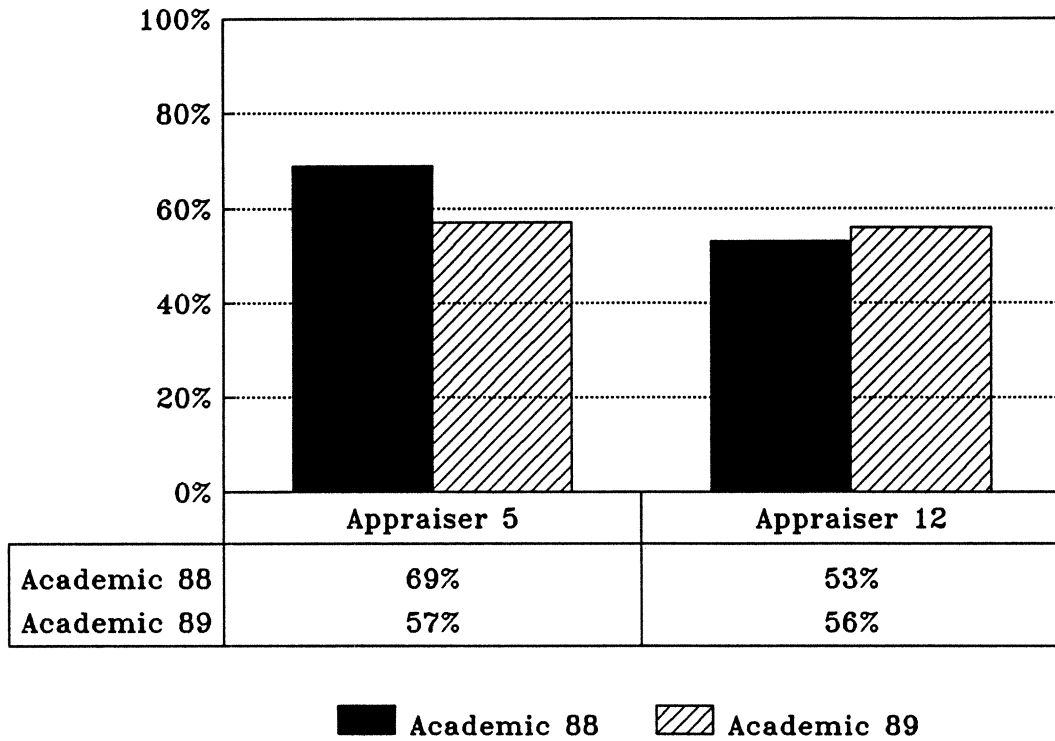
Figure 9. School H: Comparison of % of possible E Q granted



both years. Schools J and K had one teacher supervisor while L had three. The teacher supervisor with the "other" appraiser represent the two required evaluators within the TTAS evaluation system. In School K the teacher supervisor evaluated the total faculty both years while the "other" appraisers divided the staff in half and each one evaluated half of the faculty one year and the other half the next year.

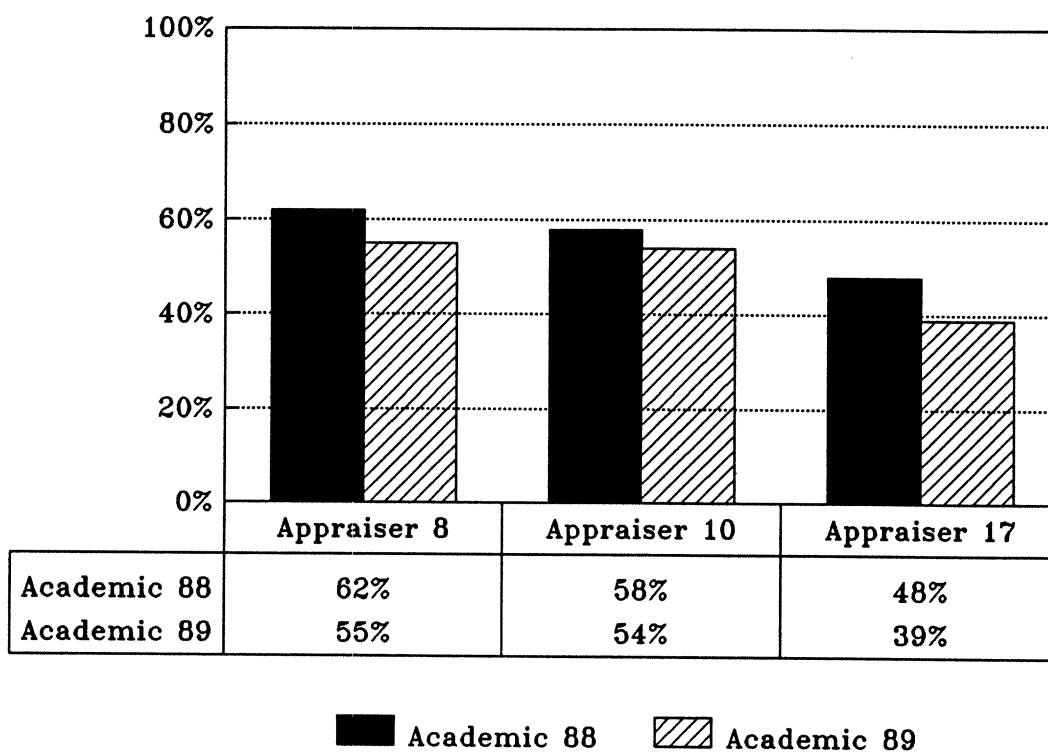
To compare the "other" appraisers with that of the teacher supervisors in School J and L, the writer noted the

Figure 10. School I: Comparison of % of possible E Q granted



similarity between the teacher supervisor's scores and the average of the "other" appraisers'. Since they were evaluating the same teachers, the consistency of the building appraisers was compared both years of the study. In School J, the average of the two "other" appraisers was 53 in 1987-88 and 46.5 in 1988-89 while the teacher's supervisor's score was 62 in 1987-88 and 55 in 1988-89. In School K the average of the two "other" appraisers was 35.5 in 1987-88 and 36.5 in 1988-89 while the teacher supervisor's average was 31 and 44 respectively. In School K, the "other" ap-

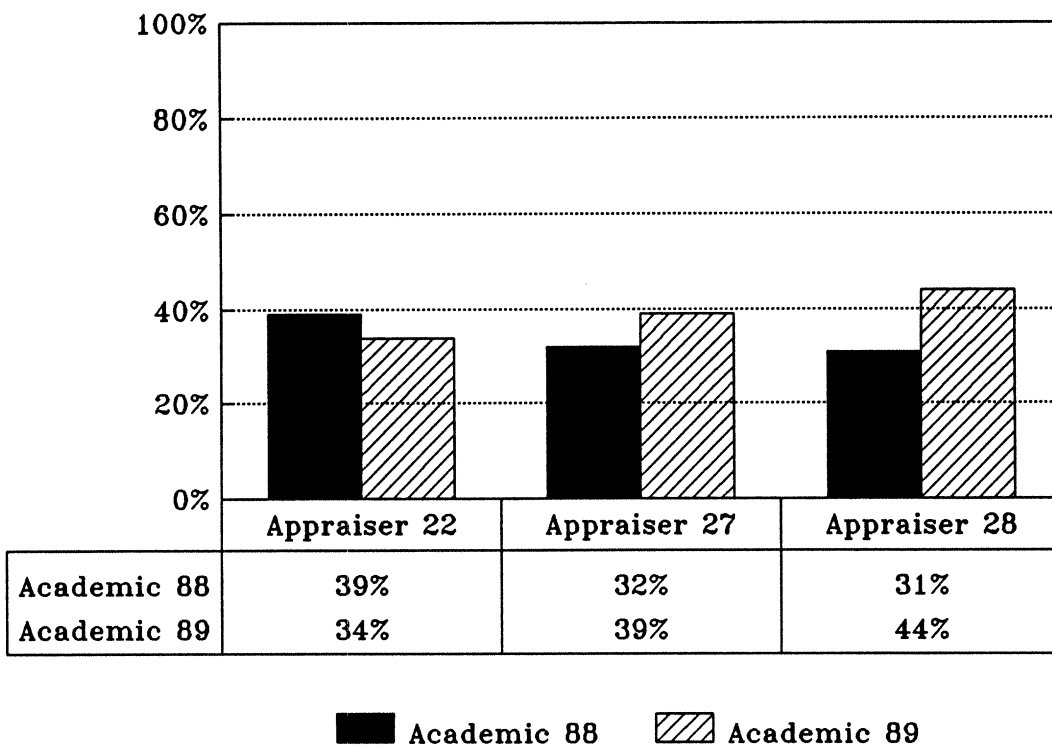
Figure 11. School J: Comparison of % of possible E Q granted



praisers scored more consistently with each other than with the teacher supervisor. All three appraisers for the secondary School J reduced their percentage of exceptional quality points for the second year.

The appraisers within the elementary schools evaluated the same set of teachers each year. The secondary School L experienced personnel shifts with its "other" appraisers so three teacher supervisors and only two "other" appraisers are included in the data analysis. Secondary School K divided the teachers between the two "other" appraisers and

Figure 12. School K: Comparison of % of possible E Q granted

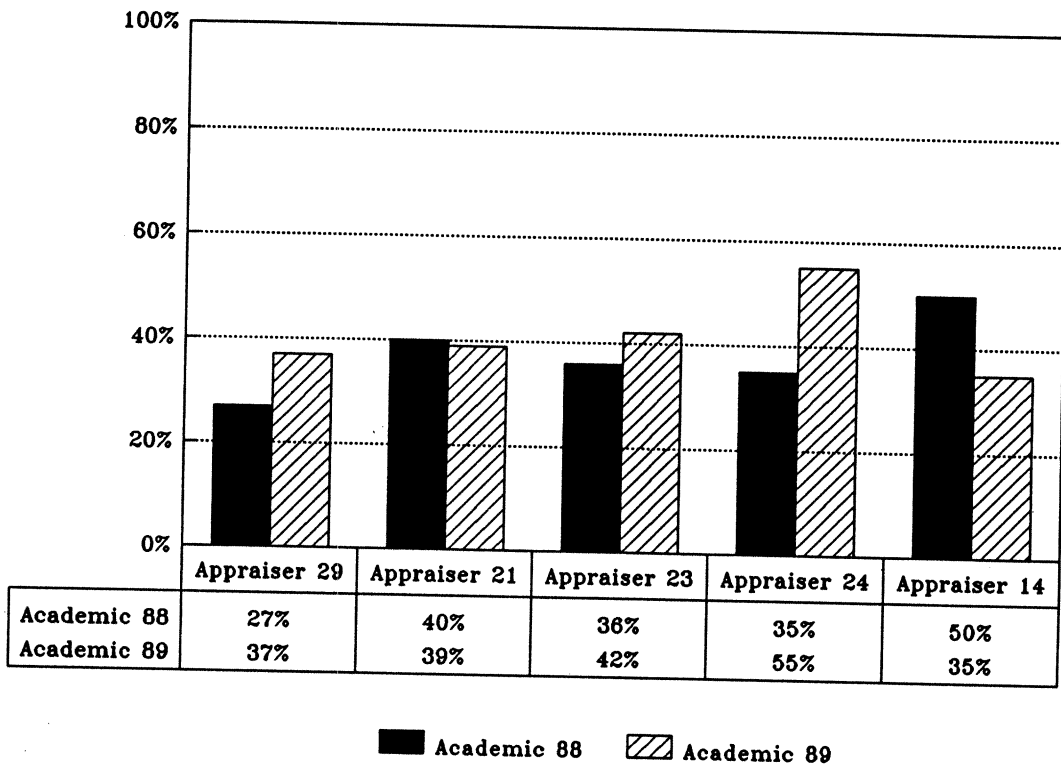


rotated the two lists of teachers each year. Secondary School J divided its teachers between the two "other" appraisers but did not change the lists each year.

Research Question Two

The second research question asks, "Do elementary appraisers have similar or different patterns of evaluation from secondary appraisers?" Figure 14 illustrates the collective elementary and secondary appraisal scores. The elementary total percentage of possible exceptional quality

Figure 13. School L: Comparison of % of possible E Q granted



points (Figure 15) in 1987-88 was 58.67% and in 1988-89 was 55.11%. The secondary percentage in 1987-88 was 46.33% and in 1988-89 was 45%. The shift in scores in the second year shows that the elementary appraisers awarded fewer positive points than they awarded during the first year. The secondary appraisers awarded fewer positive points to their teachers than they did the previous year. In both years the elementary appraisers awarded higher scores to their teachers than did the secondary appraisers.

Figure 14. Percent of possible E Q points granted by school

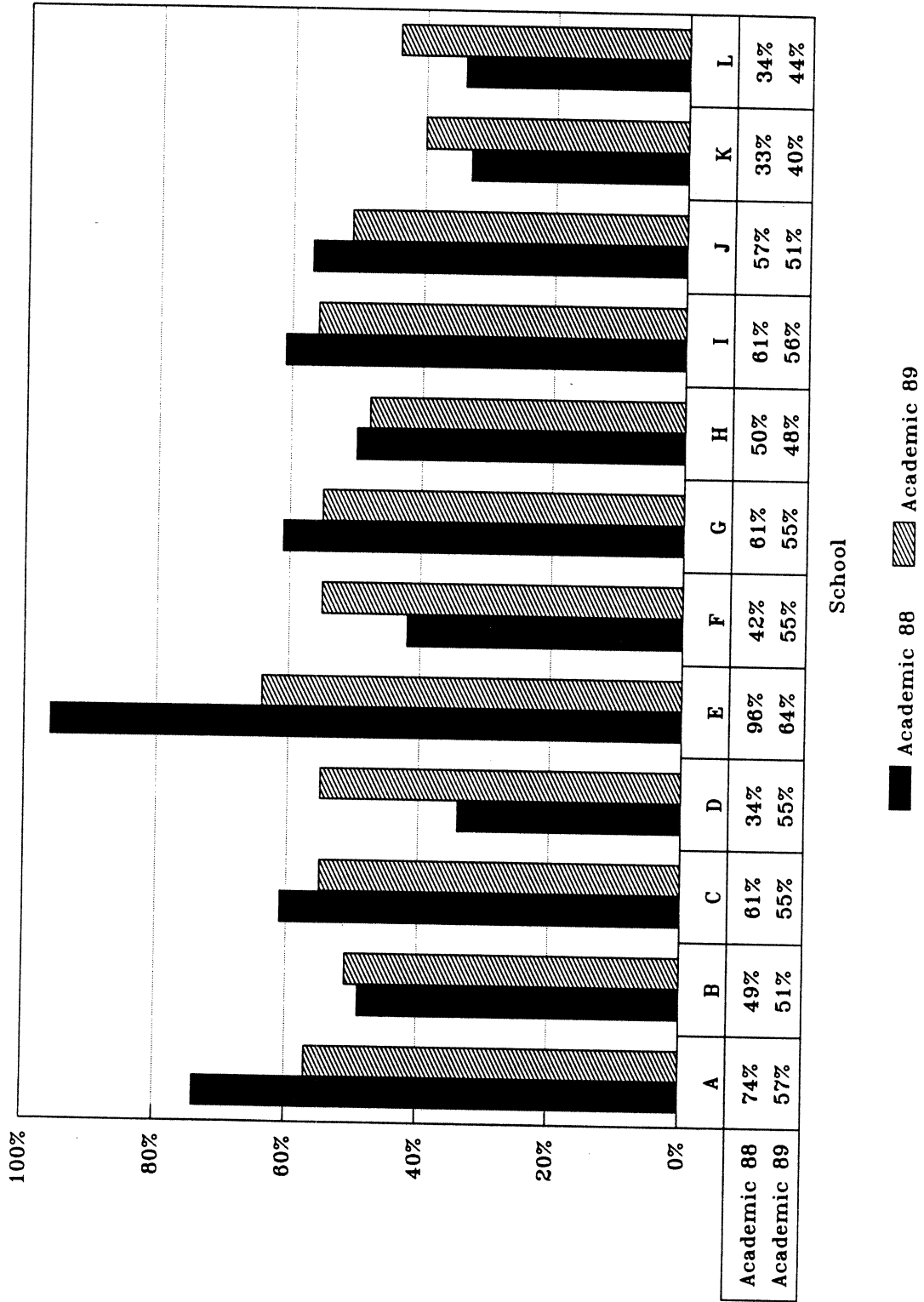
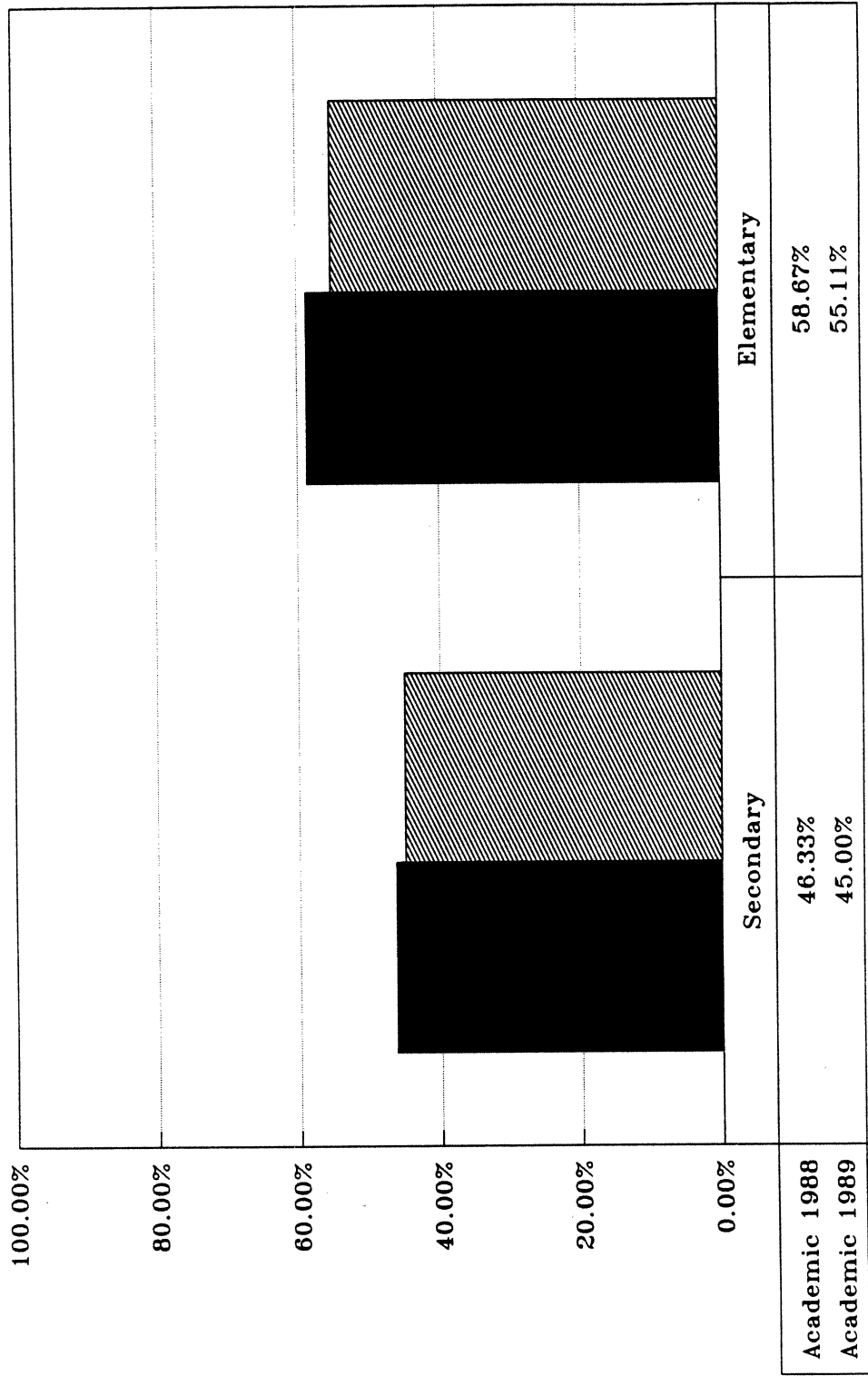


Figure 15. Percent of possible E Q points granted



Academic 1988
 Academic 1989

Figures 16 and 17 compare the teachers' "overall summary performance scores" by school and by elementary or secondary settings. These summary performance scores reflect the sum of the total scores awarded teachers by the appraisers, including the exceptional quality scores. These scores incorporate both the teacher supervisor's and the "other" appraiser's observation data. The resulting score which is placed on an appraisal record for each teacher is the "overall summary performance score." The highest score that a teacher may receive is 184. These performance scores determine the need for the development of an improvement plan for a teacher and placement on the career ladder. The five TTAS ratings with their corresponding "overall summary performance scores" are clearly outstanding (160.0-184.0), exceeds expectations (136.0-159.9), meets expectations (104.0-135.9), below expectations (80.0-103.9), and unsatisfactory (0.0-79.0)

Research Question Three

Research question 3 submits, "Considering the exceptional quality scores between and within the two-year period, what is the range from least to most positive and the percentage of maximum exceptional quality points awarded teachers among the appraisers?" The purpose or commitment of the district was to have appraisers score more consistently and fairly in 1988-89 than in 1987-88. Figure 18 illustrates the percentage and range of exceptional quality

Figure 16. Comparison of performance scores by school

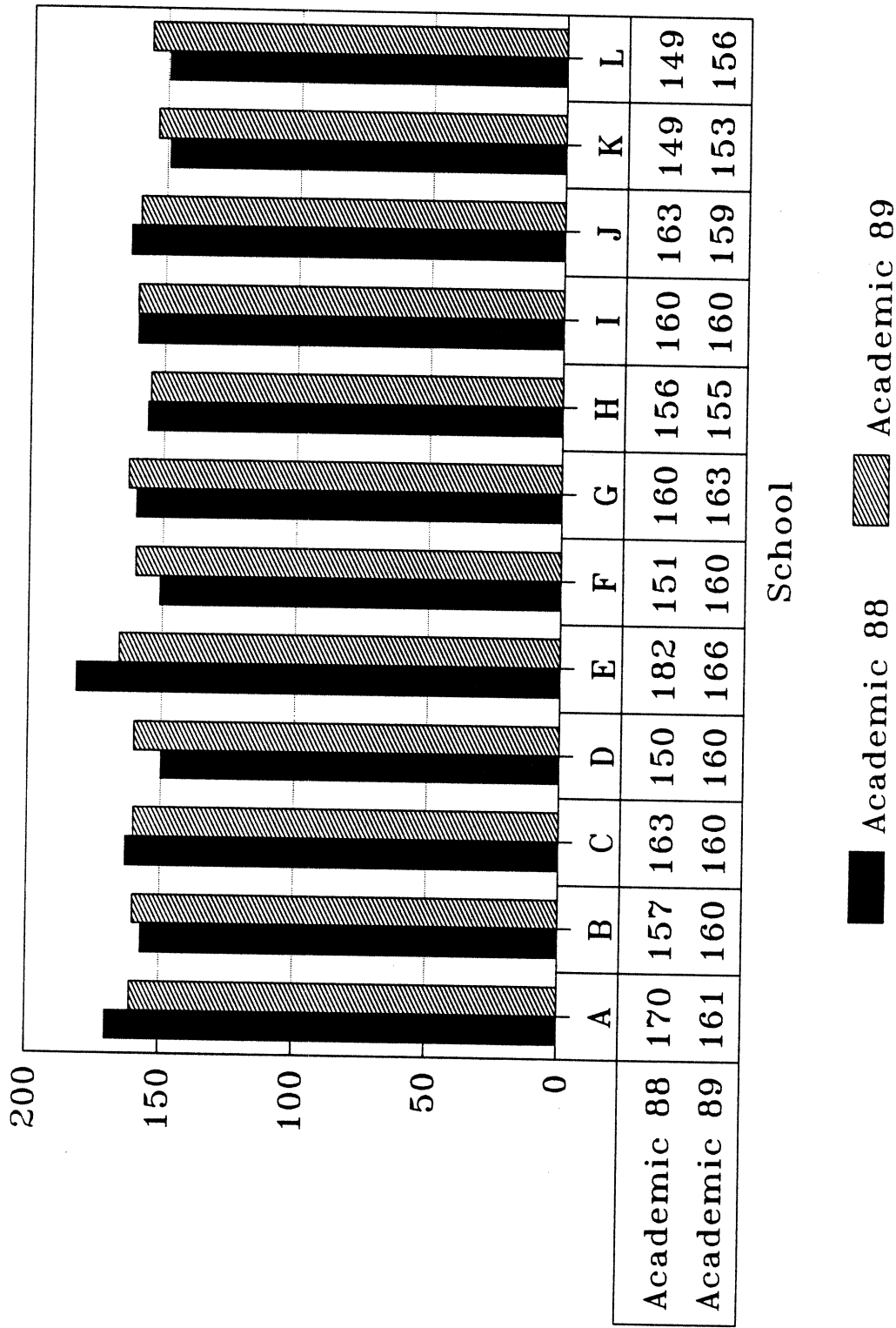
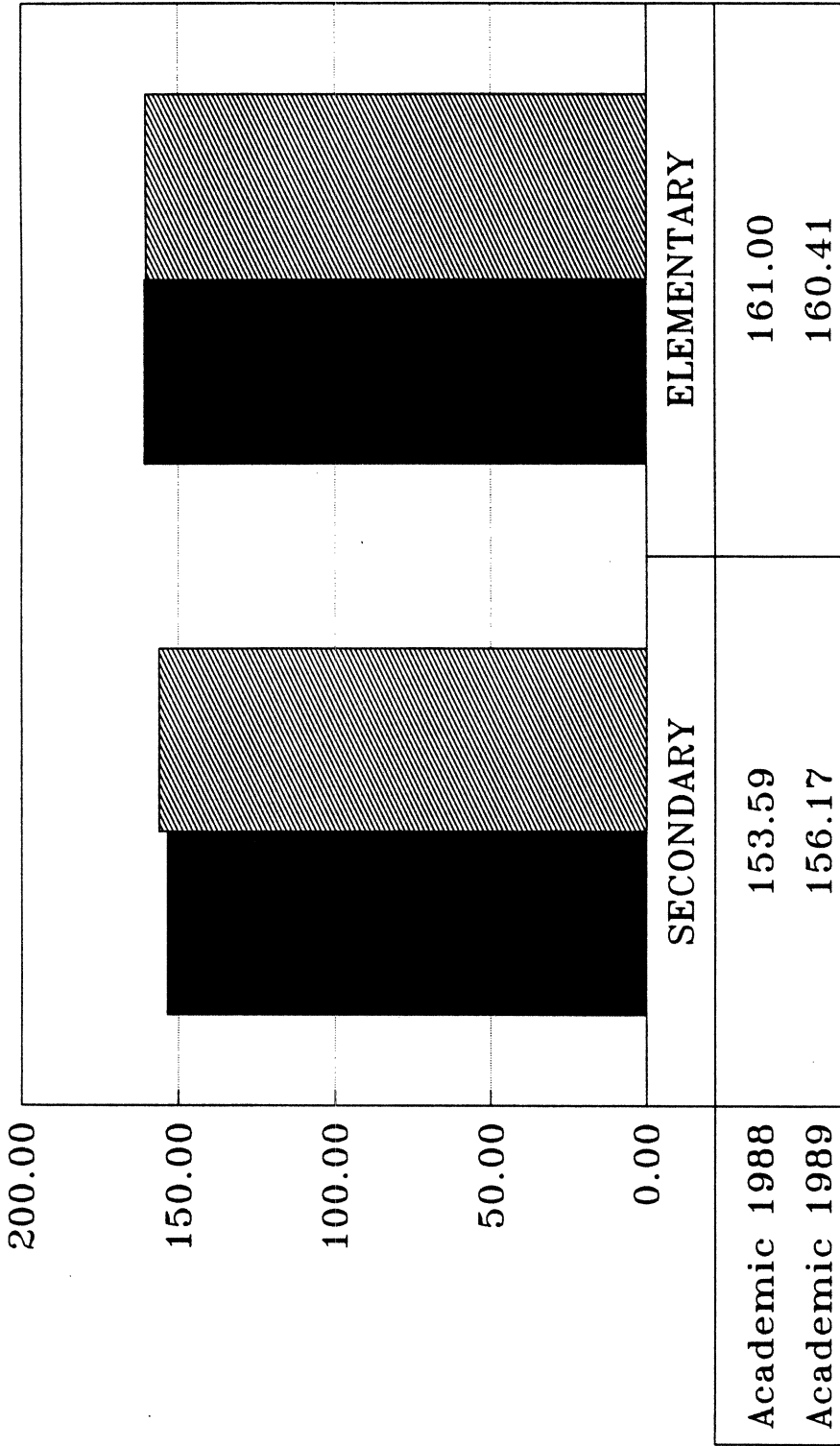
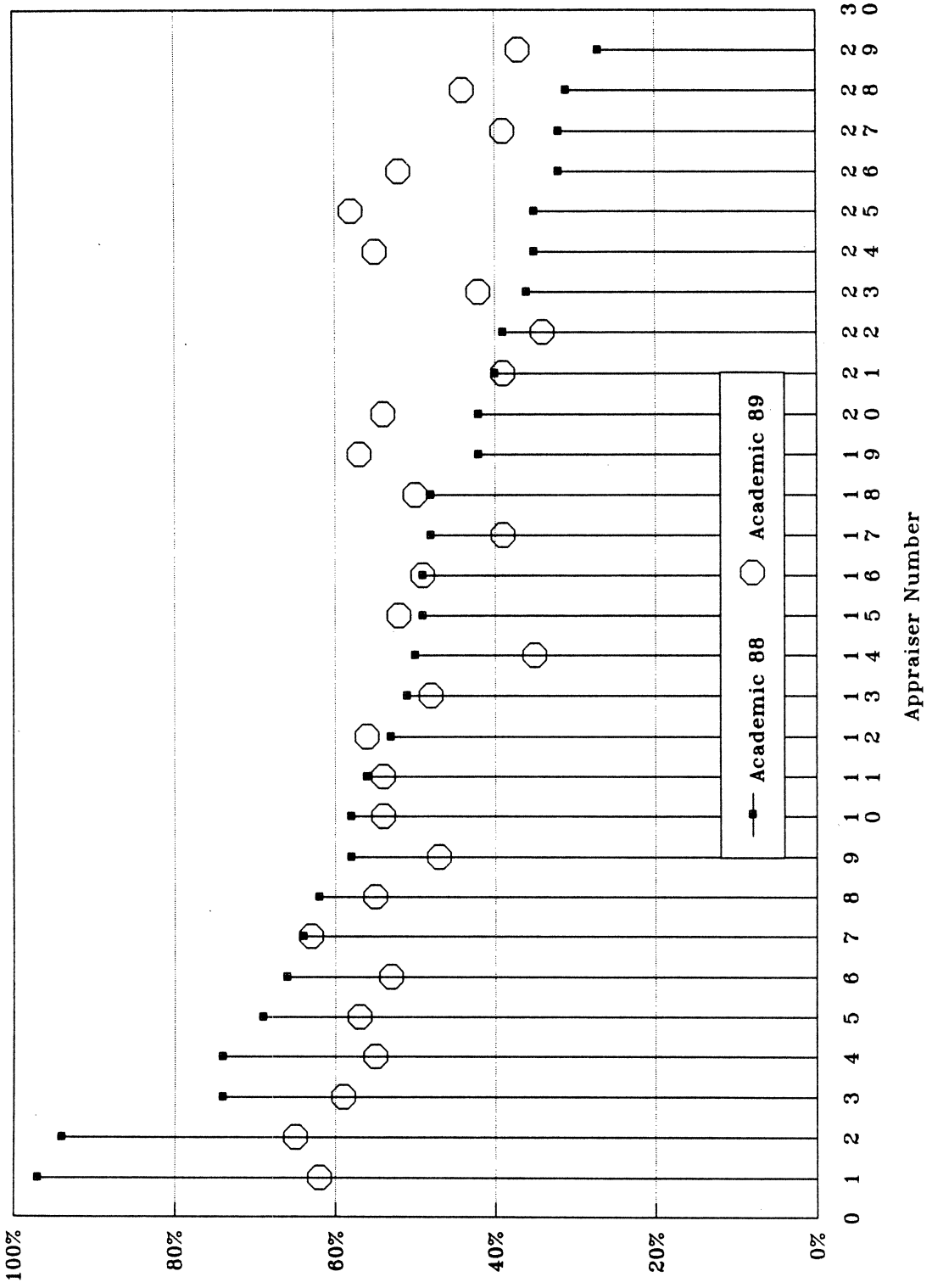


Figure 17. Comparison of performance score by grade level



 Academic 1988
  Academic 1989

Figure 18. Percent of possible E Q by appraiser



points awarded teachers by appraiser. The scores indicate a more consistent pattern among the 29 appraisers in 1988-89 than in 1987-88. The range of exceptional quality points in 1987-88 was from 97, the most positive, to 27, the least positive for a 70 point differential. The range of exceptional points in 1988-89 was from 74 to 34 for a 40 point differential. There was a reduction of 30 points in the high/low differential.

Figure 14 notes the range and the percentage of exceptional quality scores by school. The graphs of the schools illustrate a more consistent scoring in 1988-89. The range was 74 to 34 in 1987-88 and 57 to 44 in 1988-89.

Research Question Four

Research question 4 states, "Did the range and percentage of possible exceptional quality scores granted to teachers change from 1987-88 to 1988-89? Are the changes reflected in some, most, or all of the appraisers considered in the study group?" Table 2 ranks and compares the appraisers each year from the least to the most positive in awarding exceptional quality points. Seventeen of the 19 appraisers changed their rank with each other year to year by three or more positions. For example appraiser 4 ranked in position 4 in 1987-88; however, this appraiser ranked in position 9 in 1988-89. For the purposes of this study, the author considered the range of 40 to 60 percent of possible points was designated as the acceptable or consistent scor-

Table 2

Comparison of Appraiser Ranks in Awarding Exceptional
Quality Points from Most to Least Positive

Appraiser Number	Appraiser Rank 1987-1988	Appraiser Rank 1988-1989
1	1	3
2	2	1
3	3	4
4	4	9
5	5	6
6	6	15
7	7	2
8	8	10
9	9	21
10	10	12
11	11	13
12	12	8
13	13	20
14	14	28
15	15	16
16	16	19
17	17	24
18	18	18
19	19	7
20	20	14
21	21	25
22	22	29
23	23	23
24	24	11
25	25	5
26	26	17
27	27	26
28	28	22
29	29	27

ing range. Table 3 indicates that fewer appraisers scored outside this acceptable range in 1988-89 than in the previous year. While 16 scored out of the acceptable range in 1987-88, this was reduced to 9 in 1988-89.

Research Question Five

Question 5 asks, "Because of additional local training in 1988-89, did appraisers at extreme ends of the range of exceptional quality scores in 1987-88 evaluate consistently with the other appraisers in 1988-89?" The district determined 40 to 60 percent was the acceptable or consistent range for the percentage of possible exceptional quality scores granted. Scores outside of this range were considered extreme or inconsistent in this study; therefore, 15 of the 29 appraisers scored outside of this designated range in 1987-88 (Table 3). These were the inconsistent or extreme appraisers in 1987-88. Nine of these appraisers were elementary while seven were secondary.

In 1988-89 only 9 of the 29 appraisers scored inconsistently from the others. Three were elementary while six were secondary appraisers. Two of the three elementary appraisers were from School E while five of the six secondary appraisers were from Schools K and L. All but two of the secondary appraisers scored more positively in the second year. Table 2 notes that there were no elementary appraisers in the least positive scoring positions the second year and no secondary appraisers in the highly posi-

Table 3

Appraisers With Extreme Scoring Patterns

1987-88 Rank	EQ Freq.	Grade Level	School	Work Experience	1988-89 Rank	EQ Freq.	EQ Change
With EQ Frequency Above 60							
1	97	Elem	E	10	3	62	-35
2	94	Elem	E	19	1	65	-29
3	74	Elem	A	30			
4	74	Elem	A	12			
5	69	Elem	I	24			
6	66	Elem	G	19			
7	64	Elem	C	14	2	63	-1
8	62	Sec	J	14			
With EQ Frequency Below 40							
14	50	Sec	L	28	28	35	-15
17	48	Sec	J	25	24	39	-9
21	40	Sec	K	25	25	39	-1
22	39	Sec	K	17	29	34	-5
23	36	Sec	L	10			
24	35	Sec	L	17			
25	35	Elem	D	7			
26	32	Elem	D	32			
27	32	Sec	K	21	26	39	7
28	31	Sec	K	23			
29	27	Sec	L	20	27	34	7

tive scoring positions. Of the nine out of range evaluators in 1988-89, only six were so in 1987-88. Three secondary appraisers reduced their frequency of award in 1988-89 and fell out of the acceptable range.

Research Question Six

Research question six states, "Do data from buildings with a high percentage of teachers with masters' degrees indicate the greater frequency of exceptional quality scores?" The five schools with the greatest percentage of teachers with masters degrees in 1988-89 (Figure 19 and Figure 20) were L, E, F, A, and C. The five schools with the lowest percentages of teachers with masters degrees were G, I, K, H, and B and J with the same score. The five schools with the highest percentages of exceptional quality points in 1988-89 were E, A, I, C, and D. The five schools with the lowest percentages of exceptional quality points were K, L, H, J, and B. Three of these sets of schools with the highest percentages of exceptional quality and the highest percentages of teachers with masters degrees were the same both years. Three of the sets of schools with the lowest percentages were the same both years.

The mean of exceptional quality frequency of the 12 schools is 52.6. Their mean of masters degrees held is 41.3. Seven schools were above the mean of exceptional quality. Five of the seven had percentages of masters degrees above the district mean. Five of the schools had

Figure 19. Percentage of teachers with masters degrees

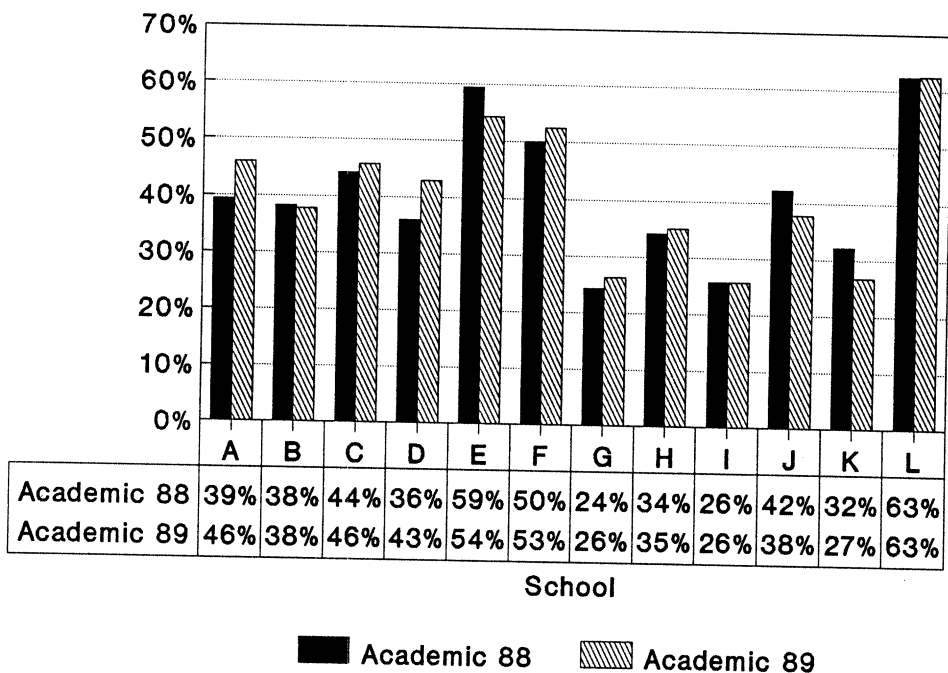
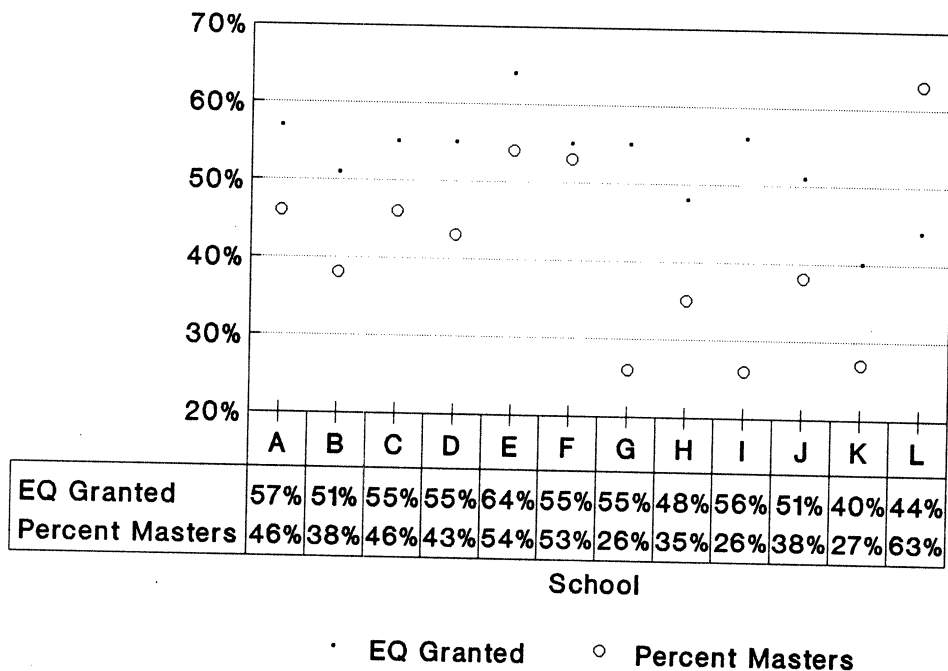


Figure 20. Frequency of E Q compared to % of masters held



percentages below the mean of exceptional quality. Four of these five had percentages of teachers with masters degrees below the district mean.

Other Data Findings

The educational work experience which includes the teaching and administrative years of the appraisers was examined in Table 4 to determine if any relationship between

Table 4

Educational Work Experience of Appraisers in 1988-89

Appraiser Number	Years of Experience
1	11
2	20
3	31
4	13
5	25
6	20
7	15
8	15
9	22
10	6
11	14
12	15
13	14
14	19
15	19
16	24
17	25
18	18
19	12
20	15
21	12
22	18
23	11
24	18
25	8
26	33
27	22
28	24
29	21

experience and scoring could be noted. Table 4 lists the 29 appraisers by their years of experience in teaching and administration. The mean of the total years of experience among the appraisers was 17.9 years in 1988-89. The mean elementary experience was 18.3 years. In Table 4 the mean secondary experience was 17.4 years. The average for the five most positive appraisers in 1988-89 was 19.4 years. The mean for the five least positive appraisers in 1988-89 was 24.6 years. The least and the most positive appraisers had greater educational experience than the district mean.

CHAPTER 5

SUMMARY OF FINDINGS, IMPLICATIONS AND RECOMMENDATIONS

This study considered the appraisal patterns of 29 district appraisers over a two-year period and a single district's efforts to assist its appraisers to evaluate teachers more consistently. The study compared the frequency and range of the exceptional quality scores and the overall performance scores assigned to teachers evaluated by a single school district's appraisers in 1987-88 and 1988-89. It also compared the most and the least positive appraisers of this group each year by individual schools and grade levels. The study considered the effectiveness of additional local training on the appraisers' consistency in evaluating teacher performance.

Summary of Findings

The findings of study are as follows.

1. The total spread or range among the 29 appraisers' exceptional quality and performance scores assigned to teachers was less in 1988-89 than 1987-88. Among the 29 appraisers the range of awarded possible exceptional quality points was 27% to 97% the first year and 37% to 65% the second year of the study. The mean overall performance

scores by school ranged from 149 to 182 the first year and 153 to 166 the second year of the study.

2. There was a difference between the elementary appraisers and secondary appraisers in scoring for both years. The elementary appraisers were more positive than the secondary ones in awarding exceptional quality points and the overall performance scores.

3. Although most of the appraisers scored similarly both years, the elementary appraisers scored less positively the second year of this study in overall performance scores and in awarding exceptional quality points. While the secondary appraisers scored the overall performance scores higher, the awarding of exceptional quality points was less positive than the first year. In 1987-88 the 29 district appraisers awarded 51.2% of the possible exceptional quality points; however, in 1988-89, they awarded 50.7% of these possible exceptional quality points (See Table 1).

4. Some appraisers evaluated extremely differently from their fellow appraisers for both years of the study. In 1987-88, 16 appraisers scored outside the acceptable range. In 1988-89, 9 appraisers scored outside this range.

5. In five of the 7 school with extreme appraisal scores the percentage of masters degreed teachers was above the district mean. Four of the 5 schools with the least positive exceptional quality scores had percentages of teachers with masters degrees below the district mean.

6. The educational experience of appraisers indicated that the least and the most positive appraisers had greater educational experience than the district mean.

7. Field teams of elementary and secondary appraisers could provide much information to a local district. The district should organize the format of the information needed before the schedule for observation begins. Appraisers must follow this schedule. If the schedule can not be followed, the appraiser should write an explanation of the absence and send it to the superintendent. Districts should require that all discussions and debriefings be mandatory for all team observers.

8. Once all district appraisers understand these procedures and district requirements, they will adhere to them more diligently. The district can utilize the data from these appraisal team placements to analyze and to follow the scoring records of each appraiser as compared to team members and future teams throughout the district apart from their assigned school teams.

9. Teachers are working without a clear picture of what exemplifies exceptional quality. The student reactions to these teacher behaviors provide the keys to an appraiser awarding these exceptional quality points. Considering the importance of student response, should not both appraisers of a teacher observe the same type of class? When one appraiser observes an honors class and the other observes a

basic class, will the student reaction be the same? Should the appraisers know the students as well as the teachers they appraise? Should future appraisal training focus on student reactions at different grade-levels?

10. The particular local efforts of this district involved several areas of related research. Although this district concluded these were beneficial to its local needs, other combinations of research items mentioned in Chapter 2 could assist other districts. First, the district immersed the individual appraiser in the appraisal system and provided data for each appraiser to develop an individual plan for greater consistency among district appraisers. Second, the group training enhanced similar common ground and organizational devices among all the appraisers. The discussions of previous sets of ratings and the observation of the same and different teaching segments focused the appraisers on descriptive instead of evaluative behaviors. The district used the staff development and viewing of the videotapes for teachers as well as the appraisers in explaining the concept of student reactions in assessing exceptional quality points. This district attempted to balance the actual with the videotaped observation of district teachers to establish similar images of exceptional quality.

11. Group training and appraisal team field practice provided much close personal contact among the practicing appraisers of the district which Pritchett (1986) argued

should improve consistency in scoring. Despite this training, an assistant superintendent agreed with Bulzer (1984) that personality differences and the uniqueness of human raters caused some of the variance among the district's appraisers.

12. Each year the appraisers reviewed the past year's results and experienced refresher training as a group. This focus according to research assisted the appraisers to begin each new appraisal year with a clearer purpose for observations and evaluations.

13. The district did not subject the extreme scoring appraisers, although identified, to retraining. However, all extreme appraisers' scores moved closer to the mean the second year of the study. Since each appraiser had a copy of the range of scores by criterion, as instructed by an assistant superintendent, appraisers analyzed their own scores by items on the TTAS instrument criterion and noted those areas in need of greater consistency. Each appraiser and school developed individual improvement plans or means to become more consistent.

14. The district's commitment to hold conferences with all teachers assisted in providing the positive atmosphere necessary for the perception of fairness in the appraisal system. The use of district teachers as the focus for observation and evaluation helped define the district's image of exceptional teacher behaviors. The district's

voiced expectation that appraisers become more and more consistent illustrates a commitment to an appraisal system with high standards for both teachers and appraisers.

15. If Popham's (1985) conclusions that appraisers are affected by their own teaching and appraising are correct, then this district will be shaping the style of its future appraisers since most of its administrators are hired from within the district. A teacher's perception of the appraisal system could predict that teacher's future appraisal style.

Implications and Suggestions for Further Research

Based on the findings of this study the following implications may be drawn.

1. Without training by school districts for appraisers, in addition to the training mandated by the state, it should be anticipated that wide variations in scores may occur and that consistency in scores may not occur.

2. Efforts by school districts to provide training to appraisers in order to improve the consistency of appraisals may reduce the range of appraisal scores given by appraisers and in general may lower the scores given by the most "positive" appraisers and increase the scores given by the least "positive" appraisers. The average score for the districts may remain about the same with or without additional training for appraisers.

3. The average of scores given by appraisers to elementary teachers may be higher than those to secondary teachers and scores given to high school teachers may be higher than those given to middle/junior high school teachers.

4. On the average teachers with advanced degrees may earn higher appraisal scores than will teachers without advanced degrees.

5. Appraisers who are more experienced may produce more consistent appraisals over time than will less experienced appraisers. However, the more experienced appraisers may tend to produce appraisals which are more deviant (higher or lower) from the district mean than will less experienced appraisers. In other words, appraisers with less experience tend to produce appraisal scores with wide fluctuations in their range of scores, but they also are more likely to have average scores which are at the mean of district-wide scores.

Recommendations

School districts should require retraining of appraisers when they score at the extreme ranks each year in relation to the other district appraisers. Through training these appraisers may develop individual plans for narrowing the extreme range of their scores for upcoming years. A district should require that retraining include all appraisers who score outside of the established accepted range such

as 40 to 60 percentage of possible exceptional quality points. Some appraisers should appraise in teams on different campuses, or they should appraise only after additional training results indicate that their evaluations have a more consistent scoring pattern. Possibly most positive and least positive scoring appraisers could be teamed together.

This study offers a simple system of appraiser analysis for a district to implement since districts have access to their appraisers' scoring data. Medium to larger school districts could analyze their appraisal data using computer technology. The needs and resources of each district determine the level of sophistication of these analyses. Even a simple descriptive statistic such as an ordinal rank each year presents many avenues for research and staff development, while even more elaborate analyses could generate individual item agreement or more complex rater comparisons. Locally developed questionnaires on problems with observing teachers or implementing TTAS instrument could help these extreme scoring appraisers.

Districts should consider adjusting the actual score to some adjusted mean score to allow for appraiser differences. Possibly some district administrators should not function as appraisers. Some job descriptions may need to be changed to allow those in the central office to perform as appraisers.

Some appraisers may need to be assigned to a different campus.

A district could demonstrate a commitment to more consistent appraisal scoring by adding the appraisers' consistency in evaluating teachers as a basic component of their own administrative performance evaluations. Appraisers who consistently score in the extremes of the district rankings could receive fewer evaluative points within their own appraisal system than appraisers who have more consistent rankings. While teachers' career ladder pay reflects teaching behaviors, a district could tie administrators' salaries in part to their role as consistent appraisers.

Each local district should review the problems, issues, and concerns of its appraisers each year. Local districts should pursue methods to identify and assist appraisers who implement the TTAS instrument inconsistently. There needs to be more direction from the local district as to what is expected in scoring patterns. For example, a range of acceptable appraisal scores could be identified by the district. Confusion over the purposes and direction of evaluation can cause problems with the time appraisers spend on evaluations.

The state may continue to offer only a half-day update for general state issues; however, each district should analyze its own data to predict the local effects of the TTAS yearly appraisals, especially the financial impact of the career ladder on the local budget. The means to assign

teachers to the higher career ladder steps needs to be identified now. The effectiveness of this evaluation system will determine the future numbers and specific teachers placed on the career ladder.

A district should consider distributing its teachers according to their degrees for the purpose of enhancing the chances for greater consistency among the appraisers and schools. A district should choose to redistribute its more experienced teachers within the district to achieve a better balance of experience in each of its schools. As new schools open within this district, the opportunities to attempt these new distribution patterns arise.

Local school districts should develop an ongoing analysis of the consistency of their teacher appraisers. Their insistence upon consistency among their appraisers will enhance the fairness of their evaluation systems. If districts are faced with a quota system for the teacher career ladder in the future, more consistent evaluations among schools will reinforce the concept of the district's fairness in evaluation and the acceptance by its teachers of its decisions for or against their career ladder placement.

Furthermore, local districts should develop an ongoing training for local appraisers to instill district goals for the importance of consistency and fairness. Information and research concerning steps to determine and to improve interrater reliability and consistency suggest that the local

district take a decisive role in the instruction of its appraisers.

Involving the higher career ladder teachers as teaching models as well as instructional leaders for the administrative appraisal training could further develop a mutual commitment between teachers and administrators for the instructional purposes of evaluation. This instructional preparation of the appraisers with the help of teachers could affect both the teachers and the appraisers positively. Districts should videotape examples of what model district's standards of exceptional quality for teachers on the TTAS instrument.

District staff development programs, local retraining efforts, and central office staff support may assist in fulfilling the responsibility of school districts to provide a positive atmosphere for evaluation. Then as the district's appraisal practices are added to the velcro strip of educational reform within each classroom, the chances of them clinging will vastly be improved.

Appendix A
INDEPENDENT SCHOOL DISTRICT
GUIDELINES FOR INTERPRETING
EXCEPTIONAL QUALITY PERFORMANCE

FOREWARD

In October, 1986, a committee was appointed at the direction of the Superintendent of Schools to discuss and analyze the concept of exceptional quality as it relates to the Texas Teacher Appraisal System and to develop guidelines which more specifically define the parameters of the EQ rating. The committee, composed of three classroom teachers, four appraisers, and a central office administrator, included an equal representation of elementary and secondary personnel. This document represents the culmination of a task requiring numerous hours of intense thought and discussion. It was the intent of the committee to establish, within the confines of the instrument, a standard for excellence which would be both reasonable and obtainable for the instructional staff as well as flexible to meet the diversity of circumstances under which it would be applied.

Committee representatives considered each of the performance indicators for exceptional quality in terms of three qualifiers: preponderance, quality, and variety. Critical attributes for these indicators were developed, where appropriate, on the basis of the strength or weight of positive evidence, the quality or uniqueness of the effort, and the variety or number of different ways a teacher might perform a particular behavior. Not all qualifiers were found to be applicable in every case. At the same time, the committee found it difficult to distinguish between the standard expectation and exceptional quality for every indicator.

As you review these guidelines, please note that the critical attributes have been expressed in a question format...questions which will be asked each time an appraiser considers the rating for an indicator. Most exceptional quality attributes include two or more questions separated by an AND or an OR. When an affirmative AND appears between two attributes, both questions must be answered in the affirmative to award credit. When an OR appears, a positive response for either or both will qualify for credit. For critical attributes where no distinction is made between standard expectation and exceptional quality, the phrase, "If SE, EQ earned", is used to so note that EQ credit is to be awarded.

This document was developed as a supplement to the official appraisal instrument and is not intended to serve as the sole basis for the interpretation of indicators. When used in concert with the official instrument, we believe that the process will be greatly enhanced and that all parties to the system will benefit from a higher degree of accuracy and consistency.

NOTE: During the summer of 1987, the SBOE made several minor modifications to the TTAS instrument. These guidelines have been edited to reflect those changes with no significant revision to the content or intent of the original document.

 DOMAIN 1. INSTRUCTIONAL STRATEGIES

Criterion 1. provides opportunities for students to participate actively and successfully

a. appropriately varies activities

SE - Did the teacher vary instruction/lesson activities for the purpose of enhancing student participation?

EQ - Did the teacher maintain a high level of student interest/involvement/success throughout the lesson?

b. interacts with students in group formats as appropriate

SE - Did the grouping format(s) provide opportunities for the teacher to interact with student(s) effectively?

EQ - Did the teacher interact frequently and consistently with students?
 AND
 Did the interaction positively affect the student engagement?
 OR
 Did the teacher orchestrate a variety of groups successfully?

c. solicits student participation

SE - Did the teacher frequently pursue individual student contributions appropriate to the lesson?

EQ - Did the teacher provide an opportunity for a majority of the students to contribute?
 OR
 Did the teacher solicit student contributions with a variety of techniques?

d. extends students' responses/contributions

SE - Did the teacher take advantage of more than one opportunity to extend?

AND

Were extensions related to lesson objectives?

EQ - Did the teacher take advantage of a majority of the opportunities to extend?

AND

Were extensions beyond the comprehension level? (i.e., Bloom's Taxonomy: higher level thinking skills)

e. provides ample time for students to respond to teacher questions/solicitations and to consider content as it is presented

SE - Did the teacher allow adequate wait time for students to think and respond?

EQ - Did the teacher consistently allow for adequate wait time?

AND

Did the teacher vary wait time according to the abilities of the students and the difficulty of the questions?

f. implements instruction at an appropriate level of difficulty

SE - Did the teacher present the lesson at an appropriate level of difficulty?

EQ - As the lesson progressed, did the teacher monitor and adjust the level of difficulty as needed?

Criterion 2. evaluates and provides feedback on student progress during instruction

a. communicates learning expectations

SE - Did the teacher make a clear statement to students regarding what was to be learned?

EQ - Did the teacher refocus students on intended learning outcomes?

b. monitors students' performance as they engage in learning activities

SE - Did the teacher use an appropriate technique to check for student understanding?

EQ - Did the teacher use the inputs gathered through MONITORING to adjust instruction?

OR

Did the teacher check student progress with a variety of techniques?

c. solicits responses or demonstrations from specific students for assessment purposes

SE - Did the teacher assess student progress by individually questioning students or asking them to demonstrate a skill or process?

EQ - Did the teacher use the input gathered through ASSESSING to clarify or reteach?

OR

Did the teacher use a variety of techniques to assess individual student progress?

d. reinforces correct responses/performances

SE - Did the teacher recognize and reinforce correct responses/performances?

EQ - Did the teacher use appropriate opportunities to reinforce individual responses by telling the student WHY THE RESPONSE OR PERFORMANCE WAS CORRECT?

e. provides corrective feedback/clarifies, or none needed

SE - Did the teacher provide corrective feedback/clarification for incorrect responses, if needed?

EQ - Did the teacher PLAN the lesson to allow opportunities for guided practice, corrective feedback, or clarification prior to independent practice or lesson progression?

f. reteached, or none needed

SE - Did the teacher reteach when necessary?

EQ - No reteaching was necessary
OR

If reteaching were necessary,
did the teacher reteach using
an alternative method(s) of
instruction?

 DOMAIN II. CLASSROOM MANAGEMENT AND ORGANIZATION

Criterion 3. organizes materials and students

<u>Performance Indicators</u>	<u>Critical Attributes</u>
a. secures student attention, or students are attending	SE - Did the teacher secure a majority of students' attention when giving directions? EQ - Did the teacher secure and maintain a majority of the students' attention throughout the lesson/activities?
b. uses administrative procedures and routines which facilitate instruction	SE - Did the teacher manage routine administrative procedures efficiently? EQ - Was there evidence that the teacher had planned and organized routine administrative procedures to develop student responsibility?
c. gives clear administrative directions for classroom procedures or routines, or none needed	SE - Did the teacher give clear, concise administrative directions? OR Was there evidence to suggest that clear, concise administrative directions had been given? AND Did Students understand and respond appropriately to administrative directions? EQ - Was there evidence to suggest that classroom procedures or routines were understood throughout the lesson?

- d. maintains seating arrangement/grouping appropriate for the activity and the environment
- SE - Did the teacher arrange the room/area such that students could easily focus on the instruction without undue distraction?
- EQ - Did the room arrangement contribute significantly to keeping students on task?
OR
Were seating arrangements/groupings changed to more effectively accommodate various lesson activities?
- e. has materials, aides, and facilities ready for use
- SE - Did the teacher have materials, equipment, and facilities ready for use prior to the beginning of the lesson?
- EQ - Did the teacher effectively utilize the time resulting from prior planning/-preparation?

Criterion 4. maximizes amount of time available for instruction

- a. Begins promptly/avoids wasting time at the end of the instructional period
- SE - Did the teacher begin the instruction/activity promptly?
AND
Did the instruction or other planned activities extend to the end of the period?
- b. implements appropriate sequence of activities
- SE - Did the teacher sequence lesson activities in such a way that students had the necessary background and information to follow instructions or complete assignments?
- c. maintains appropriate pace
- SE - Did the teacher appropriately pace instruction according to the level of difficulty and range of student abilities?

d. maintains focus

SE - Did the teacher remain focused on the lesson objectives?
 AND
 Did the teacher maximize the instructional time by avoiding teacher initiated delays, digressions, or lengthy transitions?

e. keeps students engaged

SE - Did the teacher keep most of the students engaged during the instructional time?

Criterion 5. manages student behavior

a. specifies expectations for class behavior, or none needed

SE - Did the students' behavior indicate that the teacher's expectations for behavior had been made clear?

OR

If inappropriate behavior occurred during instruction, did the teacher restate or clarify expectations?

b. uses techniques to prevent off-task behavior, or none needed

SE - Did the teacher demonstrate specific strategies for preventing off-task behavior? (passive/alone)

c. uses techniques to redirect/stop inappropriate/disruptive behavior, or none needed

SE - Did the teacher accurately identify and promptly redirect/stop inappropriate/disruptive behavior, if needed? (active/may effect others)

d. applies rules consistently and fairly, or none needed

SE - Did the teacher apply class rules fairly and consistently, when needed?

EQ - Were class rules effective and age and learner appropriate?

e. reinforces appropriate behavior

SE - When and if inappropriate behavior occurred, did the teacher reinforce desired behavior with specific praise to the individual or group? (reinforcement may be non-verbal)

EQ - Did the teacher reinforce desired behavior, even if there was no evidence of inappropriate behavior?

 DOMAIN III. PRESENTATION OF SUBJECT MATTER

Criterion 6. teaches for cognitive, affective, and/or psychomotor learning and transfer

<u>Performance Indicators</u>	<u>Critical Attributes</u>
a. begins instruction/activity with an appropriate introduction	SE - Did the lesson introduction focus student attention on the content/purpose of the lesson? EQ - Did the lesson introduction CAPTURE student interest and focus attention on the content/purpose of the lesson?
b. presents information in an appropriate sequence	SE - Did the teacher present the lesson content in an appropriate sequence? EQ - If SE, EQ is earned.
c. relates content to prior or future learning	SE - Did the teacher relate the content to prior or future learning? EQ - Did the teacher CONSISTENTLY take advantage of opportunities to relate content to prior OR future learning? OR Did the teacher relate in a particularly meaningful way the content to prior or future learning?
d. provides for definition of concepts and description of skills and/or attitudes and interests	SE - Did the teacher provide sufficient DEFINITION(S) for content appropriate to age and ability of students? EQ - Did the teacher or student(s) expand definitions with concise details to enhance student understanding? OR Did the teacher use a variety of techniques to define new concepts, skills, or attitudes and interests?

- e. provides for elaboration of critical attributes of concepts, skills, and/or attitudes and interests
- SE - Did the teacher or students elaborate the KEY POINTS of the lesson
- EQ - Did the teacher use a variety of strategies to elaborate the key points of the lesson?
- f. stresses generalization, principle, or rule as a relationship between or among concepts, skills, or attitudes/interests
- SE - Did the teacher stress the rule(s), principle(s), or generalizations(s) to enhance student comprehension and retention of lesson content?
- EQ - Did the teacher encourage students to attempt to analyze synthesize, or evaluate rule(s), principle(s), or generalization(s) from the content?
- g. provides opportunities for application
- SE - Did the teacher provide an opportunity(ies) for students to apply skills or concepts in activities requiring thinking skills beyond the comprehension level? (Credit may be given even though application activities were not completed during the observed lesson.)
- EQ - If SE, EQ is earned.
- h. closes instruction appropriately
- SE - Did the teacher close the lesson with a brief summary of the key points presented?
- EQ - Did the teacher summarize at the conclusion of EACH lesson segment?
- OR
- Did the teacher encourage students to participate in summarizing the lesson or lesson segments?

Criterion 7. uses effective communication skills

- a. makes no significant errors
- SE - No teacher errors occurred which interfered with student learning.
- OR
- If errors occurred, did the teacher recognize and correct them and clarify the content?
- b. explains content and/or learning tasks clearly
- SE - Did the teacher explain the content and learning tasks so that they were understood by the majority of the students?
- EQ - Did the teacher explain or demonstrate the steps to be followed, provide examples, and identify potential areas of difficulty?
- c. stresses important points and dimensions of content
- SE - Did the teacher stress or emphasize important points in the lesson?
- EQ - Did the teacher use several strategies to stress or demonstrate important points?
- OR
- Did the teacher use a particularly effective/unique method to emphasize important points(s)?
- d. uses correct grammar
- SE - Did the teacher use correct grammar?
- OR
- No more than one significant grammatical error was noted which might affect student learning.
- e. uses accurate language
- SE - Did the teacher avoid excessive use of false starts, interrupters, qualifiers, or distractors?
- f. demonstrates skill in written communication
- SE - Did the teacher demonstrate skill in written communication by having no more than one error in spelling, grammar, or sentence structure?

 DOMAIN IV. LEARNING ENVIRONMENT

Criterion 8. uses strategies to motivate students for learning

Performance Indicators

Critical Attributes

a. relates content to student interests/experiences

SE - Did the teacher deliberately relate content to students' interests and/or experiences?

EQ - Did the teacher more effectively relate the content to student interest(s) by involving the students and using a variety of materials and/or methods in the presentation?

b. emphasizes the value/importance of the activity or content

SE - Did the teacher make a general statement regarding the value or importance of the activity or content?

EQ - Did the teacher stress the value or importance of the activity or content and relate it to society, the lesson objective, or to the students personally?

c. reinforces/praises learning efforts of students

SE - Did the teacher take advantage of appropriate opportunities to acknowledge and encourage learning efforts or progress?

EQ - Did the teacher use a variety of methods to communicate an awareness of and appreciation for students' efforts?

d. challenges students

SE - Did the teacher challenge students by communicating learning expectations that may require extra effort?

EQ - Did the teacher monitor and provide guidance to enhance opportunities for students' success?

AND

Did the teacher assure students that they are capable of meeting the learning expectations?

Criterion 9. maintains supportive environment

- a. Avoids sarcasm and negative criticism
- b. establishes climate of courtesy and respect

SE - Did the teacher avoid sarcasm and negative criticism?

SE - Was there evidence that the teacher had established a classroom climate of courtesy and respect?

- c. encourages slow and reluctant students

EQ - Did the teacher maintain a climate of courtesy and respect throughout the lesson?

AND

Did the teacher model courtesy and respect in interactions with students?

SE - Did the teacher encourage slow and reluctant students?

EQ - When student responses were incorrect, did the teacher dignify student errors?

OR

Did the teacher demonstrate patience and positively reinforce student effort?

- d. establishes and maintains positive rapport with students

SE - Did the teacher relate to students in a pleasant manner and secure cooperation from the class?

EQ - Was there evidence to suggest that there was a high degree of cooperation between students and teacher?

OR

During the instruction, were students secure and willing to risk a response?

APPENDIX B
INDEPENDENT SCHOOL DISTRICT
DISTRICT GUIDELINES
TEXAS TEACHER APPRAISAL SYSTEM
1987-88

The following guidelines for the administration of the Texas Teacher Appraisal System have been approved for the 1987-88 school year.

I. APPRAISAL PERIODS

The two appraisal periods have been designated to coincide with the semester break approved in the 1987-88 calendar:

Appraisal Period #1	September 16, 1987-January 15, 1988
Appraisal Period #2	January 19, 1988-May 14, 1988

Appraisal Period #1 and #2 include 74 and 75 instructional days, respectively, upon which observations may be scheduled. Formal observations may not be conducted during the first two weeks of instruction nor on the last instructional day before any official school holiday. Therefore, the following dates are excluded from the appraisal schedule:

<u>Appraisal Period #1</u>	<u>Appraisal Period #2</u>
September 1-15, 1987	March 11, 1988
November 25, 1987	March 31, 1988
December 18, 1987	

During the first appraisal period, teacher supervisors will convert their observations reports to evaluation reports from December 7, 1987, through January 15, 1988. Second period evaluation reports will be completed April 18-May 13, 1988.

II. DESIGNATED APPRAISERS

Each teacher will be assigned two appraisers, one to be designated as the "teacher supervisor" and the other as "other appraiser." Teachers serving more than one campus will be appraised at their home campus. Building administrators will be assigned accordingly:

<u>Campus Level</u>	<u>Teacher Supervisor</u>	<u>Other Appraiser</u>
Elementary	Building Principal	Assistant Principal
Jr. Hi. School	Building Principal	Assistant Principal
Sr. Hi. School	Building Principal	Assistant Principal
	Associate Principal	Assistant Principal

Appraisal responsibilities at the Sullivan-Keller Instructional Center and other special education facilities will be shared by the interim Assistant Director of Special Education and another certified appraiser to be designated at a later date. Building principals will serve as "teacher

supervisors" for professional personnel employed in the Denton Regional Day School Program (deaf education) with the program coordinator assigned as the "other appraiser."

As a rule, teachers will be assigned to two appraisers only; however, in the event of an unforeseen crisis, an alternate appraiser may be assigned by the Office of Support Services to fulfill appraisal duties in the absence of the appraiser originally appointed.

III. NUMBER OF REQUIRED OBSERVATIONS

HB 173, which was recently passed in a special session of the legislature, provides that teacher appraisal may be conducted only once during each school year for teachers having attained Career Ladder Level II, or higher, and whose most recent appraisal was scored as exceeding expectations or clearly outstanding. One annual appraisal by two appraisers will be required for teachers meeting these criteria. The appraisal may be conducted during either Appraisal Period #1 or #2, with scheduling left to the discretion of the assigned appraisers.

Two appraisals will be required for all other teachers.

IV. SCHEDULING OF OBSERVATIONS

All required observations for the 1987-88 school year will be scheduled with a minimum of three days notice. Notifications must be documented in a written form (i.e., memorandum, note, schedule, etc.) with a specific date and time of day indicated.

An observation may be rescheduled by an appraiser in the event of unforeseen situations which may be of an emergency or critical nature. Should it be necessary to reschedule an observation, a minimum of three days notice will be required unless waived by the teacher.

With the exception of required scheduled observations, appraisers may observe in a classroom on an unscheduled basis at any time, and the results may become a part of the appraisal record so long as the established procedures for documentation and conferencing with the teacher are followed.

V. OTHER CONSIDERATIONS

All formal observations will be at least 45 minutes in duration with the provision of waiver only in extraordinary circumstances which must be documented and approved by the Superintendent of Schools.

Professional personnel other than classroom teachers will continue to be evaluated with the approved instruments used in prior years. Librarians, counselors, and other support personnel who have been advanced to Level II in the district sponsored career ladder programs and who were rated as exceeding expectations or clearly outstanding in the most recent appraisal may be evaluated only once during the school year.

Appendix C
INDEPENDENT SCHOOL DISTRICT
FIELD TEAM PRACTICE SCHEDULE
1988-89

Field Team Practice Schedule 1988-89

Week	Team	Campus	Team Members
September 12-16, 1988	1	C	<u>9</u> , 13, 20, 24, 25
	2	H	1, 7, 4, 8, <u>16</u> , 21
	3	D	2, 10, 18, <u>26</u> , 29
	4	B	11, 12, <u>15</u> , 17
	5	K	4, 5, 14, <u>20</u> , 28
	6	A	<u>3</u> , 6, 19, 22, 23
November 7-11, 1988	1	E	<u>2</u> , 3, 21, 25, 28
	2	L	1, 9, 11, 22, <u>29</u>
	3	K	7, 18, 20, <u>27</u>
	4	G	<u>6</u> , 8, 12, 14, 26
	5	J	4, <u>10</u> , 13, 5, 23
	6	I	<u>5</u> , 6, 17, 19, 24
January 23-27, 1989	1	J	5, <u>8</u> , 11, 25, 29
	2	E	<u>1</u> , 3, 10, 20
	3	B	6, 9, 14, 17, <u>18</u>
	4	K	<u>7</u> , 12, 13, 23, 28
	5	A	<u>4</u> , 16, 22, 24, 26
	6	F	2, 15, <u>19</u> , 21, 27
February 27- March 3, 1989	1	D	14, 15, 20, 22, <u>25</u>
	2	L	1, 5, 6, <u>24</u> , 27
	3	H	3, 8, <u>13</u> , 18, 23
	4	I	9, 10, <u>12</u> , 16, 21
	5	C	2, 4, <u>7</u> , 17, 29
	6	G	<u>11</u> , 19, 26, 28

Team Leader is Underlined

N = Appraiser Number

Appendix D
INDEPENDENT SCHOOL DISTRICT
DISTRICT GUIDELINES
TEXAS TEACHER APPRAISAL SYSTEM
1988-89

INDEPENDENT SCHOOL DISTRICT
DISTRICT GUIDELINES
TEXAS TEACHER APPRAISAL SYSTEM
1988-89

The following guidelines for the administration of the Texas Teacher Appraisal System have been approved for the 1988-89 school year.

I. APPRAISAL PERIODS

Two appraisal periods have been designated and scheduled to coincide with the first and second semesters as approved in the 1988-89 calendar:

Appraisal Period #1	September 15, 1988-January 18, 1989
Appraisal Period #2	January 23, 1989-May 19, 1989

Appraisal Period #1 and #2 include 75 and 76 instructional days, respectively, upon which observations may be scheduled. Formal observations may not be conducted during the first two weeks of instruction nor on the last instructional day before any official school holiday. Therefore, the following dates are excluded from the appraisal schedule:

<u>Appraisal Period #1</u>	<u>Appraisal Period #2</u>
September 1-14, 1988	January 19, 1989
November 22, 1988	March 10, 1989
December 16, 1988	March 23, 1989

During the first appraisal period, teacher supervisors will convert their observation reports to evaluation reports from December 5, 1988, through January 20, 1989. Second period evaluation reports will be completed April 17-May 19, 1989.

II. DESIGNATED APPRAISERS

Each teacher will be assigned two appraisers, one to be designated as the "teacher supervisor" and the other as "other appraiser." Building administrators will be assigned accordingly:

<u>Campus Level</u>	<u>Teacher Supervisor</u>	<u>Other Appraiser</u>
Elementary	Building Principal	Assistant Principal
Jr. Hi. School	Building Principal	Assistant Principals
Sr. Hi. School	Building Principal	Assistant Principals
	Associate Principals	

Teachers assigned half time to two campuses will be ap-

praised on each campus with building principals serving alternately as the "teacher supervisor" and "other appraiser." Appraisal arrangements for itinerant teachers scheduled less than 50% of the time on two or more campuses will be determined on a case by case basis.

Appraisal responsibilities at special education facilities will be shared by the Assistant Director of Special Education and another certified appraiser.

As a rule, teachers will be assigned to two appraisers only; however, in the event of an unforeseen crisis, an alternate appraiser may be assigned by the Office of Support Services to fulfill appraisal duties in the absence of the appraiser originally appointed.

III. NUMBER OF REQUIRED OBSERVATIONS

The Texas Administrative Code, Section 149.43(c)(1) provides that "teachers on levels two, three or four of the career ladder whose performance was exceeding expectations or clearly outstanding on the most recent overall summary performance score shall be appraised at least once each year." One annual appraisal by two appraisers will be required for teachers meeting these criteria. **The appraisal will be uniformly extended throughout both scheduled appraisal periods with one observation completed in the fall and one in the spring.**

Two appraisals will be required for all other teachers.

IV. SCHEDULING OF OBSERVATIONS

All required observations for the 1988-89 school year will be scheduled with a minimum of three days notice. Notification must be documented in written form (i.e., memorandum, note, schedule, etc.) with a specific date and time of day indicated.

An observation may be rescheduled by an appraiser in the event of an unforeseen situation which may be of an emergency or critical nature. Should it be necessary to reschedule an observation, a minimum of three days notice will be required unless waived by the teacher.

With the exception of required scheduled observations, appraisers may observe in a classroom on an unscheduled basis at any time, and the results may become a part of the appraisal record so long as the established procedures for documentation and conferencing with the teacher are followed.

V. OTHER CONSIDERATIONS

All formal observations will be at least 45 minutes in duration with provision of waiver only in extraordinary circumstances which must be documented and approved by the Superintendent.

Professional personnel other than classroom teachers will continue to be evaluated once each year with locally developed and approved instruments.

REFERENCES

- Atkinson, D., & Murray, M. (1987, March). Improving interrater reliability. Paper presented at the annual meeting of the Conference on College Composition and Communication, Atlanta, GA.
- Alfonso, R.J., Firth, G., & Neville, R. (1984). The supervisory skill mix. Educational Leadership, 41(7), 16-18.
- Andrews, H. A. (1985). Evaluating for Excellence. Stillwater, OK: New Forum Press.
- Association of Teacher Educators. Commission on Teacher Assessment. (1988). Teacher Assessment. Reston, VA: Author.
- Banks, C. G., & Murphy, K. R. (1985). Toward narrowing the research-practice gap in performance appraisal. Personnel Psychology, 38, 335-345.
- Barnes, S. (1987, April). The development of the Texas teacher appraisal system. Paper presented at the annual meeting of the American Research and Education Association, Washington, DC.
- Becker, B. E., & Cardy, R. I. (1986). Influence of halo error on appraisal effectiveness: A conceptual and empirical reconsideration. Journal of Applied Psychology, 71(4), 662-671.
- Bernadin, H. J. (1978). Effects of rater training on leniency and halo effects in student ratings of instructions. Journal of Applied Psychology, 63(3), 301-308.
- Bernadin, H. J., & Pence, E. C. (1980). Effects of rater training: Creating new response sets and decreasing accuracy. Journal of Applied Psychology, 65(1), 60-66.
- Bernadin, H. J., & Walter, C. S. (1977). Effects of rater training and diary-keeping on psychometric error in ratings. Journal of Applied Psychology, 62(1), 64-69.
- Borman, W. C. (1975). Effects of instructions to avoid halo error on reliability and validity of performance evaluation ratings. Journal of Applied Psychology, 60(5), 556-560.
- Borman, W. C. (1979). Format training effects on rating

accuracy and rater errors. Journal of Applied Psychology, 64(4), 410-421.

- Bradley, R. C. (1983). Preparing teachers for the principal's instructional visits. Educational Administration and Supervision, 4(1), 15-17.
- Braskamp, L. A., Brandenburg, D., & Ory, J. C. (1984). Evaluating Teaching Effectiveness: A Practical Guide. Beverly Hills, CA: Sage Publications.
- Bridges, E. M. (1974). Faculty evaluations-a critique and a proposal. Administrator's Notebook, 22, 1-4.
- Brody, J. A. (1977). A good teacher is harder to define than find. The American School Board Journal, 164, 25-28.
- Brown, E. M. (1968). Influence of training, method, and relationship on the halo effect. Journal of Applied Psychology, 52(3), 195-199.
- Buck, S., & Tiene, D. (1989). The impact of physical attractiveness, gender, and teaching philosophy on teacher evaluations. Journal of Educational Research, 82(3), 172-177.
- Bullock, W., & David, J. (1985). Interpersonal factors that influence principals' ratings of teachers performance. Planning and Changing, 16(1), 3-11.
- Bulzer, W. K. (1988). Biases in the recording of performance-related information: The effects of initial impressions and centrality of the appraisal task. Organizational Behavior and Human Decision Processes, 37, 329-347.
- Cathoun, J. G. (1988). Evaluating medical student clinical skills performance: Relationships among self, peer, and expert ratings. Evaluation and the Health Profession, 7(2), 201-212.
- Cardy, R. L., & Dobbins, G. H. (1986). Affect and appraisal accuracy: Liking as an integral dimension in evaluating performance. Journal of Applied Psychology, 7(4), 672-678.
- Cardy, R. L., & Kehoe, J. F. (1984). Raters selective attention ability and appraisal effectiveness: The effects of a cognitive style on the accuracy of differentiation among ratees. Journal of Applied Psychology, 69(4), 589-594.

- Cleveland, J. N., Murphy, K. R., & Williams, R. E. (1989). Multiple uses of performance appraisal: Prevalence and correlates. Journal of Applied Psychology, 74(1), 130-135.
- Clive, F. (1984). What's new in performance appraisal? Personnel Management, 16, 20-23.
- Constable, E., & Andrich, D. (1984, April). Inter-judge reliability: Is complete agreement among judges the ideal? Paper presented at annual meeting of the National Council on Measurement in Education, New Orleans, LA.
- Cooper, W. H. (1981). Conceptual similarity as a source of illusory halo in job performance ratings. Journal of Applied Psychology, 66(3), 302-307.
- Cooper, W. H. (1981). Ubiquitous halo. Psychological Review, 90(2), 218-244.
- Crain, J. A. (1990). Inter-rater reliability of the Texas Teacher Appraisal Instrument. Unpublished doctoral dissertation, University of North Texas, Denton.
- Crain, J. A., Worley, E. N., & Dunn, K. S. (1987). Exceptional quality: A second step. Abilene, TX: Worley Crain Associates.
- Darling-Hammond, L., Wise, A. E., & Pease, S. R. (1983). Teacher evaluation in the organizational context: A review of the literature. Review of Educational Research, 53(3), 285-328.
- DeNisi, A. S., Robbins, T., & Cafferty, T. P. (1989). Organization of information used for performance appraisals: Role of diary-keeping. Journal of Applied Psychology, 74(1), 124-129.
- Dobbs, L. (1989). A quasi-experimental study of inter-rater reliability when awarding exceptional quality points on the Texas Teacher Appraisal Instrument. Unpublished doctoral dissertation, University of North Texas, Denton.
- Doyle, K. O., Jr. (1983). Evaluating teaching. Lexington, MA: D. C. Heath and Company.
- Edwards, M. R. (1983). Productivity improvement through innovations in performance appraisal. Public Personnel Management, 12(1), 13-14.
- Feldman, J. M. (1981). Beyond attribution theory: Cognitive processes in performance appraisal. Journal of Applied Psychology, 66(2), 127-148.

- Ferguson, H. L., & Enger, J. M. (1985). Missouri performance based teacher evaluation program: Principal perceived ratings and projected job targets. (ERIC Document Reproduction Services No. ED 265 205)
- Friedman, M. G. (1986). 10 steps to objective appraisals. Personnel Journal, 65(6), 66-71.
- Frick, T., & Semmel, M. I. (1978). Observer agreement and reliabilities of classroom observational measures. Review of Educational Research, 48(Winter), 157-184.
- Furman, R. L. (1987). Supervising Evaluation of Teaching. New York: Vantage Press.
- Gaugler, B. B., & Thornton, G. C., Jr. (1989). Number of assessment center dimensions as a determinant of assessor accuracy. Journal of Applied Psychology, 74(4), 611-618.
- Glasman, N. S., & Paulin, P. J. (1982). Possible determinants of teacher receptivity to evaluation. The Journal of Educational Administration, 20(2), 148-171.
- Graves, J. P. (1982). Let's put appraisal back in performance appraisal: Part 1. Personnel Journal, 61(11), 844-849.
- Gray, P. J. (1982). A study of the effects of training and providing evidence on the accuracy, reliability, and attitudes of evaluation of teachers (Doctoral dissertation, Boston University School of Education, 1982). Dissertation Abstracts International, 44, 4405A.
- Greenberg, J. (1986). Determinants of perceived fairness of performance evaluation. Journal of Applied Psychology, 71(2), 340-342.
- Greenfield, W. (1987). Instructional leadership: Concepts, issues, and controversies. Boston: Allyn and Bacon.
- Guion, R. M. (1965). Personnel testing. New York: McGraw-Hill.
- Hall, J. L., Posner, B. Z., & Harder, J. W. (1989). Performance appraisal systems: Matching practice and theory. Group and Organization Studies, 14(1), 51-69.
- Harris, B. M. (1986). Developmental teacher evaluation. Newton, MA: Allyn and Bacon.
- Hauenstein, N. M., & Foti, R. (1989). From laboratory to practice: Neglected issues in implementing frame-of-reference rater training. Personnel Psychology, 42(2), 359-378.

- Independent School District. (1987, August). Guidelines for interpreting exceptional quality performance (rev. ed.). TX: Author.
- Ivancevich, J. M. (1979). Longitudinal study of the effects of rater training on psychometric error in ratings. Journal of Applied Psychology, 64(5), 502-508.
- Kaplan, M. F. (1973). Stimulus inconsistency and response dispositions in forming judgements of other persons. Journal of Personality and Social Psychology, 24, 58-64.
- Keefe, J. W. (Ed.). (1988). Profiling and Utilizing Learning Styles. Reston, VA: NASSP.
- Landy, F. J., & Farr, J. L. (1983). The measurement of work performance. New York: Academic Press.
- Latham, G. P., Wexley, K. N., & Purcell, E. D. (1975). Training managers to minimize training errors in the observation of behavior. Journal of Applied Psychology, 60(4), 550-555.
- LeBrun, P. F., Jr. (1986). Appraising teacher performance: A catalyst to improvement. NASSP Bulletin, 70(492), 56-60.
- Lemley, R. E. (1986). Training the assessor: Design ensures sharpening assessment skills. NASSP Bulletin, 70(486), 26-29.
- Lewis, J. (1973). Appraising teacher performance. West Nach, NY: Parker Publishing.
- Ligon, G., & Ellis, J. (1986, April). Adjusting for rater bias in teacher evaluation: Political and technical realities. Paper presented to the annual meeting of the American Educational Research Association, San Francisco, CA.
- Littlefield, J. H. (1985, March-April). Metacognition of performance raters. Paper presented to the annual meeting of the American Educational Research Association, Chicago, IL.
- Littlefield, J. H., & Trondle, G. R. (1987, April). Effects of rating task instructions on consistency and accuracy of expert raters. Paper presented to the annual meeting of the American Educational Research Association, Washington, DC.
- McConnell, J. H. (1971). The assessment center: A flexible program for supervisors. Personnel, 48, 35-40.

- McIntyre, R. M., Smith, D. E., & Hassett, C. E. (1984). Accuracy of performance ratings as affected by rater training and perceived purpose of rating. Journal of Applied Psychology, 69(1), 147-156.
- McGreal, T. L. (1983). Improved classroom observation skills in successful teacher evaluation. Alexandria, VA: Association of Supervision and Curriculum Development.
- Medley, D. M., Coker, H., & Soar, R. S. (1984). Measurement-based evaluation of teacher performance. New York: Longman.
- Medley, D. M. & Mitzel, H. E. (1963). Measuring classroom behavior by systematic observation. In N. L. Gage (Ed.), Handbook of research on teaching (pp. 247-328). Chicago: Rand McNally.
- Murphy, K. R., & Balzer, W. K. ((1986). Systematic distortions in memory-based behavior ratings and performance evaluations: Consequences for rating accuracy. Journal of Applied Psychology, 71(1), 39-44.
- Murphy, K. R., & Balzer, W. K. (1989). Rater error and rating accuracy. Journal of Applied Psychology, 74(4), 619-624.
- Murphy, K. R., Gannett, B., Herr, B., & Chen, J. (1986). Effects of subsequent performance. Journal of Applied Psychology, 71(1), 427-431.
- Murphy, K. R., Balzer, W. K., Kellam, K. L., and Armstrong, J. C. (1984). Effects of the purpose of rating on accuracy in observing teacher behavior and evaluating teaching performance. Journal of Educational Psychology, 76(1), 45-54.
- Napier, N. K., & Latham, G. P. (1986). Outcome expectations of people who conduct performance appraisals. Personnel Psychology, 39(4), 827-837.
- Nowakowshi, J. R. (1983). On educational evaluation: A conversation with Ralph Tyler. Educational Leadership, 40(8), 24-29.
- Ormrod, J. E. (1986). Predictors of faculty dissatisfaction with an annual performance evaluation. Journal of the College and University Personnel Association, 37(3), 13-17.
- Peterson, K. D. (1989). Costs of school teacher evaluation in a career ladder system. Journal of Research and Development in Education, 22(2), 30-36.

- Peterson, K. D. (1988). Reliability of panel judgements for promotion in a school teacher career ladder program. Journal of Research and Development in Education, 21(4), 94-99.
- Popham, W. J. (1985, April). Teacher evaluations: Mission impossible. Paper presented at the meeting of the American Educational Research Association, Chicago, IL.
- Pritchett, J. (1986). Training the assessor: Process can enrich management techniques. NASSP Bulletin, 70(486), 34-35.
- Pulakos, E. D. (1984). A comparison of rater training programs: Rater error and accuracy training. Journal of Applied Psychology, 69(4), 581-588.
- Pulakos, E. D. (1986). The development of training programs to increase accuracy with different rating tasks. Organizational Behavior and Human Decisions Processes, 38, 76-91.
- Smither, J. W., Barry, S. R., & Reilly, R. R. (1989). An investigation of the validity of expert true score estimates in appraisal research. Journal of Applied Psychology, 74(1), 143-151.
- Smither, J. W., Collins, H., & Buda, R. (1989). When rater satisfaction influences performance evaluations: A case of illusory correlation. Journal of Applied Psychology, 74(4), 599-605.
- Splitt, D. (1985). How to conduct the evaluation processes. (ERIC Document Reproduction Service No. ED 256 040)
- Stow, S. B., & Frudden, S. J. (1985). Teacher evaluation: School administrators' perceptions of their competence at teacher performance evaluations. National Forum on Education Administration and Supervision, 2(3), 49-54.
- Sullivan, J. J., & Walker, J. J. (1981). How would your teachers rank you? NASSP Bulletin, 65, 35-37.
- Taylor, E. K., Parker, J. W., & Ford, G. L. (1959). Rating scale contest: Predictability of structured and unstructured scales. Personnel Psychology, 12, 247-266.
- Texas Education Agency (1986). Executive summary of the final report on the development of the Texas Teacher Appraisal system. Austin, TX: Texas Education Agency (ERIC Document Reproduction Service No. ED 294 323)

- Texas Education Agency. (1986). Teacher appraisal system: Teacher orientation manual. Austin, TX: Texas Education Agency.
- Tsui, A. S. (1983). Qualities of judgemental ratings of teaching performance: Dimensionality and stability. A version of the paper presented at the annual convention of the American Psychological Association, Anaheim, CA.
- Vance, R. J., Winne, P. S., & Wright, S. E. (1983). A longitudinal examination of rater and ratee effects in performance ratings. Personnel Psychology, 36(3), 609-620.
- Weare, J., Moore, J., & Woodall, F. (1985). Interrater reliability: A selected and annotated bibliography of articles concerning interrater reliability. (ERIC Document Reproduction Service No. ED 280 898)
- Wells, R. C. (1982). Guidelines for effective and defensible performance appraisal systems. Personnel Journal, 61(9), 776-782.
- Williams, K., DeNisi, A., Blencoe, A., & Cafferty, T. (1981). The role of appraisal purpose: Effects of purpose on information acquisition and utilization. Organizational Behavior and Human Decision Processes, 35, 314-339.
- Wise, A. E., Darling-Hammond, L., McLaughlin, M. W., & Bernstein, H. T. (1984). Case studies for teacher evaluation: A study of effective practices. Santa Monica, CA: Rand Corporation.
- Yap, K. C., & Capie, W. (1985, March-April). The influence of same day or separate day observations on the reliability of assessment data. A paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.
- Zedeck, S., & Casio, W. (1982). Performance decision as a function of purpose of rating and training. Journal of Applied Psychology, 67(6), 752-758.
- Zuick, R. (1986). Another look at inter-rater agreement. Princeton, NY: Educational Testing Service.