

76-1
10

ABSTRACT OF APPLIED PROJECT

James M. Brown, M.A. in Education

Graduate School
Morehead State University
1983

TEACHER EVALUATION CRITERIA:
A CONSENSUS IN EASTERN KENTUCKY

ABSTRACT OF APPLIED PROJECT

An applied project submitted in partial fulfillment of
the requirements for the degree of Education
Specialist at Morehead State University

by
James M. Brown
Committee Chairman: Dr. Ronald Mersky
Professor of Education
Morehead, Kentucky
1983

APP-KY THESES
371.144
B878A

ABSTRACT

This research surveyed teacher evaluation criteria in an Eastern Kentucky setting. Using the Berliner and Tikunoff instrument, which a review of literature showed had a strong foundation; teachers, principals, and parents were surveyed as to how they felt on the twenty-one items of teacher evaluation criteria. All of the respondents came from five counties in Eastern Kentucky: Floyd, Johnson, Lawrence, Martin, and Pike. Teachers and principals were limited to those who serve on the elementary level and the parents were those who had children in the elementary grades. One variable that was included with the educators was that of the number of years they have been employed in education. Those with five or less years of experience could then be compared to those with more experience to see if longevity affects one's attitudes. This made a total of five categories that a respondent could be placed into: new teachers, older teachers, newer principals, older principals, or parents. From these five categories four comparisons were made: new teachers vs. older teacher, newer principals vs. older principals, teachers vs. principals, and professional educators vs. parents. Every comparison produced from three to five significant differences. Concerning the teachers, experience appeared to make the teacher more decisive in rating negative and positive behavior; thus implying that the older teacher has more confidence in the evaluation process. Older teachers also saw greater need to check on

83/12/16 J. Brown list

student progress which could indicate that they place more value on student output and extrinsic evaluations.

For principals, experience was not a major factor as almost all of them had more than five years of experience.

Teachers differed from principals on four points of teacher evaluation criteria. In every case the teachers gave higher ratings to all four items placing the emphasis on making sure the student is learning. Here teachers appear to be more "product" oriented than the principals. Carrying this idea to the extreme it might show that principals lean more in the direction of running an efficient school by working on the real goals rather than the stated goals which are given to the public.

Parents differed in four areas with professional educators. These items were centered around the type of instruction a teacher gives as to whether or not spontaneity is appropriate. Parents leaned toward a well defined and structured environment to learn in; while the teachers saw a need for using happenings and occurrences that arise naturally for teaching experiences. Parents also indicated they wanted their child treated with more individuality than collectivity as compared with what the professional educators thought necessary.

This study had a total of sixteen significant differences of

opinion among the noted categories. It should be remembered, however, that the groups do not necessarily disagree with each other but rather that they differ on the amount of emphasis to place upon certain items of teacher evaluation criteria. These differences should be kept in consideration when the evaluation process is going on. This idea is supported by the review of the literature which shows that the evaluation process benefits by an exchange of ideas concerning what and how a teacher is to be evaluated.

Perhaps a conclusion that can be drawn from the review of literature is that parents have for too long a period of time been left out of the evaluation process. By including parents in the process, research indicates that their consensus will facilitate a smoother teacher evaluation.

This study indicates that parents and educators have similar perceptions as to what constitutes good teaching. The study suggests that schools would be wise to show the parents that they are interested in similar goals. By showing parents that their goals are aligned with the school's goals, the bonds of community relations can be strengthened. As for the four items which showed a discrepancy, administrators probably should communicate the differences so that a dialogue can be established.

Lastly, the reader can conclude that there is a solid foundation

of criteria available here for teacher evaluations that is acceptable to all concerned parties. This study confirms other studies that have shown administrators need not be arbitrary in their evaluations. By the use of presently available material that is acceptable to principals, teachers and parents, principals can evaluate on valid criteria. Although there were a few areas of disagreement, this study confirmed that there is a significant amount of agreement among all three groups.

Accepted by:

Ronald Neely, Chairman

Robert C. Keenan

William J. Moore

APPLIED PROJECT

James M. Brown, M.A. in Education

Graduate School
Morehead State University
1983

TEACHER EVALUATION CRITERIA:
A CONSENSUS IN EASTERN KENTUCKY

APPLIED PROJECT

An applied project submitted in partial fulfillment of
the requirements for the degree of Education
Specialist at Morehead State University

by
James M. Brown
Committee Chairman: Dr. Ronald Mersky
Professor of Education
Morehead, Kentucky
1983

TABLE OF CONTENTS

	Page
ABSTRACT	2
CHAPTER I INTRODUCTION	10
DEFINITION OF TERMS	18
CHAPTER II REVIEW OF RELATED LITERATURE	20
CHAPTER III METHODOLOGY	33
SELECTION OF THE SAMPLE	33
VARIABLES	35
INSTRUMENT	36
DATA COLLECTION AND PROCESSING	36
CHAPTER IV RESULTS	40
DESCRIPTIVE STATISTICS	40
INFERENCEAL STATISTICS	44
EVALUATION OF INFERENCEAL STATISTICS	47
Professional Educators vs Parents	47
New Teachers vs Older Teachers	49
New Principals vs Older Principals	52
Teachers vs Principals	53
ANSWERING QUESTIONS	55
RESPONDING TO THE NULL HYPOTHESIS	58
CHAPTER V SUMMARY	60
CONCLUSIONS	62
RECOMMENDATIONS	64
APPENDIX	66
TABLES	68
REFERENCES	84

CHAPTER I

INTRODUCTION

Teacher evaluations have always been a sore point between teachers and those evaluating them. Part of the problem is that they cannot agree upon what constitutes good teaching. If they had a standard that was agreed upon before the evaluation process was begun then evaluators would know exactly what behavior to look for and the teachers would know exactly what was expected of them, thus relieving much of the tension that exists. This problem has been borne out in several studies over the recent years. (Deever, 1971: Combs, 1965: Flanders, 1967: Broudy, 1969: Berliner, 1976: and Heitzman and Starpoli, 1975).

Different definitions of teacher competence have been developed, and positions for and against teacher evaluation and specific evaluation tools have been explained. One facet of teacher evaluation that seemingly has been bypassed by many administrators is whether or not the evaluator (whether he is the principal or the supervisor) and the teacher agree upon what constitutes teacher competence for an elementary teacher. If teachers and administrators can agree upon definitions of teaching effectiveness, then it is likely that a purposeful teacher evaluation process will emerge. Conversely, if the principals and teachers maintain an adversarial role with each

other it is likely that the evaluation process will lack usefulness.

Studies have shown (Grant and Carvell, 1980) that in some cases there appear to be no significant difference between the two groups of principals and teachers concerning their opinions on what topics constitute desired teaching behavior. However, weaknesses within the Grant and Carvell study can be seen in their sampling and methodology if it is being used as a basis to judge the "collective whole." The purpose of their study was to determine whether or not teachers and principals agree on what constitutes desirable and undesirable teaching behaviors and techniques. Their sampling took place in one school district in the Midwest from a relatively homogeneous setting. Only one hundred teachers were polled and of that number only seventy-three responded. Of their principals, twenty-six respondents were found out of twenty-nine polled. Not only was there a small number of total respondents but there were no other variables controlled that might account for the attained results. Suppose, for example, that new principals disagree with teachers but the older principals have turned to thinking like the teachers. If the sample contains predominately older principals then the results are likely to be biased. Or perhaps newer teachers view the criteria in a manner similar to the principals but older teachers have become more militant and changed their opinions on the criteria. Even an interaction effect is possible when the time variable enters into the picture. In any case it is wise to consider the number of years that a person has been in education to determine whether or not it

effects the respondents views on the teacher evaluation criteria.

With this extra information coming from a new independent variable, simple T-Tests would not be sophisticated enough to handle the data properly. An ANOVA would be a more delicate instrument to use showing results that otherwise might remain hidden.

Those in the public schools owe it to the parents not to neglect their opinions since parents have a vested interest in their childrens education. It also seems prudent to enlarge the scope of research on teacher evaluation by considering the needs and desires of the clients. Historically, public schools exist for the benefits of the parents and their children. Although parents are generally not all professionally trained educators, they still have expectations of what should be occurring in the classroom. Because of this philosophy, this study will also include parents' evaluations of the teaching criteria.

It is not surprising in one sense to note that there is a great deal of similarity between principals and teachers, since they have similar training and both work for a living in a school setting. Most principals have even spent many years being a classroom teacher. (Kentucky law requires three years of successful classroom experience for every principal.) Given this information one can readily see how differences might be minimized. However, one does not need to spend four to six years in college to become a parent, neither does

he develop any bias by becoming absorbed into the system and becoming dependent upon the system such as the paid professionals do. In this sense one might expect to find a different perspective from the parents. They would not be looking for the most efficient method of running a classroom but rather the most effective way of educating their children.

Sociologists of complex organizations such as schools (Etzioni, 1961) have found that often an organization will have a stated or public goal and then in practice have a real goal that undermines the stated goal. This is known as goal displacement. For instance, the schools (teachers and administrators) might say, "We are here to give the students the best education possible," whereas in reality the goal is, "We are looking for the easiest way to get a student through twelve years of school with the fewest complaints and problems." The real goal does not necessarily rule out the stated goal because the best way may be the easiest way; but when it is not the easiest way something more convenient will be sought after and found.

The implications for this study could mean that teachers and principals presently look for efficient uses of time, money, and/or personnel rather than effective uses which would meet the stated goals. Priorities that are shifted away from the stated goals for whatever reason would then be seen as decreasing the quality of the education promised in the stated goals. This is not to say that

parents are more qualified than professionally trained educators but merely that they bring with them a fresh perspective that should be considered in the evaluation of teacher behavior. Perhaps parents can help point out discrepancies between the real and the stated goals.

Several important assumptions provided the underlying framework for this study. One assumption being that teachers, principals, and parents are capable of forming valid opinions of theory that are transmittable to practice. Another assumption implicit in this study is that the twenty-one item survey by Berliner and Tikunoff is as universally valid as the previous studies have shown it to be (Berliner and Tikunoff, 1971; Grant and Carvell, 1980.) Lastly, this study assumed that teaching behavior can be objectively measured by these twenty-one items.

This study answered or lent insight to the following questions:

1. Do teachers and principals embrace similar philosophies regarding what constitutes desirable teaching behavior?
2. Does the number of years one spends in education effect the results of the survey?
3. Do parents agree with professional educators as to what constitutes desired teaching behavior?

4. Do East Kentucky teachers and principals differ from their counterparts in the Midwest?

5. What are the patterns of significant differences that exist, if any?

6. How are these twenty-one items clustered, and if they are, could they be grouped together to make a smaller instrument?

Answers to these questions are important for several reasons. Such an exploratory analysis provides one indication of the compatibility of ideas among these three groups. If all three groups agree upon what constitutes desired teaching behavior then indeed, this would be a go ahead signal to use this model or an abbreviated form in our classrooms. However, if there is disagreement then negotiations are needed to rectify the situation. Differences here could be a warning to principals to clear the air with teachers or parents before the year begins as to what is to be expected within the classrooms. Much tension and many hostilities could thus be eliminated. In administration and supervision, these questions have never been addressed in an empirical study.

In this study, a review of the literature determined that teachers and principals will usually agree upon criteria of teacher evaluation. What was not found in the literature was any indication as to why they usually agree. If educators agree, and parents

disagree on certain items in this study, then there are two basic variables which could account for the difference. They are the difference in educational attainment and/or the fact that teachers and principals are on the inside of education's complex organization seeking real goals instead of stated goals.

Null Hypotheses

In inferring differences between these three groups of respondents the following null hypotheses were tested:

1. There is no difference between teachers and principals concerning their philosophy of desired teaching behavior in the classroom as measured by the Berliner and Tikunoff instrument.

2. The number of years a teacher or principal spends in education does not affect the results of the survey.

3. There is no difference between professional educators and parents concerning their philosophy of desired teaching behavior in the classroom as measured by the Berliner and Tikunoff scale.

4. There is no significant difference between East Kentucky teachers and principals and their counterparts in the midwest as studies by Grant and Carvell, 1980.

DEFINITION OF TERMS

Three categories of respondents were surveyed. One category was composed of elementary principals or head teachers from schools which (at least) contained grades one to six (1-6) and sometimes kindergarten to eighth (K-8). All principals mentioned in this study came from one of Kentucky's five most eastern counties, as did the teachers and parents. Another category was that of elementary teachers. In this report a teacher is defined as one employed in a public school to teach students that are in grades K-six. The third category was that of parent. For the purpose of this study only parents of children in grades K-six were used.

Other definitions used in this study are:

Effective--that which best meets the needs of the stated goals to their ultimate completion.

Efficient--that which is most cost expedient to save on materialistic matters and gives the greatest output for the least input. Efficient best meets the real goals of the organization (Etzioni).

Respondent--a person from one of the three categories previously defined.

Professional educator--an elementary teacher or principal.

New--those educators with five or less years of experience.

Older--those educators with more than five years of experience.

The next chapter presents a review of the literature with emphasis on other instruments of teacher evaluations and their effectiveness along with where the input comes from in constructing such an instrument. Also of equal emphasis will be the results of these other examples. Chapter 3 is on methodology and describes the sample used in this study, the instrument, and the statistical treatment employed. Chapter 4 describes the results. Chapter 5 concludes this study with a discussion of the findings and implications for future study.

CHAPTER II

Review of Related Literature .

A principal's job of evaluating the performance of teachers has been complicated by the absence of a satisfactory yardstick to measure teaching effectiveness. Research on performance evaluation indicates a lack of consensus on the most appropriate method of evaluation. History has shown however, that this job need not rest solely upon the principal (Pembroke and Goedert, 1982).

Teachers have traditionally been evaluated, both formally and informally by parents and educational administrators (Ovard, 1975, p. 87). In fact, evaluation of teachers can be traced back as far as the Middle Ages (Eastridge, 1976, p. 48) and can be seen more recently in the contractual agreement between teachers and administrators in which provisions have been made for teacher evaluation (National Educational Association, 1969). Such evaluation is considered necessary in order to provide students with the quality education they need and deserve while providing teachers with the chance of developing into a master teacher (Suprina, 1978, p. 54). Unfortunately in spite of its important role in education, teacher evaluation is often based on nebulous criteria and administered by unqualified personnel (Goens, 1976, p. 15). Such conditions may be due to the lack of quantitative research (Watson, 1963, Chapter 3)

upon which to base effective teacher evaluation or the development of suitable measuring instruments in education.

"In examining thirty-nine research studies Barr, in 1952, (Quite possibly his findings are now outdated but still worth noting) reported the following conclusions:

1. No one appears to have developed a satisfactory working plan or system that can be used by personnel officers who must make judgments about teacher effectiveness.

2. Little has been done in evaluating the non-classroom responsibilities of the teacher-his activities as a friend and counselor of pupils, his activities as a member of a school staff, his activities as a member of the school community, and his activities as a member of the profession.

3. Very little has been done in differential measurement and prediction. Concern seems to have been chiefly with the general merit of teachers. Administrators often need teachers with special abilities.

4. Teaching effectiveness generally has been treated as something apart from the situation giving rise to it. More needs to be known about the situational determiners of effective teaching." (Barr, 1952)

Berliner and Tikunoff heavily emphasized Barr's priorities listed in conclusion number two. They overlap almost point for point. Lemley has shown (1983) that by actively involving teachers in the performance evaluation process the key to their commitment is attained. Performance evaluation should not be something done to teachers, but rather something that is done with teachers, both in developing a process for evaluation and conducting evaluation sessions. Performance evaluation done in conjunction with teachers connotes a positive and growth-related, rather than judgmental experience. Pembroke was also careful to note with the increasing trends toward merit pay, performance-based salary programs, and teacher accountability, that there is a growing need for some solid foundational guidelines to evaluate teacher performance. In summary, Pembroke points out that teacher involvement is the key to teacher commitment.

Before the theoretical model of teacher evaluation is approached it should be prefaced with a definition of "evaluation". The word "evaluation" derives from a combination of Latin and French. The derivation suggests an act that draws value from, or fixes the worth of a thing. Thus, engaging in evaluation, one is attempting to fix the worth or value of what we evaluate--teacher performance (Lemley, 1983).

In the beginning of a teacher's career, during the months of student teaching, studies have shown many successful methods for

evaluating teaching from Ned Flander's system (1965) of verbal interaction to a more comprehensive analysis using microteaching (Allen and Ryan, 1969). Reinhartz and Beach (1982) advocate using these same methods throughout the teacher's career. Their findings show that principals have a tendency to evaluate preservice teachers in different manners than they do inservice teachers. Preservice teachers were evaluated with more consistency than inservice teachers and at a higher rating too. Beach suggested that this could be explained because of more open dialogue concerning the preservice teacher while the evaluation of an inservice teacher is often done on a one-to-one basis. What this implies for the present study is that when teachers and parents can enter the evaluation process with principals, dialogue will increase and so should teacher ratings.

Marsh (1982) has also completed studies showing that dialogue or feedback from ratings, coupled with a frank discussion of their implications with an external consultant, can be a practical intervention for improving teaching effectiveness.

From the various methods of evaluating teachers there can be seen two broad classifications, product-oriented evaluation and process-oriented evaluation. Weaknesses in the process-oriented evaluation, for example, would rate teachers highly if they can lecture smoothly and eloquently. But if the student learns little, then such teachers are not effective (Popham, 1982).

On the Berliner and Tikunoff instrument of teacher evaluation most of the items are centered around process-oriented factors but there are about five of the twenty-one items that are product-oriented, numbers 13, 14, 18, 19, and 20. (In this sense it would be practical to put these five items in a cluster). Although Popham tends to find product-oriented evaluations at the crux of the matter Pembroke and Goedent agree with Berliner and Tikunoff in seeking a balance. "If an evaluation system is based solely on changes in students' behavior (products), the effect of the teacher on those changes, or absence of change, cannot be isolated from the effect of all other individuals with whom the student interacts. Consequently, an evaluation system that focuses on what the teacher does, as well as what the teacher accomplishes, provides a better balance" (Pembroke and Goedent, 1982).

Gerald Pine (1975) establishes certain conditions for evaluating teachers. "The teacher is not only accountable to himself, his students, and his colleagues, but also to administrators, parents, school boards, and the community at large. It is good to remember that these people are not unreasonable in their quest for data that demonstrate the value of teaching. Parents and community members have a right to know whether teaching is effective in their schools.

To effectively evaluate and supervise teachers so that they will improve their teaching skills and approaches, to render more

meaningful and effective assistance to youngsters, and to generate data that will be helpful to parents and the community, it seems that the following constitute minimal and necessary conditions for teacher evaluation:

The use of appropriately designed evaluative instruments that include criteria reflecting the body of theoretical and empirical knowledge derived from professional literature and research.

The establishment of evaluative criteria flexible enough to encompass varied theoretical positions and individual styles of teaching, i.e., individualized evaluation to teachers.

A statement of criteria understandable to teachers, administrators, supervisors and parents.

A plan of evaluation that includes judgments from both the internal and external frames of reference.

A continuous process of evaluation with established monitoring points so that the teacher and appropriate supervisory personnel have some specific time reference for gauging and discussing individual progress.

A plan of evaluation consistent with democratic and psychological

principles of supervision.

A clearly stated philosophy and rationale for evaluation and supervision derived from the contributions of teachers, supervisors and parents.

A clearly defined but flexible methodological procedure for collecting data to test evaluative criteria for the evaluation of each teacher, such as:

a. Teacher and supervisor listen to and analyze the tapes of the individual teacher's classes.

b. Teacher and colleagues view and analyze the tapes of the individual teacher's classes.

c. Supervisor observes the teacher and confers with the teacher individually.

d. Teacher conducts personal research regarding his effectiveness and shares the results for critique with his supervisor and/or colleagues. For example, this might be accomplished through the use of questionnaires or surveys of pupils, colleagues, and parents.

e. Teacher writes a self-evaluation, and supervisor writes

an evaluation of the teacher periodically. Together, they discuss the results.

A plan of evaluation that includes an annual review by teachers and supervisors of evaluative processes and criteria.

An annual orientation by supervisory personnel and teachers to inform school boards, parents, and the public of how teachers are evaluated.

A plan of evaluation characterized more by a horizontal supervisory relationship between teacher and supervisor than by a vertical relationship.

A plan of evaluation that has been developed by teachers and supervisors working together, and which has evolved from a free and open discussion of the philosophical, theoretical, and empirical considerations that influence the work of the teacher.

A plan of evaluation that takes into consideration local conditions, needs, resources, and principles.

A plan of evaluation which encourages openness of the teacher's self rather than concealment.

If these conditions are met, evaluation will be a facilitating

and enhancing process characterized by respect for the teacher, concern for the students, and response to the community. Under such conditions confrontation and differences of opinion become constructive forces and teachers invest themselves more fully and openly in a collaborative and interactive process of evaluation designed to promote professional growth and facilitate student accomplishment." (Pine, 1975)

Margaret Verble's studies have shown that rating scales, however popular, carry serious limitations as evaluation, feedback, and grading tools (1979). Their appearance of objectivity is an illusion. When rating a specific, isolated competence, the criteria of difference between, for instance, "superior" and "outstanding" or "2" and "3" are not always possible to articulate in reference to specific performances, and as distinctions become finer, objective criteria become harder to define. Additionally, the sum total of the evaluation ratings assigned the parts of a performance do not always represent a fair evaluation of the whole. Most people who have worked with rating scales know that many times an individual will achieve several isolated competencies, but the over-all performance will leave much to be desired. In cases like this the halo effect needs to be guarded against. More often, performances of individual competencies may not be outstanding, but the total performance will be such that all viewers can agree upon its excellence.

"In inservice teaching situations, we generally identify and agree on good teachers, so-so ones, and bad ones. Teaching is a public activity that refers directly to those taught, rather than to a list of criteria or competencies. In any given school and no matter how they may be rated on a pseudo-objective scale, good teachers are those whom students think are good and bad teachers are those who are considered bad by their students. And everybody in the school--and the community--knows who they are" (Verble, 1979). This implies that parents can be good judges of what goes on in the classroom.

John F. Huntley, in an article for the "Harvard Educational Review," discusses the difference between extrinsic and intrinsic classification methods and in doing so throws light upon the process by which one discerns quality:

In the abstract, three considerations distinguish the extrinsic from the intrinsic case, sorting traits are single, simple, and space bound, and therefore subject to quick discernment and easy notation: in the intrinsic, they are relational configurations or time bound processes which are brought into focus by prolonged systematic observation. Second, for extrinsic classification, the sorter's personal touchstone, metaphorically speaking, is readily calibrated against some extrinsic, public, and absolute standard: for intrinsic, the sorting criterion must be reaffirmed and made conceptually manifest by looking for contrastive differences in the population itself. And finally, the extrinsic process may proceed quickly and assuredly through a large population on a one-by-one, each-as-they-come basis: but the intrinsic sorters must restudy the whole group before segregating its members (Huntley, 1976).

Through intrinsic evaluation, judgments regarding quality are decided about complex entities such as teaching, and while rating scales attempt extrinsic evaluation, public and private school teachers actually are evaluated by their students by intrinsic methods: students observe their teachers over a period of time, they compare them to other teachers they have had, and they, generally unconsciously, evaluate them and re-evaluate former teachers in the light of such comparisons.

In the light of these considerations, the Berliner and Tikunoff instrument appears to be both an intrinsic and an extrinsic evaluation procedure for the evaluation of the performance of teachers. The goals of the evaluation methodology were to: (1) insure that the best performances were evaluated as such: (2) stimulate as nearly as possible the methods by which teaching performance is actually measured.

Verble's experiment obtained some extremely high positive correlations substantiating Pirsig's assertion that people can agree upon what quality is (Pirsig, 1974). More specifically and importantly for the purpose of this study, the correlations support the idea that people can, in concrete situations agree upon what good teaching is and what it is not. These correlations also indicate that intrinsic shared evaluation offers a legitimate way to evaluate teachers. We can agree upon the relative merits of teaching performances. We can make distinctions among performances,

and consensus can confirm the legitimacy of those distinctions. The confirmation evidenced in the correlations removes the evaluation of teaching from the pseudo-objectivity of the instructor, translating a subjective judgment into numbers on a rating scale whose competencies can never be defined completely, and places it in the same realm as, for instance, evaluations of the fine arts such as painting, ceramics, or music.

This type of evaluation procedure more actively engages the participants in the evaluation process and the problems of evaluation. Using "teachers in such considerations is, in and of itself, valuable. On the other hand, rating scales employed by students are not so instructive--they, in effect, categorically define for students what good teaching is and then offer them a pat solution to the problems of evaluation (Verble, 1979)." It is for this reason that this study has not included students in the classification of respondents.

The public agreement upon the quality of teaching is ultimately the test by which teachers in the schools survive or perish. Good teachers are those whom their constituency considers good, and bad teachers are those whom their constituency considers bad. (Constituency being defined as those whom the teachers serve, i.e., administrators, parents, and students). Other criteria, such as competencies and skills performed, are not the final measure. Yet they are the rulers by which educators have measured teaching

performances. It would seem more logical to judge teaching performance by a combination of intrinsic and extrinsic public evaluation but this idea is not too common today (Verble).

Supervisory ratings are by far the most widely used measure of the performance of subordinates. It is therefore in the interest of both the organization and the individual to maximize the accuracy of ratings. A great deal of effort has gone into the development of alternative rating systems and scale formats, but to date there is little consistent evidence that this approach has materially affected the quality of performance ratings (Landy & Farr, 1980). Several recent reviews (Cooper, 1981; Feldman, 1981; Landy & Farr, 1980) have suggested that the key to improving ratings lies in studying the rating process rather than attending exclusively to rating outcomes. Logic has it then, that an instrument which could meet the approval of all concerned parties would be that key to improving ratings.

Observation and categorization of ratee behavior is the first step in making judgments about performance (Borman, 1978; Cooper, 1981). One way of increasing the accuracy of performance appraisals might be to increase raters' accuracy in observing ratee behavior. Both Bernardin and Walter (1977) and Thornton and Zorich (1980) showed that rater errors are related to accuracy in behavioral observation, Thornton and Zorich also showed that training increases raters' accuracy in observing and recalling specific behavioral events.

CHAPTER III

Methodology

This section describes the makeup of the sample, the selection of variables concerning the research, the statistical methods used, the data collection procedures, and other methodological factors.

Selection of the Sample

The people surveyed fell into one of three categories: teachers, principals, or parents. All three categories came from five counties in East Kentucky: Martin, Johnson, Floyd, Lawrence, and Pike. These counties are assumed typical of rural Appalachia because they represent a wide range of economy and social mores. Thus, these results should be generalizable to other rural sections of Appalachia. Cost prevented this study from expanding to many more counties.

Teachers were collected from Martin County by distributing a survey in every teachers mailbox who works at one of the seven elementary schools. This included one hundred twelve teachers. The teachers in the other counties were reached by means of lists obtained from the superintendents when this project was explained

to them. Randomly selected from the lists names were chosen and surveys were mailed to them with an enclosed self-addressed stamped envelope. The envelopes were not marked in any way as to give away the identity of the respondent. Because Paintsville Independent Schools are separate from Johnson County Schools, the surveys were administered the same way here as they were in Martin County, one in every elementary teachers mail box. All together in the counties outside of Martin there were one hundred fifty-seven (157) teacher surveys sent out bringing the total of teachers approached to two hundred sixty-nine (269). From this number one hundred thirty-nine (139) responded or fifty-two percent (52).

Principals were collected from Martin County by distributing a survey in their mailboxes. All principals were approached with a survey in this manner. In the other counties every elementary principal that was on the lists from the superintendents was mailed a survey along with a self-addressed stamped envelope. In all, fifty-six (56) principals were approached and thirty-nine (39) responded for a percentage of seventy (70).

All of the parents who responded were from Martin County. Lists of parents to choose from randomly were not available. The parents selected were ones who live in the Inez, Warfield, and Turkey Creek areas of the county who had visited the schools during the Spring semester of 1983. It should be understood that some of a sampling bias might exist here in this case but it must be

remembered that these parents are samples of the ones that teachers deal with anyway.

Other restrictions placed upon the parent group are that they had to be parents of elementary children and their children had to be presently enrolled in grades K-6. This was done to ensure the possibility that they would have a vested interest in what goes on in the elementary classroom. The whole purpose of this method of selecting the parents was to find parents who would be expected to qualify as having a concern with the items on the survey. A total of twenty-seven (27) parents responded to the survey out of forty-eight (48) for a percentage of fifty-six (56).

Variables

In this research design there are two independent variables. First, is the classification of the respondent whether he/she is a principal, teacher, or parent. The second variable is concerned with only the professional educators, that is the number of years the respondent has been in education. The professional educators were divided into two categories, those with five or less years of education experience and those with more. Five years was chosen as a breaking point because normally this will allow the educator to obtain tenure and become fully entrenched in the educational system. This study will have twenty-one dependent variables. They are the items on the Berliner and Tikunoff instruments to measure teacher

behavior. These dependent variables will be examined separately and collectively to the degree that clustering similar items is possible. (See appendix 1 for a copy of the instrument.)

Instrument

The instrument itself consisted of twenty-one items from a list formulated by Berliner and Tikunoff. Berliner and Tikunoff (1971) had conducted a study for the California Commission for Teacher Preparation and Licensing. Using ethnographers (anthropologists and sociologists) as observers in forty California classrooms, Berliner and Tikunoff sorted out twenty-one behaviors that discriminated between more-effective and less-effective teachers. These twenty-one behaviors sufficed as the formulation for the twenty-one items in the survey administered in the study.

Data Collection and Processing

All surveys that were collected were fed into the 550 computer at Morehead State University and the data was examined using the Statistical Analysis System (SAS) provided by the University of Kentucky's IBM 370 and the Statistical Package for the Social Sciences (SPSS) from the same source.

There were five groups that a respondent might fall into: Parent, Teacher with less than or equal to five years experience,

Teachers with more than five years experience, Principals with less than or equal to five years experience, and Principals with more than five years experience. Due to the low N (number) of principals with five or less years of experience (4) this group will not be examined quite as extensively as the others.

For these groups the means, standard deviations, standard errors of measurement, skewness, and kurtosis will be derived and compared using combination of Z scores, T-tests and ANOVAS. Z scores will show significant differences between the mean of the scores obtained on each of the twenty-one dependent variables. T-test for independent samples will examine the differences between two means at a time. For example, the means of teachers having more than five years experience with the means of teachers having less than five years experience.

Lastly, an Analysis of Variance will be used to determine whether there is a significant difference, between all means at the selected probability level of an alpha $<.05$. Most statistic texts outline the procedure to be followed so this paper shall not go into the mathematics of any of these operations. The ANOVA that will be performed is a nonparametric procedure for testing that the distribution of a variable has the same location parameter across different groups. This procedure is called NPAR1WAY for short and it performs analysis of variance on ranks and certain rank scores of a response across a one way classification.

Most nonparametric tests are derived by examining the distribution of rank scores of the response variable. The rank scores are simply functions of the ranks of the response, where the values are ranked from low to high, which is exactly how the Berliner and Tikunoff scale is composed. Statistics defined as linear combinations of these rank scores are called linear rank statistics. The NPAR1WAY was used to calculate the Z-scores used in this study.

With these statistical procedures it is recognized that they will have more power than the Grant study because of the larger N. Another factor increasing the power of the ANOVA is that this is a test which reduces within-group variance or error (Gay, p. 256).

Assumptions and Limitations of the Study

In the cases where an elementary teacher also happened to be an elementary parent, (there were no cases of principals being elementary parents) the preference was to always classify these people as teachers. This did decrease the number of potential parent respondents but it was assumed these people had also been through four or more years of higher education and would also be influenced by the fact they are on the inside of this formal complex organization. It is also assumed that all respondents do have a vested interest with the occurrences of an elementary classroom and would take this survey seriously.

This study, even though larger than the Grant study, is still limited by the number of respondents. It did work with a larger geographic area, but of course anytime a study can have a larger N it is bound to be less limited.

Another obvious limitation is that the manual gathering and inputting of so much data is subject to error. Although intense scrutiny is exercised to avoid error, it is, however, assumed that errors would be randomly distributed and would not adversely affect any particular result or finding.

CHAPTER IV

Results

Descriptive Data

A brief overview of the data shows that there are one hundred thirty-nine (139) teachers, seventy-six (76) with more than five years of experience and sixty-three (63) with five or less years of experience. Principals have thirty-nine (39) respondents and all but five of them have more than five years of experience. Years in education were not considered with parents so this will result in missing values for this variable. A total of twenty-seven (27) parents responded to the survey. (See table 1)

Table 2 shows the means, standard deviations, standard error of means, skewness, and kurtosis for all twenty-one dependent variables regardless of who the respondent was in the survey. It should be noted that teachers were the most predominant group in making up this data since they composed 67.8 percent of the sample. Thus it is possible to see many similarities between this table and table number 3 which is just the teachers. Tables 4 and 5 show the same descriptive statistics for principals and parents respectively. There was a high degree of consistency among all three groups on certain items such as B, D, S, and T which received high

ratings and A and U which received low ratings. Those variables, C and H, with means close to the middle number three and those variables showing relatively high standard deviations indicate that these items were not as decisively pronounced by their respective group of respondents. A significant amount of variability within the group could suggest that the item is not a generally accepted point of teacher evaluation. The items that would end up making ideal criteria are those that show a consistently high or low ranking across the board to all groups of respondents and have a relatively low standard deviation. For example, respondents regardless of classification items (B, D, G, J, M, P, R, S, and T) are all ranked positively and possess a standard deviation of less than 1.000. Item U is also ranked negatively and has a low standard deviation of about 1.1.

The skewness and kurtosis of each dependent variable show at which end most of the respondents lie on that item as well as the extremity, if any, of the minority. A normal distribution is symmetrical and the value of the mean, mode, and median are the same. However, when a distribution is skewed, as all of these items are to one degree or another, the mean, mode, and median will have different values. These differences will not be important for our purposes unless they are significantly different from a normal distribution. If the skewness number is positive this means that the extreme scores are at the upper end of the scale; and if the skewness number is negative then the extreme scores are at the

lower end of the scale. One assumption to remember with extremely high or low scores in skewness is that of normality. Normality is required for many statistics to be valid including all ANOVAs. It means that the population is normally distributed over the scale.

In examining the skewness results for the sample means, each skewness number was tested by changing it to a Z-score with the following formula: $Z = \text{skewness} / \sqrt{6/n}$. Those items which have an absolute value of Z at /1.96/ are duly noted with an asterisk. It is necessary to remember that one of the assumptions for further inferential statistics is that the skewness of the sample mean should not be significantly different from the general population mean.

In cases of a survey like this, one cannot assume that the general population will have a normally distributed bell curve with no skewness. On a one-to-five scale it would be highly improbable to find the general population always centered with mean, mode, and median on the three. The general population, therefore, is somewhat skewed to an uncertain degree. In cases where the skewness is significantly different from the normal curve, one should remember the assumptions of the Mann-Whitney U test. These differences might account for the skewness.

The kurtosis of the normal curve is three (3) and is called mesokurtic. Those means which have a kurtosis greater than three

are said to be leptokurtic while those less than three are platykurtic (Glass, p. 98). It can be determined whether or not the distribution is significantly kurtic by applying the following formula: $Z = \text{kurtosis} / \sqrt{24/n}$. A Z value greater than an absolute value of /1.96/ indicates a significant departure from normality. Cases where the kurtosis is significantly different are noted with an asterisk realizing this could play a role in comparing this mean with others.

The standard error is used when there is a difference found between sample means. It will determine whether or not the difference is a result of sampling error or a reflection of a true difference. An interesting characteristic of sampling errors is that they are normally distributed over the scale. The standard error of the mean will indicate how much the sample means could be expected to differ if other samples from the population were used.

The major factor affecting the standard error of the mean is sample size. As the size of the sample increases, the standard error of the mean decreases. The smaller the standard error of the mean is, the better the research. A smaller standard error indicates less sampling error (Gay, p. 282).

In using an ordinal set of data on a one-to-five scale some possible difficulties arise with the assumptions. One assumption of a normally distributed curve is that the median, in this case

three, will also be the mode and the mean. If that were the case, on every item of teacher evaluation the general population would have to be neutral. Of course if that were true there would be nothing to study and there would be no criteria on which to evaluate teaching. Since there is skewness in the general population there should be a similar skewness in any smaller population. If there is no skewness this would only indicate that the smaller population is centered around the midpoint of three. In the category of "All Respondents", only one item was not skewed significantly (H). This shows that the general population has not agreed on whether this is a positive point of evaluation or negative. It also indicates that the general population has agreed that the other twenty (20) items are valid criteria.

To summarize the value of the skewness, it will show the direction that the population is skewed, the degree that the mean varies from the median, or the point at which no difference exists for an item. The largest skewness figures available coincide with the highest means of 4.87 and 4.90 on the older teachers items D and T.

Inferential Statistics

As reported earlier the type of ANOVA that was performed was a nonparametric one way analysis of variance. The nonparametric test chosen was the Mann-Whitney U test because its assumptions

met perfectly the criteria of the data. The assumptions for the Mann-Whitney U are few in number and relatively straightforward. First the Mann-Whitney U assumes that there are two independent groups--that is, a single subject contributes one score to one of two groups. Second, scores must be ranked in order to perform the Mann-Whitney U test. Therefore, they must be measured on at least an ordinal scale. Thirdly, most nonparametric including this one assume the underlying distributions are continuous. If the distribution is continuous, ties should not occur. However, since our measurements are discrete, ties frequently do occur but they are corrected for in the program by making the alpha level more stringent. Lastly, if assumptions are met the Mann-Whitney U test evaluates differences in central tendency and/or shape of the general population distributions. In using the Mann-Whitney U test, the skewness and kurtosis should not be significantly different from the general population to obtain accurate results.

The Mann-Whitney U test is the most powerful nonparametric alternative to the parametric "independent T" test. For this reason other T-tests will not be performed on this data. The Mann-Whitney U makes no assumptions about normality or homogeneity and has an "asymptotic relative efficiency" of 95.5 percent in comparison to an Independent T-test. If the data are not normal, then the Mann-Whitney U may be even more powerful than the independent T-test. Consequently, the test may be used routinely in place of the T-test.

The Mann-Whitney test is commonly used to determine whether or not the measure of central tendency (mean or median) for two independent groups differ significantly. Actually, the Mann-Whitney U is sensitive to differences in distribution for the two populations. When the two populations being compared have identical shapes, the test compares the median for the populations. When the distributions are also symmetrical (i.e., no skewness present), then the tests can evaluate differences between sample means.

For the purposes of this study, the Mann-Whitney U test was performed four times. First, to compare teachers of five or less years experience with teachers of more than five years experience. Second, to compare principals of the two experience levels. Third, to compare teachers with principals regardless of experience; and fourth, to compare the professional educators with the parents.

Does this data meet the assumptions of the Mann-Whitney U test? Considering the first assumption there are two separate groups of subjects in each of the four Mann-Whitney tests performed. This assumption was satisfied with the teachers and principals. The Mann-Whitney U tests required at least an ordinal scale of measurement. Each of the dependent variables were ranked on a one-to-five scale making them ordinal, thus satisfying this assumption. Continuous distribution is a requirement of the Mann-Whitney U test. The dependent variables are continuous variables, however, since there are only five choices, ties do occur--SPSS will correct for ties with

large samples. Symmetry is also required, or at least comparably shaped distributions, in order to evaluate central tendency. Examination of the high numbers of samples without having significant differences suggests that this too will not be a major problem. Meeting this criteria, it seems likely that any differences between the groups are in central tendency.

Evaluation of Inferential Statistics

1. Professional Educators vs. Parents

In the comparison of the means of professional educators and the parents, there are four significant differences to be noted (see table 7). They are on items A, C, H, and I which correspond numerically to items 1, 3, 8, and 9 on the Berliner and Tikunoff instrument listed in appendix 1. In brief form the items and means are listed here.

A. Teacher makes a statement whose consequences would be ridiculous if carried out. Pr. Ed. = 2.12 Parents = 1.33

C. Teacher displays unanticipated switching of activity, e.g., from instruction to classroom management to behavior management to instruction, etc... Pr. Ed. = 3.17 Parents = 2.11

H. Teacher treats the whole group as one, often to maintain

group control. Pr. Ed. = 3.11 Parents = 2.52

I. Teacher capitalizes instructionally on unexpected incidents that arise during class time. Pr. Ed. = 3.95 Parents = 2.93

In every case parents were rating the questioned behavior lower than the professional educators. This could be because educators are more lenient or understanding with some of their behavior. These items could be indicative that educators are seeking goals which are more conducive to the smooth running of the system (real goals) rather than the best education for the students (stated goals).

In the push toward individualized instruction parents are still ahead of professional educators if that is indeed what item H is indicating, wanting their children to be treated as an individual rather than as a group. One could almost expect this difference as parents see only the individual child and do not have to deal with the group.

Items C and I are somewhat related in that they both deal with teaching by using natural occurrences and happenings rather than artificially creating the learning experience. For example, if an elephant came by the class this could make an excellent teaching aid. But, if the clock shows that it is 1:30 and time for

zoology, the students might not have the desire just then to study elephants. As a result the teacher might be forced to resort to artificial methods of teaching about elephants such as books or films. Parents appear to like the more structured method of artificial material. Whereas the educators more readily see the need to capitalize on using what becomes available, even at the cost of switching in the middle of something else. There is no contradiction between these two items because they are related significantly with a positive correlation. One could expect that if parents would rate lower on one they would also rate lower on the other. (Correlation = +.24 and the .0005 significance level) Thus these two items are related.

2. New Teachers vs. Older Teachers

Teachers who have five or less years of experience (new teachers) were different from older teachers in five (5) different categories: items F, H, J, T, and U. (See table 10)

These differences like all differences reported are significant according to the Mann-Whitney U test which performed the Z tests. All five of these differences had Z scores lower than -1.96.

F. Teacher scolds child in front of others. <5 years = 2.87 >5 years = 2.30.

H. Teacher treats whole group as one often to maintain group control. <5 years = 3.42 >5 years = 2.90

J. Teacher reacts constructively to students' feelings and attitudes. <5 years = 4.57 >5 years = 4.74

T. Teacher checks on student progress regularly and adjusts instruction accordingly. <5 years = 4.65 >5 years = 4.90

U. Teacher calls attention to self for no apparent instructional purpose. <5 years = 1.94 >5 years = 1.65

Probably item H is the most interesting one here in this case because there was also a significant difference on that item between professional educators and parents. It can be noted that the older teachers come closer to siding with the parents than any other group. The newer teachers were the farthest from the parents with the principals coming in between at 3.05. One must raise the question, "Why is there so much difference between newer teachers and parents on the issue of treating the whole group as one?" A possible explanation is that newer teachers who struggle to maintain authority in the classroom, resort to treating the whole group as one; while the more experienced teachers are able to develop other means of maintaining control.

Newer teachers gave a higher rating to scolding a child in front

of others than the older teachers. This, too, could be indicative that older teachers have refined their methods of control through their years of experience. A further topic of research is raised in this issue concerning the value of experience in teaching. What value is there in years of teaching experience? These items point to some possibilities.

Although both groups gave low ratings to a teacher calling useless attention to self, newer teachers were less likely to think anything wrong with that. A possible explanation is that older teachers have developed other means to obtain the children's attention. This item in particular, but other items as well, (F and H), could have further implications. Newer teachers do not have tenure and in many cases are evaluated more than tenured teachers. They could possibly see evaluations as more of a threat since they do not have the security of tenure. As a result they seek to be less critical of certain teacher behavior patterns which they might have a tendency to resort to. On all 3.00 of these items that were below the midpoint of three the younger teachers were less critical of the questioned behavior.

On items J and T, the positively rated items, the newer teachers gave less of a positive rating than the older teachers. Again, it is as if they are less sure of themselves and, as a group, do not want to commit themselves too much on points where they could be evaluated.

The 4.90 rating by older teachers on item T was the highest mean rating of any group on any item. It is interesting that this is an item measuring extrinsic methods. This item shows that older teachers place a great deal of emphasis on the output from each child as well as the intrinsic values each child encounters. As Verble's research showed, it does take a combination of both intrinsic and extrinsic methods to teach effectively and evaluation of such teaching should include items for evaluating both. These results could be evidence that older teachers have come to realize this to a greater extent than the newer teachers.

3. New Principals vs. Older Principals

Not much emphasis should be placed upon the results of these groups as the N of newer principals was only 4. Years of experience is often required to enter this field thus the size of these two groups is biased. Even though three items were shown to have significant differences the results could have easily been different if even one newer principal would change his responses (see table 13).

The three items that differed were: B, P, and S.

B. Teacher expresses positive, pleasant, optimistic attitudes and feelings. <5 years = 3.50 >5 years = 4.83

P. There is a warm and family-like quality to classroom.

interaction and good feelings between teacher and students and between students and students. <5 years = 4.00 >5 years = 4.49

S. Students express eagerness to participate, appear actively and productively involved in learning activities. <5 years = 3.50 >5 years = 4.54

The reader should note that all three of these items measure intrinsic values and in every case the older principals placed more emphasis on the intrinsic aspect. Newer principals appear to be more concrete, looking to evaluate more on items easier to objectively record.

4. Teachers vs. Principals

The Kansas teachers and principals that were compared by Grant and Carvell showed no significant differences on any item. In Eastern Kentucky the same two groups had four points of difference: Items D, N, R, and S. (see table 14)

D. Teacher seems confident in teaching a given subject and demonstrates a grasp of it. T = 4.84 P = 4.59

N. Teacher encourages students to take responsibility for their own classwork. T = 4.61 P = 4.33

R. Teacher actively listens to what students are saying, reading or reciting. T = 4.75 P = 4.44

S. Students express eagerness to participate, appear actively and productively involved in learning activities. T = 4.71 P = 4.44

All responses, even though significantly different, are still at the positive end of the scale showing that both teachers and principals approve of these items as valid teaching evaluation criteria but in every case teacher more heartily approved with a higher positive rating.

Three out of four, N, R, and S, are dealing directly with the students, focusing on them as the primary purpose of education. The teachers acknowledge this fact more than the principals by rating these items higher. In contrast to the principal who travels between classrooms, the teacher, who spends much time with the same children, apparently becomes more conscious of the actual and specific needs of the children.

The evidence suggests that older teachers are more student oriented reinforces the idea that years of teaching will make the teacher more child centered in philosophy. The longer one teaches the more true this becomes.

On item S in particular, teachers did differ significantly with

principals but it is interesting to note that the principals differed themselves on this item with the newer principals even lower than the older principals. This lower score shows that newer principals are significantly lower in their opinion of seeing students eager to learn. In that case it would mean newer principals are not as concerned with one of the stated goals of schools.

Answering Questions

Question number 1 asked, "Do teachers and principals embrace similar philosophies regarding what constitutes desirable teaching behavior?" In examining the only four significant differences between teacher and principals both sides evaluated the items on the positive end of the scale showing that they do agree that the items are valid teaching evaluation criteria, they only differ in the degree to which they support these items. Why they differ is left open to debate but principals should be made aware that teachers do tend to place a great deal of emphasis on the actual student aspects of teaching.

"Does the number of years one spends in education affect the results of the survey?" According to this survey it appears there are several items that experience will change in teachers. Teachers develop a growing awareness of student needs the longer they teach and seek to gear their instruction more around the student rather

than just teach for the sake of teaching. Older teachers also appear more confident of themselves and/or their teaching methods than newer teachers as is indicated by the stronger stand they take on lower ratings of negative behavior. Younger teachers can look with hope into the years that the evaluation process will become less of a threat due to their increasing teaching skills which help them better cope with the problems they encounter.

"Do parents agree with professional educators as to what constitutes desired teaching behavior?" In most cases yes, but there is a pattern that can be seen in the few items they do differ on. Parents seek continuity in the classroom and behavior that is deliberate and expected. Teaching that is vague and hard to follow is undesired more so with parents. Teachers can learn from this that parents want clear and understandable directions. Parents also expect their child to be treated as an individual rather than as a group. This can be related back to the other points in that this way the child is sure to understand his responsibilities.

"Do East Kentucky professional educators differ from their counterparts in the Midwest?" Statistically there is no way to prove this without the data cards from the Grant and Carvell study, but for the sake of visual examination of the means, Table 16 was prepared. There are some relatively large differences in the means which might be significant. A general rule of thumb might be those means differing more than .5 probably are significant.

"What patterns of significance are there?" These have already been discussed where they are applicable.

"How are these twenty-one items clustered, and if they are, could they be grouped together to make a smaller instrument?" Some differences pointed to the need of an area for student centered evaluation criteria. Items F, G, H, J, K, M, N, P, Q, R, S, and T already focus on the student but with some rewording these twelve items could be combined into a smaller number of items.

Some items can be clustered together that focus on extrinsic or intrinsic evaluations. Items G, M, N, R, S, and T all are concerned with student output making them accessible to extrinsic teacher evaluation methods. What they all condense into is the question, "How well are the students learning?" Perhaps a standardized test of student achievement would serve this purpose as well. It would surely be more objective. The intrinsic items like B, D, J, and P measure the affective climate of the classroom. They are making sure that the teacher will not neglect the non-academic aspects of the children such as physical, emotional, and social needs. The wording of an all-inclusive item like this will not enter into the scope of this paper. It might even take several items if they cannot all condensed into one. Wordings which should be avoided are those of items C and H where the means tend to be close to the middle ground of three. Items like that are seen as neither positive or negative criteria by which to evaluate a teacher.

Responding to the Null Hypotheses

In answering these questions, clues have been given to the proper responses of the null hypotheses. The first null, there is no difference between teachers and principals concerning their philosophy of desired teaching behavior in the classroom as measured by the Berliner and Tikunoff scale, needs to be rejected on those four items previously discussed (D, N, R, and S), at least in Eastern Kentucky. On the other seventeen items it would not be wise to reject the null.

The second hypothesis stated, "The number of years a teacher or principal spends in education does not affect the results of the survey." This too needs to be rejected on at least five accounts for the teachers and possibly three with the principals. Teachers do change over the years in their attitudes as do their philosophies of education.

The same must be said for the third null hypothesis, because it, too, had four points upon which professional educators and parents differed on. This is not to say these or any other differences reported here constitute a disagreement but merely a difference in the degree to which the group would place emphasis on a particular item. One of the main conclusions that can be drawn from this research is that all respondents agree upon what constitutes valid teaching evaluation criteria. They only

occasionally differ as to the degree of support for an item. In no case does one group rate an item high and another group rate it low. The differences occur on the same side of the scale.

The last hypothesis will remain unanswered. "There is no significant difference between East Kentucky educators and their counterparts in the Midwest." The scope of this study's research collection was limited as the data from the Midwest was unattainable. The decision to reject or accept will be left to the reader as he compares the means of each group.

CHAPTER V

Summary

This study examined teacher evaluation criteria in an East Kentucky setting. Using the Berliner and Tikunoff instrument, which a review of literature showed had a strong foundation, teachers, principals, and parents were surveyed as to their professional opinions on the twenty-one items of teacher evaluation criteria. All of the respondents came from five counties in Eastern Kentucky: Floyd, Johnson, Lawrence, Martin, and Pike. Teachers and principals were limited to those who serve on the elementary level and the parents were those who had children in the elementary grades. One variable which divided the educators was the number of years they have been employed in education. Those with five or less years of experience could then be compared to those with more than five years experience to see if longevity affects one's attitudes. This made a total of five categories that a respondent could be placed into: new teachers, older teachers, new principals, older principals, or parents. From these five categories, four comparisons were made: new teachers vs. older teachers, new principals vs. older principals, teachers vs. principals, and professional educators vs. parents. Every comparison produced from three to five significant differences.

Comparing new and older teacher data, experience appeared to

make the teacher more decisive in rating negative and positive behavior, thus implying that the older teacher has more confidence in the evaluation process. Older teachers also saw greater need to check on student progress which could indicate that they place more value on student output and extrinsic evaluations. The difference between teachers shows their thoughts on the safety of tenure.

In examining the differences between principals with more than five years experience and those with less, a few significant differences were found but they should not be considered with any great value as the newer principal group only had an N of four and would not be as reliable as it would be with more respondents.

Teachers differed from principals on four points of teacher evaluation criteria. In every case the teachers gave higher ratings to all four items placing the emphasis on making sure the student is learning. Here teachers appear to be more "product" oriented than the principals. Carrying this idea to the extreme it might show that principals lean more in the direction of running an efficient school by working on the real goals rather than the stated goals which are given to the public.

Parents differed in four areas with professional educators. These items were centered around the type of instruction a teacher gives as to whether or not spontaneity is appropriate. Parents

leaned toward a well defined and structured environment to learn in, while the teachers saw a need for using happenings and occurrences that arise naturally for teaching experiences. Parents also indicated they wanted their child treated with more individuality than collectively as compared to what the professional educators thought necessary.

Conclusions

The purpose of this study was to determine whether there were any significant differences between principals, teachers, and/or parents concerning what constitutes desirable and undesirable teaching behaviors and techniques. Based on the results of the survey of thirty-nine elementary school principals, one hundred and thirty-nine teachers, and twenty-seven parents it was concluded that there was agreement among the groups concerning what constitutes both desirable and undesirable teaching behaviors.

This study had a total of sixteen significant differences of opinion among the noted categories, but as earlier stated it must be remembered that the groups do not necessarily disagree with each other but rather that they differ on the amount of emphasis to place upon certain items of teacher evaluation criteria. These differences should be kept in consideration when the evaluation process is going on. This idea is supported by the review of the literature which shows that the evaluation process benefits by an

exchange of ideas concerning what and how a teacher is to be evaluated.

A conclusion that can be drawn from the review of literature is that parents have for too long been left out of the evaluation process.

This study indicates that parents and professional educators perceive good teaching in similar ways. This is not to say educators or parents have all the answers but that they can agree and this agreement should be built upon. The schools would be wise to show the parents that they too are interested in the same goals and objectives as they seek. By showing parents that their goals are similar to the school's goals, the bonds of community relations can only be strengthened. Concerning the four items which showed a discrepancy, a principal would be wise to show that these are points of interest and concern.

Lastly, the reader can conclude that there is definitely a solid foundation of criteria available here for teacher evaluations that is acceptable to all concerned parties. This study confirms other studies that have shown administrators need not be arbitrary in their evaluations. By the use of available material that is acceptable to principals, teacher, and parents, principals can evaluate on valid criteria.

Recommendations

In addition to answering several questions, this study also raises questions and suggests that additional research on the topic is desirable. A central question is the need to explain the few differences that do exist. Are they local differences or are they universal? Why do parents differ from educators on certain items? Why does experience make a difference between teacher's professional opinions? How can administrators help newer teachers compensate for their lack of experience, since this study shows they do lack confidence? Do Kentucky or Kansas results apply to other areas of the country? Should attempts be made to minimize the existing differences or just to understand them?

This data also suggests that further investigations concerning concurrence of principals, teachers, and parents on criteria for teacher evaluation is warranted. Finally, this data suggests that there is a common core of behaviors on which all parties can agree.

For all practical purposes the next step to this study is actual implementation. Aside from the suggestions already made, these main ideas covered in the Berliner and Tikunoff instrument could be grouped together by areas they cover, and then the ambiguous items removed. A program of continual and regular idea exchange should be made involving the principal and each teacher concerning teacher behavior. Although it is often hard to get many parents

involved in school matters principals should make an effort to solicit input from parents concerning what they expect. It should not be recommended that these implementations would be as valid for an area outside of East Kentucky until a similar pilot study was completed.

APPENDIX 1

THIS IS A SURVEY GIVEN TO DETERMINE WHETHER OR NOT THE FOLLOWING ITEMS ARE CONSIDERED VALID TEACHER EVALUATION CRITERIA. PLEASE INDICATE YOUR PROFESSIONAL OPINION BY CIRCLING 5 ON THE ITEMS THAT ARE MOST IMPORTANT OR DESIRABLE AND 1 ON THOSE LEAST IMPORTANT OR UNDESIRABLE.

1. TEACHER MAKES A STATEMENT WHOSE CONSEQUENCES WOULD BE RIDICULOUS IF CARRIED OUT.
1 2 3 4 5
2. TEACHER EXPRESSES POSITIVE, PLEASANT, OPTIMISTIC ATTITUDES AND FEELINGS.
1 2 3 4 5
3. TEACHER DISPLAYS UNANTICIPATED SWITCHING OF ACTIVITY, e.g., FROM INSTRUCTION TO CLASSROOM MANAGEMENT TO BEHAVIOR MANAGEMENT TO INSTRUCTION, ETC.
1 2 3 4 5
4. TEACHER SEEMS CONFIDENT IN TEACHING A GIVEN SUBJECT AND DEMONSTRATES GRASP OF IT.
1 2 3 4 5
5. TEACHER GIVES DIRECTION OF THREAT AND FOLLOWS THROUGH WITH IT.
1 2 3 4 5
6. TEACHER SCOLDS CHILD IN FRONT OF OTHERS.
1 2 3 4 5
7. TEACHER SEEMS TO PERCEIVE LEARNING RATE OF STUDENTS AND ADJUSTS TEACHING PACE ACCORDINGLY.
1 2 3 4 5
8. TEACHER TREATS WHOLE GROUP AS ONE, OFTEN TO MAINTAIN GROUP CONTROL.
1 2 3 4 5
9. TEACHER CAPITALIZES INSTRUCTIONALLY ON UNEXPECTED INCIDENTS THAT ARISE DURING CLASS TIME.
1 2 3 4 5
10. TEACHER REACTS CONSTRUCTIVELY TO STUDENTS' FEELINGS AND ATTITUDES.
1 2 3 4 5
11. STUDENTS COOPERATE WITH OTHER STUDENTS AND WITH THE TEACHER.
1 2 3 4 5

12. ADULTS OTHER THAN TEACHER ARE ALLOWED TO INSTRUCT.
1 2 3 4 5
13. TEACHER PREPARED STUDENTS FOR LESSON BY REVIEWING, OUTLINING,
EXPLAINING OBJECTIVES AND SUMMARIZING.
1 2 3 4 5
14. TEACHER ENCOURAGES STUDENTS TO TAKE RESPONSIBILITY FOR THEIR
OWN CLASSWORK.
1 2 3 4 5
15. TEACHER FILLS EMPTY TIME PERIODS WITH BUSY WORK.
1 2 3 4 5
16. THERE IS A WARM AND FAMILY-LIKE QUALITY TO CLASSROOM INTERACTION
AND GOOD FEELINGS BETWEEN TEACHER AND STUDENTS AND BETWEEN
STUDENTS AND STUDENTS.
1 2 3 4 5
17. A STUDENT OPENLY RESISTS TEACHER DIRECTION AND REFUSES TO
COMPLY.
1 2 3 4 5
18. TEACHER ACTIVELY LISTENS TO WHAT STUDENT IS SAYING, READING,
OR RECITING.
1 2 3 4 5
19. STUDENTS EXPRESS EAGERNESS TO PARTICIPATE, APPEAR ACTIVELY
AND PRODUCTIVELY INVOLVED IN LEARNING ACTIVITIES.
1 2 3 4 5
20. TEACHER CHECKS ON STUDENT PROGRESS REGULARLY AND ADJUSTS
INSTRUCTION ACCORDINGLY.
1 2 3 4 5
21. TEACHER CALLS ATTENTION TO SELF FOR NO APPARENT INSTRUCTIONAL
PURPOSE.
1 2 3 4 5

ARE YOU AN ELEMENTARY PRINCIPAL? _____
 ARE YOU AN ELEMENTARY TEACHER? _____
 HOW MANY YEARS HAVE YOU BEEN EMPLOYED IN EDUCATION? _____
 ARE YOU THE PARENT OF A CHILD IN GRADES K-6? _____

TABLE 1
 DESCRIPTIVE STATISTICS
 ALL RESPONDENTS (205)

	TEACHERS	PRINCIPALS	PARENTS
< 5 years	62	35	
			27
> 5 years	77	4	
	139	39	27

TABLE 2
DESCRIPTIVE STATISTICS
ALL RESPONDENTS (205)

	MEANS	SD	SE	SKEW	KURTOSIS
A	2.02	1.39	.097	1.06 *	-0.25
B	4.70	0.81	.056	-3.22 *	10.56 *
C	3.02	1.49	.104	-0.09	-1.32 *
D	4.77	0.79	.055	-3.96 *	15.35 *
E	4.19	1.14	.079	-1.48 *	1.52 *
F	2.52	1.39	.097	0.39 *	-1.05 *
G	4.63	0.87	.061	-3.01 *	9.21 *
H	3.03	1.43	.100	-0.03	-1.24 *
I	3.82	1.27	.088	-0.90 *	-0.17
J	4.61	0.87	.061	-2.91 *	8.70 *
K	4.27	1.19	.083	-1.50 *	1.03 *
L	2.52	1.39	.097	0.41 *	-1.06 *
M	4.51	1.03	.072	-2.25 *	4.21 *
N	4.49	1.11	.078	-2.38 *	4.63 *
O	2.47	1.42	.099	0.56 *	-0.98 *
P	4.49	0.98	.069	-2.26 *	4.71 *
Q	2.56	1.61	.112	0.45 *	-1.37 *
R	4.66	0.87	.061	-3.03 *	8.81 *
S	4.63	0.82	.057	-2.53 *	6.10 *
T	4.72	0.83	.058	-3.54 *	12.35 *
U	1.77	1.20	.084	1.46 *	1.09 *

IF SKEWNESS = /.34/ THEN Z = /1.96/

IF KURTOSIS = /.67/ THEN Z = /1.96/

TABLE 3
DESCRIPTIVE STATISTICS

ALL TEACHER RESPONDENTS (139)

	MEANS	SD	SE	SKEW	KURTOSIS
A	2.08	1.42	0.12	0.98 *	-0.42
B	4.74	0.76	0.06	-3.83 *	15.72 *
C	3.15	1.43	0.12	-0.24	-1.17 *
D	4.84	0.69	0.06	-5.05 *	25.42 *
E	4.17	1.13	0.10	-1.38 *	1.29 *
F	2.55	1.38	0.12	0.36	-0.99 *
G	4.68	0.75	0.06	-3.55 *	14.52 *
H	3.13	1.37	0.12	-0.19	-1.05 *
I	3.99	1.20	0.10	-1.15 *	0.52
J	4.66	0.75	0.06	-3.41 *	13.67 *
K	4.27	1.17	0.10	-1.39 *	0.60
L	2.45	1.34	0.11	0.40	-0.96 *
M	4.55	0.89	0.08	-2.20 *	4.36 *
N	4.61	1.01	0.09	-3.00 *	8.07 *
O	2.40	1.36	0.12	0.63 *	0.78
P	4.53	0.89	0.08	-2.35 *	5.75 *
Q	2.70	1.60	0.14	0.28	-1.43 *
R	4.75	0.76	0.07	-3.92 *	16.03 *
S	4.71	0.70	0.06	-2.66 *	6.81 *
T	4.78	0.66	0.06	-4.10 *	18.76 *
U	1.78	0.50	0.06	1.35 *	0.76

IF SKEWNESS = /.41/ THEN Z = /1.96/

IF KURTOSIS = /.81/ THEN Z = /1.96/

TABLE 4
DESCRIPTIVE STATISTICS
ALL PRINCIPAL RESPONDENTS (39)

	MEANS	SD	SE	SKEW	KURTOSIS
A	2.28	1.41	0.23	0.83	-0.49
B	4.69	0.66	0.11	-1.96 *	2.44 *
C	3.21	1.66	0.26	-0.09	-1.65 *
D	4.59	0.94	0.15	-3.10 *	10.04 *
E	4.13	1.30	0.21	-1.46 *	1.11
F	2.64	1.25	0.20	0.31	-0.94
G	4.59	0.99	0.16	-2.82 *	7.74 *
H	3.05	1.45	0.23	0.18	-1.31
I	3.82	1.12	0.18	-0.81 *	0.23
J	4.49	1.00	0.16	-2.49 *	6.37 *
K	4.23	1.11	0.18	-1.46 *	1.69 *
L	2.80	1.59	0.26	0.31	-1.51
M	4.44	1.23	0.20	-2.16 *	3.35 *
N	4.33	1.03	0.17	-1.93 *	3.86 *
O	2.49	1.49	0.24	0.61	-0.97
P	4.44	1.00	0.16	-2.35 *	5.89 *
Q	2.33	1.51	0.24	0.75	0.95
R	4.44	1.05	0.17	-2.15 *	4.43 *
S	4.44	1.00	0.16	-2.35 *	5.89 *
T	4.59	0.99	0.16	-2.82 *	
U	1.87	1.30	0.21	1.46 *	1.11

IF SKEWNESS = /.77/ THEN Z = /1.96/

IF KURTOSIS = /1.54/ THEN Z = /1.96/

TABLE 5
 DESCRIPTIVE STATISTICS
 ALL PARENT RESPONDENTS (27)

	MEANS	SD	SE	SKEW	KURTOSIS
A	1.33	0.96	0.19	2.62 *	5.27 *
B	4.48	1.19	0.23	-2.10 *	2.99 *
C	2.11	1.22	0.24	0.32	-1.67
D	4.67	0.96	0.19	-2.62 *	5.27 *
E	4.44	0.93	0.18	-2.28 *	6.27 *
F	2.15	1.61	0.31	0.87	-0.99
G	4.41	1.19	0.23	-1.93 *	2.47 *
H	2.52	1.67	0.32	0.58	-1.34
I	2.93	1.44	0.28	0.06	-1.29
J	4.48	1.19	0.23	-2.10 *	2.99 *
K	4.43	1.44	0.28	-1.97 *	2.22 *
L	2.44	1.29	0.25	0.37	-1.24
M	4.41	1.34	0.26	-2.08 *	2.90 *
N	4.07	1.54	0.30	-1.29 *	-0.01
O	2.78	1.63	0.31	0.16	-1.60
P	4.37	1.37	0.26	-1.82 *	1.62
Q	2.15	1.75	0.34	1.06 *	-0.82
R	4.56	1.09	0.21	2.10 *	2.60 *
S	4.52	1.09	0.21	-1.99 *	2.24 *
T	4.56	1.28	0.25	-2.62 *	5.27 *
U	1.59	1.34	0.26	2.07 *	2.90 *

IF SKEWNESS = /.92/ THEN Z = /1.96/

IF KURTOSIS = /1.85/ THEN Z = /1.96/

TABLE 6
DESCRIPTIVE STATISTICS

ALL PROFESSIONAL EDUCATORS RESPONDENTS (178)

	MEANS	SD	SE	SKEW	KURTOSIS
A	2.12	1.42	0.11	0.94 *	-0.48
B	4.73	0.73	0.06	-3.54 *	13.87 *
C	3.17	1.48	0.11	-0.19	-1.29 *
D	4.79	0.76	0.06	-4.33 *	18.80 *
E	4.16	1.16	0.09	-1.40 *	1.22 *
F	2.57	1.35	0.10	0.34	0.98 *
G	4.66	0.81	0.06	-3.33 *	12.02 *
H	3.11	1.38	0.10	-0.10	-1.13 *
I	3.95	1.18	0.09	-1.07 *	0.39
J	4.62	0.82	0.06	-3.12 *	10.90 *
K	4.26	1.15	0.09	-1.39 *	0.74 *
L	2.53	1.40	0.11	0.41 *	-1.06 *
M	4.53	0.98	0.07	-2.26 *	4.40 *
N	4.55	1.02	0.08	-2.69 *	6.57 *
O	2.42	1.38	0.10	0.63 *	-0.83 *
P	4.51	0.92	0.07	-2.34 *	5.64 *
Q	2.62	1.58	0.12	0.37 *	-1.38 *
R	4.68	0.84	0.06	-3.28 *	10.95 *
S	4.65	0.78	0.06	-2.66 *	7.40 *
T	4.74	0.75	0.06	-3.70 *	14.49 *
U	1.80	1.18	0.09	1.37 *	0.92 *

IF SKEWNESS = /.36/ THEN Z = /1.96/

IF KURTOSIS = /.72/ THEN Z = /1.96/

TABLE 7

INFERENCEAL STATISTICS

PROFESSIONAL EDUCATORS (178) VS. PARENTS (27)

	MEANS	MEANS	Z-SCORES
A	2.12	1.33	-3.18 *
B	4.73	4.48	-0.43
C	3.17	2.11	-3.41 *
D	4.79	4.69	-0.06
E	4.16	4.44	-1.11
F	2.57	2.15	-1.77
G	4.66	4.41	-0.62
H	3.11	2.52	-1.99 *
I	3.95	2.93	-3.57 *
J	4.62	4.48	-0.55
K	4.26	4.43	-1.01
L	2.53	2.44	-0.14
M	4.53	4.41	-0.42
N	4.55	4.07	-1.06
O	2.42	2.78	-0.96
P	4.51	4.37	-0.86
Q	2.62	2.15	-1.66
R	4.68	4.56	-0.26
S	4.65	4.52	-0.24
T	4.74	4.56	-0.42
U	1.80	1.59	-1.63

*SIGNIFICANT AT THE .05 LEVEL

TABLE 8
 DESCRIPTIVE STATISTICS
 NEW TEACHER RESPONDENTS (62)

	MEANS	SD	SE	SKEW	KURTOSIS
A	2.07	1.23	0.16	0.76 *	0.69
B	4.77	0.76	0.10	-4.31 *	19.48 *
C	3.21	1.38	0.16	-0.43	-1.06
D	4.81	0.74	0.09	-4.63 *	21.87 *
E	4.26	1.06	0.13	-1.41 *	1.43 *
F	2.87	1.30	0.17	0.06	-0.86
G	4.61	0.80	0.10	-3.20 *	12.30 *
H	3.42	1.25	0.16	-0.33	-0.68
I	4.02	0.93	0.12	-0.54	-0.69
J	4.57	0.80	0.10	-2.98 *	11.11 *
K	4.21	1.10	0.14	-1.04 *	-0.44
L	2.53	1.22	0.16	0.14	-0.84
M	4.52	0.99	0.13	-2.26 *	4.65 *
N	4.55	1.11	0.14	-2.72 *	6.28 *
O	2.39	1.23	0.16	0.51	-0.70
P	4.45	0.92	0.12	-2.09 *	4.88 *
Q	2.90	1.48	0.19	0.05	-1.30 *
R	4.66	0.85	0.11	-3.28 *	11.25 *
S	4.60	0.82	0.10	-2.08 *	3.47 *
T	4.65	0.79	0.10	-3.38 *	13.28 *
U	1.94	1.11	0.14	0.87 *	-0.10

IF SKEWNESS = /.61/ THEN Z = /1.96/

IF KURTOSIS = /1.22/ THEN Z = /1.96/

TABLE 9
DESCRIPTIVE STATISTICS
OLDER TEACHER RESPONDENTS (77)

	MEANS	SD	SE	SKEW		KURTOSIS
A	2.09	1.56	0.18	1.05	*	-0.52
B	4.71	0.76	0.09	-3.55	*	14.13 *
C	3.10	1.47	0.17	-0.11		-1.21 *
D	4.87	0.66	0.08	-5.60	*	31.47 *
E	4.09	1.18	0.14	-1.36	*	1.20 *
F	2.30	1.41	0.16	0.67	*	-0.73
G	4.74	0.72	0.08	-4.01	*	18.47 *
H	2.90	1.42	0.16	-0.01		-1.23 *
I	3.96	1.38	0.16	-1.22	*	0.21
J	4.74	0.72	0.08	-4.01	*	18.47 *
K	4.33	1.22	0.14	-1.64	*	1.36 *
L	2.39	1.43	0.16	0.57	*	-0.99
M	4.58	0.82	0.09	-2.06	*	3.49 *
N	4.66	0.93	0.11	-3.35	*	10.84 *
O	2.40	1.45	0.17	0.69	*	-0.89
P	4.58	0.88	0.10	-2.65	*	7.27 *
Q	2.53	1.68	0.19	0.49		-1.41 *
R	4.82	0.68	0.08	-4.84	*	24.87 *
S	4.79	0.57	0.07	-3.50	*	13.85 *
T	4.90	0.50	0.06	-5.33	*	28.66 *
U	1.65	1.16	0.13	1.78	*	2.11 *

IF SKEWNESS = /.55/ THEN Z = /1.96/

IF KURTOSIS = /1.09/ THEN Z = /1.96/

TABLE 10

INFERENCEAL STATISTICS

NEW TEACHERS (62) VS. OLDER TEACHERS (77)

	MEANS	MEANS	Z-SCORES
A	2.07	2.09	-0.64
B	4.77	4.71	-0.84
C	3.21	3.10	-0.51
D	4.81	4.87	-0.99
E	4.26	4.09	-0.83
F	2.87	2.30	-2.63 *
G	4.61	4.74	-1.48
H	3.42	2.90	-2.12 *
I	4.02	3.96	-0.84
J	4.57	4.74	-2.06 *
K	4.21	4.33	-1.13
L	2.53	2.39	-0.87
M	4.52	4.58	-0.09
N	4.55	4.66	-0.46
O	2.39	2.40	-0.31
P	4.45	4.58	-1.22
Q	2.90	2.53	-1.54
R	4.66	4.82	-1.71
S	4.60	4.79	-1.41
T	4.65	4.90	-3.35 *
U	1.94	1.65	-1.96 *

*SIGNIFICANT AT THE .05 LEVEL

TABLE 11
 DESCRIPTIVE STATISTICS
 NEW PRINCIPAL RESPONDENTS (4)

	MEANS	SD	SE	SKEW	KURTOSIS
A	2.00	1.16	0.58	0	-6.00 *
B	3.50	0.58	0.29	0	-6.00 *
C	4.00	1.16	0.58	0	-6.00 *
D	5.00	0.00	0.00	-	----
E	4.50	0.58	0.29	0	-6.00 *
F	2.50	0.58	0.29	0	-6.00 *
G	4.50	0.58	0.29	0	-6.00 *
H	3.50	1.73	0.87	0	-6.00 *
I	3.50	0.58	0.87	0	-6.00 *
J	4.50	0.58	0.29	0	-6.00 *
K	4.00	1.16	0.58	0	-6.00 *
L	2.00	0.00	0.00	-	----
M	5.00	0.00	0.00	-	----
N	4.00	1.16	0.58	0	-6.00 *
O	3.50	1.73	0.87	0	-6.00 *
P	4.00	0.00	0.00	-	-6.00 *
Q	2.00	1.16	0.58	0	-6.00 *
R	4.00	1.16	0.58	0	-6.00 *
S	3.50	0.58	0.29	0	-6.00 *
T	4.50	0.58	0.29	0	-6.00 *
U	1.50	0.58	0.29	0	-6.00 *

IF SKEWNESS = /2.40/ THEN Z = /1.96/

IF KURTOSIS = /4.80/ THEN Z = /1.96/

TABLE 12

DESCRIPTIVE STATISTICS

OLDER PRINCIPAL RESPONDENTS (35)

	MEANS	SD	SE	SKEW		KURTOSIS
A	2.31	1.47	0.25	0.83	*	-0.59
B	4.83	0.51	0.09	-3.04	*	8.48 *
C	3.11	1.70	0.29	0.00		-1.72 *
D	4.54	0.98	0.17	-2.91	*	8.75 *
E	4.09	1.36	0.23	-1.36	*	0.69
F	2.66	1.31	0.22	0.27		-1.12
G	4.60	1.04	0.18	-2.82	*	7.43 *
H	3.00	1.44	0.24	0.19		-1.21
I	3.86	1.17	0.20	0.89	*	0.19
J	4.49	1.04	0.18	-2.46	*	5.94 *
K	4.26	1.12	0.19	-1.61	*	2.20 *
L	4.89	1.66	0.28	0.15		-1.67 *
M	4.37	1.29	0.22	-1.99	*	2.57 *
N	4.37	1.03	0.17	-2.19	*	5.06 *
O	2.37	1.44	0.24	0.68		-0.78
P	4.49	1.04	0.18	-2.46	*	5.94 *
Q	2.37	1.56	0.26	0.73		-1.07
R	4.49	1.04	0.18	-2.46	*	5.94 *
S	4.54	0.98	0.17	-2.91	*	8.75 *
T	4.60	1.04	0.18	-2.82	*	7.43 *
U	1.91	1.39	0.23	1.36	*	0.69

IF SKEWNESS = /.81/ THEN Z = /1.96/

IF KURTOSIS = /1.62/ THEN Z = /1.96/

TABLE 13
INFERENTIAL STATISTICS

NEW PRINCIPALS (4) VS. OLDER PRINCIPALS (35)

	MEANS	MEANS	Z-SCORES
A	2.00	2.31	-0.24
B	3.50	4.83	-4.08 *
C	4.00	3.11	-1.07
D	5.00	4.54	-1.22
E	4.50	4.09	-0.16
F	2.50	2.66	-0.10
G	4.50	4.60	-1.25
H	3.50	3.00	-0.53
I	3.50	3.86	-0.97
J	4.50	4.49	-0.63
K	4.00	4.26	-0.58
L	2.00	2.89	-0.67
M	5.00	4.37	-1.13
N	4.00	4.37	-0.79
O	3.50	2.37	-1.39
P	4.00	4.49	-2.29 *
Q	2.00	2.37	-0.34
R	4.00	4.49	-1.08
S	3.50	4.54	-2.95 *
T	4.50	4.60	-1.25
U	1.50	1.91	-0.16

*SIGNIFICANT AT THE .05 LEVEL

TABLE 14

INFERENCEAL STATISTICS

ALL TEACHERS (139) VS. ALL PRINCIPALS (39)

	MEANS	MEANS	Z-SCORES
A	2.08	2.28	-1.01
B	4.74	4.69	-0.75
C	3.15	3.21	-0.30
D	4.84	4.59	-3.15 *
E	4.17	4.13	-0.26
F	2.55	2.64	-0.50
G	4.68	4.59	-0.12
H	3.13	3.05	-0.43
I	3.99	3.82	-1.15
J	4.66	4.49	-0.89
K	4.27	4.23	-0.54
L	2.45	2.80	-1.12
M	4.55	4.44	-0.15
N	4.61	4.33	-2.71 *
O	2.40	2.49	-0.21
P	4.53	4.44	-0.62
Q	2.70	2.33	-1.19
R	4.75	4.44	-2.33 *
S	4.71	4.44	-2.11
T	4.78	4.59	-1.05
U	1.78	1.87	-0.26

*SIGNIFICANT AT THE .05 LEVEL

TABLE 15

DESCRIPTIVE STATISTICS

A COMPARISON OF ALL FIVE GROUP MEANS (205)

	< 5 T (62)	> 5 T (77)	< 5 P (4)	> 5 P (35)	PARENTS (27)
A	2.07	2.09	2.00	2.31	1.33
B	4.77	4.71	3.50	4.83	4.48
C	3.21	3.10	4.00	3.11	2.11
D	4.81	4.87	5.00	4.54	4.67
E	4.26	4.09	4.50	4.09	4.44
F	2.87	2.30	2.50	2.66	2.15
G	4.61	4.74	4.50	4.60	4.41
H	3.42	2.90	3.50	3.00	2.52
I	4.02	3.96	3.50	3.86	2.93
J	4.57	4.74	4.50	4.49	4.48
K	4.21	4.33	4.00	4.26	4.43
L	2.53	2.39	2.00	2.89	2.44
M	4.52	4.58	4.00	4.37	4.41
N	4.55	4.66	4.00	4.37	4.07
O	2.39	2.40	3.50	2.37	2.78
P	4.45	4.58	4.00	4.49	4.37
Q	2.90	2.53	2.00	2.37	2.15
R	4.66	4.82	4.00	4.49	4.56
S	4.60	4.79	3.50	4.54	4.52
T	4.65	4.90	4.50	4.60	4.56
U	1.94	1.65	1.50	1.91	1.59

TABLE 16

DESCRIPTIVE STATISTICS

A COMPARISON OF KENTUCKY AND KANSAS MEANS

	E. KENTUCKY		KANSAS	
	TEACH	PRINC	TEACH	PRINC
A	2.08	2.28	1.54	1.50
B	4.74	4.69	4.88	4.76
C	3.15	3.21	2.70	2.79
D	4.84	4.59	4.82	4.74
E	4.17	4.13	4.19	3.72
F	2.55	2.64	2.46	2.03
G	4.68	4.59	4.75	4.73
H	3.13	3.05	2.68	2.34
I	3.99	3.82	4.42	4.65
J	4.66	4.49	4.71	4.85
K	4.27	4.23	4.68	4.85
L	2.45	2.80	3.90	4.00
M	4.55	4.44	4.39	4.54
N	4.61	4.33	4.69	4.58
O	2.40	2.49	2.04	2.08
P	4.53	4.44	4.68	4.92
Q	2.70	2.33	1.69	1.46
R	4.75	4.44	4.73	4.68
S	4.71	4.44	4.71	4.85
T	4.78	4.59	4.77	4.85
U	1.78	1.87	2.08	1.84

REFERENCES

- Allen, Dwight and Kevin Ryan, "Microteaching," Reading, Mass.: Addison-Wesley, 1969.
- Barr, Arvil, "Measurement of Teacher Characteristics and Prediction of Teacher Efficiency," Review of Educational Research, XXII, June 1952.
- Bernardin, H. & C. Walter, "Effects of Rater Training and Diary Keeping on Psychometric Error in Ratings." Journal of Applied Psychology, 62, 1977.
- Berliner, David C., "A Status Report on the Study of Teacher Effectiveness." Journal of Research in Science Teaching, 13, 1976.
- Berliner, David C. and William J. Tikunoff, "The California Beginning Teacher Evaluation Study: Overview of the Ethnographic Study." Journal of Teacher Education, 327, Spring, 1976.
- Borman, W., "Exploring the Upper Limits of Reliability and Validity in Job Performance Ratings." Journal of Applied Psychology, 63, 1978.
- Broudy, H.S., "Can We Define Good Teaching." Record, 70, 1969.
- Combs, Arthur W., The Professional Education of Teachers. Allyn and Bacon, Boston, 1965.
- Cooper, W., "Ubiquitous Halo." Psychological Bulletin, 90, 1981.
- Deever, Merwin R., The Evaluation of Teaching Competence: Workshop Manual. Arizona State University, 1971.
- Eastridge, H.E., NASSP Bulletin, 60:412, p. 48, 1976.
- Feldman, J., "Beyond Attribution Theory: Cognitive Processes in Performance Appraisal." Journal of Applied Psychology, 66, 1981.
- Flanders, Ned, "Teacher Influence, Pupil Attitudes and Achievement" (Washington, D.C.: U.S. Government Printing Office, 1965).
- Flanders, Ned, "Interaction Analysis as a Feedback System" in Edwin Amidon and John Hough (Eds.) Interaction Analysis: Theory, Research and Application. Addison-Wesley, Reading, Massachusetts, 1967.
- Gay, L.R., Educational Research, Competencies for Analysis and Application. Charles E. Merrill Publishing Company, 1976.
- Glass, Gene, and Julian E. Stanelly, "Statistical Methods in Education and Psychology," Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1970.

Goens, G.A. and R.W. Lange, NASSP Bulletin, 60:412, p. 15, 1976.

Grant, Stephen, "A Survey of Elementary School Principals and Teachers: Teacher Evaluation Criteria," Education, 100:3, Spring 1980.

Heitzman, William R. and Charles Starpoli, "Teacher Characteristics and Successful Teaching." Education, 95, Spring 1975.

Huntley, John F., "Academic Evaluation and Grading: An Analysis and Some Proposals." Harvard Educational Review, 46, 1976.

Landy, F. & J. Farr, "Performance Rating." Psychological Bulletin, 87, 1980.

Lemley, Raymond, "Evaluation: The Key to Effective Learning." NASSP Bulletin, 67:461, March 1983.

Marsh, H.W., "Student's Evaluation," The British Journal of Educational Psychology, 52:Part 1, Feb. 1982. National Education Association NEA Research Bulletin, 1969.

Mishra, S.P. "Correlates of Effective Teaching as Measured by Student Ratings." Journal of Experimental Education, 49:59, Fall 1980.

Orpen, C., "Student Evaluation of Lectures as an Indicator of Instructional Quality," Journal of Educational Research, 74:5, September 1980.

Ovard, G.F., National Association of Secondary School Principals Bulletin, 59:401, 1975.

Pembroke, Eileen, & Edmund Goedent, "What is the Key to Developing an Effective Teacher Evaluation System." NASSP Bulletin, 66:458, Dec. 1982.

Persig, Robert, "Zen and the Art of Motorcycle Maintenance." (New York: Morrow, 1974)

Pine, Gerald, "Necessary Conditions for Evaluating Teachers," NASSP Bulletin, 59:395, Dec. 1975.

Popham, James, "Catch-22 for Teachers", Principal, 61:5, May 1982.

Quereski, M. Y., "Perceptions of Traits that Identify the Best and Worst Teachers in Two Cultures." Journal of Experimental Education, 49:4, Fall 1980.

Rieff, J.C., "Evaluating Student Teacher Effectiveness." College Student Journal, 14:369, Winter 1980.

Reinhartz, Judy and Don Beach, "The Principal: A Vital Component in Assessing Teacher Education Programs," NASSP Bulletin, 66:456, Oct. 1982.

SAS Institute, Inc. SAS Introductory Guide: Statistics, 1978 Edition. Cary, North Carolina: SAS Institute, Inc., 1978.

SAS Institute Inc. SAS User's Guide: Statistics, 1982 Edition. Cary, North Carolina: SAS Institute Inc., 1982.

Suprina, R.N., NASSP Bulletin, 62:434, p. 54, 1978.

Thornton, G. & S. Zorich, "Training to Improve Observer Accuracy." Journal of Applied Psychology, 65, 1980.

Verble, Margaret S., "Evaluating the Quality of Teaching." Peabody Journal of Education, 56:2, 1979.

Victor, J.B. and J.P. Otis, "Teacher Strength and Sensitivity Behavior: Attitude Personality Correlates." Journal of Experimental Education, 49:9, Fall 1980.

Watson, F.G., In Handbook of Research on Teaching, N.L. Gage (Ed.), Chicago: Rand McNally & Co., 1963.

Accepted by the Graduate faculty of the School of Education
Morehead State University, in partial fulfillment of
the requirements for the Education Specialist
Degree in Administration and Supervision

Ronald Mery
Director of Applied Project

Applied Project Committee:

Ronald Mery, Chairman
Robert C. Newham
William J. Moore

10/3/83
Date