"Legitimizing Subjectivity: Meritorious Performance and the Professionalization of Teacher and Principal Evaluation"

Barnett Berry
Associate Director
The South Carolina Educational Policy Center
University of South Carolina

and

Rick Ginsberg
Associate Professor
Department of Educational Leadership
University of South Carolina

Columbia, South Carolina 29208

An earlier version of this paper was presented at the 1988 Annual Meeting of the American Educational Research Association, New Orleans, Louisiana.
ABSTRACT

By critiquing the status of teacher and principal evaluation and merit pay systems across the nation; and by describing findings from a series of studies of South Carolina teacher and principal reforms, the authors argue for "legitimizing subjectivity" in personnel evaluation. By this, the authors infer that evaluatees will accept and respect the professional judgments of their evaluators, "allowing" them to render evaluative decisions based on situation-specific processes and outcomes.

Quite simply, teaching and principaling is nonroutine work, whereby, standards cannot be codified and mandated from the outside; and only the authority of institutionalized expertise can develop and enforce professional norms and standards. The South Carolina data reveal that teachers and principals recognize that the complexities of their job tasks demand multiple data sources and highly trained, professional evaluators who have the time, resources, and expertise to judge their performance along high-inference variables. As more demands for merit pay schemes surface, policymakers would be wise to appreciate the pleas of practitioners in states like South Carolina, and devise programs which respect the need for subjectivity in evaluation.
"Legitimizing Subjectivity: Meritorious Performance and the Professionalization of Teacher and Principal Evaluation"

INTRODUCTION

In the 1980s, a remarkable concern for the quality of public education developed, evidenced in part, by the interest in teacher, and most recently, principal evaluation. Highly-visible commission reports such as *A Nation At Risk* (1983) and *Time for Results: The Governor's 1991 Report on Education* (1986) focused considerable attention on the dramatic need for better evaluation systems to judge the effectiveness of both teachers and principals. Policymakers' concerns about personnel evaluation stem, to a certain degree, from the effort to professionalize teaching and principaling through the promotion of merit pay programs. Simply speaking, many policymakers believe that merit pay programs can create incentives for talented individuals to enter and remain in public education. With increased resources and rhetoric for change—a by-product of the educational reform movement—many states have taken considerable strides toward improving their evaluation systems and using the information they produce to reward teachers and principals. Evaluation can control the standards and norms of any profession (Darling-Hammond, 1985). But, poorly designed and implemented systems of evaluation can have negative effects, and serve to deprofessionalize teaching and principaling.

Recently, merit pay programs in education have proliferated. For example, between 1983 and 1986, forty-six states considered some kind
of teacher performance-based compensation system. Hundreds of millions of dollars have been spent and hundreds of thousands of teachers (and to some degree, principals) have participated in a variety of performance-based pay systems. Both the pervasiveness and the financial costs of these performance-based compensation systems may very well earn these reforms the label of "the largest educational experiment in the United States today" (Cornett, 1987).

In some states, the legitimacy of the multi-million dollar investment in merit pay systems have been seriously threatened by inadequate performance evaluation programs, which have been described as "a statistical mishmash" (Olsen, 1987). Such problems have caused several state mandated teacher "merit pay" plans to be abandoned (e.g., as in Florida) and others to undergo major revisions (e.g., as in Texas).

From a historical perspective, this should not surprise anyone. Teacher merit pay plans flourished briefly during the 1920s and 1950s, but also failed due to inadequate methods for judging superior teaching (Johnson, 1984; Murname and Cohen, 1986). In those merit pay programs that have worked, Murname and Cohen (1985) found that the problems of developing defensible evaluative criteria were solved in political rather than scientific terms. In merit pay programs that work, "'inventing" teacher evaluation criteria has been generally a "collaborative project" among teacher representatives (Cohen and Murname, 1985, p. 7). In a recent assessment of the "evolution of teacher policy" in the United States, Darling-Hammond and Berry (1988)
assert:

Fair and credible evaluation processes for awarding special status to teachers are a prerequisite for both short- and long-term viability of the (merit pay) programs. In the short-term, evaluation problems can undermine both teacher and administrator support for initiatives, by creating cynicism among teachers, if decisions seem arbitrary or invalid, and resistance among administrators, if time demands and administrative headaches deflect from their central educational tasks (p. 72).

Although merit pay programs for principals have not been as pervasive, the evaluation of principals has taken on increased significance. A recently released survey by the Southeastern Educational Improvement Laboratory (Peters, 1988) found that 77 percent of the states now require (or soon will require) principal evaluation. Most state mandates call for principal evaluation to be conducted once a year. However, the actual evaluation is left in the hands of local officials. In many cases, the systems employed are being driven by the effective schools research and its emphasis on the principal as instructional leader.

Wise and his colleagues (1984) have argued that merit pay programs presuppose the existence of effective evaluation systems. Unfortunately, recent evidence strongly suggest that many teacher merit pay plans are failing because of the failure of the performance evaluation systems to capture the essence of effective teaching (Darling-Hammond and Berry, 1988); and that principal merit pay plans have little hope of succeeding without dramatic development in principal evaluation systems (Ginsberg, 1988a).

In fact, our argument is that for teacher and principal evaluation to be successfully implemented—especially for merit pay purposes—more
holistic systems must be developed and subjectivity in decisionmaking must be legitimated. By this, we mean that evaluatees will accept and respect the professional judgments of their evaluators, "allowing" them to render evaluative decisions based on situation-specific processes and outcomes. Although evidence suggest that teachers and principals are beset by what Michael Scriven (1983) calls, "valuphobia," (i.e., the fear of being evaluated), our data suggest that educators are more willing to expand the repertoire of approaches for evaluating their performance.

In this paper, we develop this argument that professional judgments can and should be a part of teacher and principal evaluation by first, critiquing the status of teacher and principal evaluation; and by second, describing--through a variety of primary and secondary data sources--teacher and principal beliefs regarding their respective evaluation systems.

TEACHER EVALUATION WITH "BUDDLES"

Since the early 1900s, educational researchers have conducted countless studies to discover and develop a valid index of teaching skill and competence (Popham, 1988). However, the review of teacher effectiveness investigations "reveals a woeful record of unfulfilled hopes and unrejected null hypotheses" (Popham, 1988, p. 274). In the latest round of this sorrowful saga, policy makers (especially in state capitals) and educators have joined forces to develop teacher evaluation instruments which would rid classrooms of incompetents,

Legitimizing Subjectivity

134
improve the performance of the average teachers, and differentiate and reward the expert pedagogue. Unfortunately, there is little evidence that many of the state-mandated teacher evaluation systems fulfill any of the above mentioned purposes.

"Champagne Evaluation." In large part, these recently developed evaluation systems utilize observation checklists in which raters (usually principals) simply look for generic indicators of teaching skill and then check off whether or not the behaviors were "observed" or "not observed."¹ In some cases, the instruments call for the observer to determine if the teaching behavior was inadequate, adequate, or, even, superior. In most cases, the observer is asked to assess as many as 50 to 100 separate indicators in a single class period.

Because the architects of these systems recognize the fragile knowledge base on which teaching rests, they "usually go to great lengths to avoid relying on human judgment in the determination of a teacher's skill" (Popham, 1988, p. 283). While having the advantage of appearing to be objective (by relying on tallies of behaviors rather than evaluator judgments), these "generic" teacher evaluation systems have been criticized as failing to capture the important features of good teaching (MacMillan and Pendlebury, 1985). In fact, research has revealed that effective teaching behaviors vary considerably for students with different socioeconomic, mental, and psychological

¹ In some cases, these instruments generally have been adopted from evaluation systems designed to assess the basic skills of beginning teachers.
characteristics (e.g., Brophy and Evertson, 1977; Cronbach and Snow, 1977; Peterson, 1976; Teddie and Stringfield, 1987) and for different grade levels and subject areas (Gage, 1978; McDonald and Elias, 1976). Perhaps, most importantly, as Wise and his colleagues (1984) note:

Teaching behaviors that have sometimes proved effective when used in moderation can produce significant and negative results when overused or when applied in the wrong circumstances. This kind of finding discourages the development of rules for teaching behaviors that can be applied generally (p. 10).

As Popham (1987) laments, the pervasive approach to teacher evaluation calls for a process to provide a "host of mini-judgments," whereby points are often assigned, depending on whether or not the teacher "stresses important dimensions of content" or "specifies expectations of classroom behavior" (p. 25). As he asserts:

....Before long, unfortunately, these aggregations are treated with respect and deference not warranted by their origins....Evaluators must often fill in myriad little bubbles on some sort of evaluation form, each little bubble representing an assignment of a specified number of points to the teacher on a particular dimension. A perusal of these appraisal forms reveals far more bubbles than is found in a glass of sparkling wine, imported or domestic....Thus, we can think of this judgment avoidance as a "champagne" teacher evaluation strategy (p. 25-26).

This critique of teacher evaluation does not even consider the significant problem of having competent evaluators fill in the bubbles. Ginsberg (1988b) argues that most principals are poorly trained in evaluation techniques, while Darling-Hammond (1985) argues that most principals--as generalists--do not have the expertise required for differentiating distinctions among teaching performance across grade levels and subject areas.

Legitimizing Subjectivity
THE FOLKLORE OF PRINCIPAL EVALUATION

Without question, the importance of principal evaluation cannot be underestimated. However, while teacher evaluation is recognized, at least, as an evolving process (Murphy, Hallinger, and Peterson, 1985), principal evaluation is lambasted for still being in the "stone age" of its development (Natriello, et. al., 1977). This is unfortunate given the growing emphasis on the school principalship.

Over the last decade, the omnipresent effective schools research has pointed to the critical importance of the principal in successful schools; and as a result, the means of principal evaluation have come under greater scrutiny. However, while a great deal has been written about principal evaluation, no definitive answers have emerged as to the best approach to use in a given situation (Ginsberg, 1988a). The research that has been conducted does not go very far beyond surveys of current practice. Such discussions of the "folklore" of principal evaluation are useful for sharing "home recipes," but without any empirical support for offering a particular method, potential users (a state education agency or local school district) have been left to their own wits in deciding what may work best.

Quite frankly, the effective schools research has been of very little help in the development of principal evaluation. For example, school effectiveness constructs such as instructional leadership, school climate, high expectations, and coordination and organization have emerged and are almost impossible to operationalize (Kroeze, 1984; Sweeney, 1982; Shoemaker and Fraser, 1981). To make matters more

Legitimizing Subjectivity

137
difficult, the principal's workday is strewn with unexpected interruptions and characterized most accurately by variety, brevity, and fragmentation (Martin and Willover, 1981). On top of this is the situational nature of principaling, with principals having to tend to the idiosyncratic demands of their particular school. Without question, schools existing within and across school districts vary greatly, each demanding differing leadership styles, and consequently, principal evaluation methods.

State of the Art, or Lack Thereof. A review of the literature does reveal trends in principal evaluation (Ginsberg, 1988a). The most common practice today is to evaluate once a year, and direct the evaluation at improving principal performance. Both self-evaluation and use of client data (teachers, parents, and students) are recommended, but are not widely incorporated. Unfortunately, evaluators (superintendents or their designers) and evaluatees (principals) often do not agree on the specific criteria and processes being used in their school districts.

Although a wide array of methods and techniques are used in principal evaluation, most systems utilize pre-set performance standards in the form of checklists. Criteria fall into one of three categories—traits, behaviors, or tasks. Behavioral criteria appear to be the most prevalent, although some principal evaluation systems use a combination of all three. Instruments tend to use either just rating scales or rating scales which allow for some comments and discussion. In a very few cases, the instruments ask only for a narrative.

Legitimizing Subjectivity
Finally, some evaluators use observations (and a conference); some just require principals to document performance through the use of a portfolio.

Much like in teacher evaluation, what seems to be missing is in principal evaluation is the reliance on human expertise and professional judgment as the key data gathering instrument. As conceptualized in Elliot Eisner's (1975) educational connoisseurship model of evaluation, humans judge and appraise the quality of complex educational processes and outcomes, just as art critics judge and appraise complex works of art.

In the past, both teachers and principals have clung to "objective" checklist-type evaluation because of their mistrust of their evaluators. However, as our data suggest, both teachers and principals may be ready to legitimate subjectivity in the course of their performance evaluation by accepting more judgments to enter into decision making. With the increased pressures of 1980s-styled accountability and pushes for merit pay schemes which require more sophisticated evaluation instruments, it is not unusual to hear both teachers and principals to exhort (as one principal in South Carolina did), "there is a lot more to what I do than is measured on these damn instruments."

TEACHER AND PRINCIPAL REFORM IN SOUTH CAROLINA

South Carolina is an interesting state to examine both teacher and principal evaluation systems since its recent reform package, the

Legitimizing Subjectivity
Education Improvement Act (1984), mandated the piloting of three statewide programs: Teacher Incentive Program (TIP), Principal Evaluation Program (PEP), and Principal Incentive Program (PIP). An earlier reform, the Educator Improvement Act was mandated in 1979. Additionally, a local school district is planning to implement a lead teacher concept as part of its effort to pilot several of the National Governors' Association "1991" school reforms. What follows is a brief description of these reforms and the respective data sources which we use to draw our conclusions.

Teacher Reform: Evaluation and Incentive Programs. Over the last several years, South Carolina has: (1) mandated a statewide beginning teacher evaluation system (Assessment of of Performance in Teaching or APT), (2) mandated that each school district adopt an evaluation system for veteran teachers which meet state specified criteria and procedures, and (3) piloted a teacher incentive program (TIP) which provides the opportunity for teachers to receive an additional $2000-3000 in compensation for superior teaching performance as well as extended service activities. Apart from the state teacher "merit pay" program, one local school district is planning--through the lead teacher concept--to identify outstanding teachers, differentiate their responsibilities, and substantially reward through higher salaries ($8000 to $10,000 above regular salary) and greater decision making authority.

In practice, most districts are evaluating both their beginning and veteran teachers with the APT which was designed to measure minimal
competency and the application of fundamental teaching skills, not teaching proficiency. The APT system is a classic example of "champagne evaluation" whereby evaluators (usually principals who were poorly trained) check off a box if a teacher performs the specified behavior. Thus, in South Carolina (including in the "lead teacher" district), it is not uncommon to find 35 percent of a district's teachers receive overall satisfactory evaluation ratings. In addition, while TIP is in its third year of piloting, only 9 percent of the eligible teachers are participating. Many teachers report that the program models cannot identify the "best" teachers, and thus, they choose not to compete for the trivial "meritorious" distinction implied by program standards. This is not surprising, given the fact that most TIP pilot districts use the APT (or an instrument with slight modifications) to evaluate its teachers.

Teacher Data Sources. First, Berry, et. al. (1987) conducted case studies of each the state's seventeen 1986-87 school districts involved in the Teacher Incentive Program. Two-hundred and fifty (250) interviews were conducted with participants (teachers and administrators) and non-participants (teachers) in the TIP pilot districts. Participants were purposely sampled and interview questions focused on key programmatic components—with particular attention paid to the development and implementation of the district's teacher evaluation system and its relationship to the merit pay program (see Berry, et. al., 1987)

Second, we have reviewed the findings of the third-party

Legitimizing Subjectivity
evaluation of the 1986-87 TIP, which relied on the surveys of 833 participating and non-participating teachers (randomly sampled) in the 1986-87 pilot districts. Several of the survey questions speak to teacher beliefs regarding the efficacy of their evaluation systems (see MGT., Inc., 1987a).

Finally, Berry and Ginsberg (1987) conducted a study to ascertain how the "lead teacher district" should selected its new teacher leaders. Numerous data were collected and analyzed, including the survey of 90% of the district's 400 teachers and 90% of the district's 25 school-level administrators. In addition, 25% of the district's teachers were interviewed. Survey and interview questions were designed to solicit information regarding the desired role of lead teachers and relevant selection criteria and processes, as well as information regarding current and needed teacher evaluation practices.

Principal Evaluation and Incentive Programs. Also, over the last several years, South Carolina has: (1) mandated a statewide principal evaluation program (PEP) as well as (2) piloted a principal incentive program (PIP). PEP requires that principals must document performance in 24 selected areas of competence by primarily preparing a portfolio which "proves" they have met state standards. The principals must document their performance from a list of appropriate data sources defined in each performance standard. After at least one observation, the supervisor meets the principal; rating each performance on a one to three scale; and later in the year, the supervisor reviews the portfolio and rates

Legitimizing Subjectivity
summatively each principal on a five-point scale. PIP requires principals to use PEP and/or a management-by-objective (MBO) process to "prove" that their performance was superior and to earn an additional $5000 in compensation.

In practice, with regard to the implementation of PEP, virtually all principals--much like the children in Lake Wobegon--are rated "above average" and poorly trained evaluators (usually superintendents) have great difficulty in "cleanly" translating the formative 3-point scale into the summative 5-point rating (see Berry and Ginsberg, 1987, for more details). With regard to the implementation of PIP, approximately 60 percent of the eligible principals are participating. Many principals do not participate or drop-out of the program because the program is perceived as "busy work" and there are "too many other priorities" to which they must attend (see MGT, 1987b and Berry, Ginsberg, and Cook, 1988 for more details).

Principal Data Sources. Berry and Ginsberg (1987b) conducted a third-party evaluation of the 1986-87 PEP implementation. A combination of quantitative and qualitative data were collected and analyzed. Survey data were collected from 100% of the superintendents (n=14), 100% of the evaluators (n=14), 95% of the principals (n=188), and 64% of the school board chairs (n=9) in the 14 participating school districts. In addition, open-ended interview data were collected from 100% of the superintendents and evaluators, 45% of the principals, and 35% of the school board chairs. The evaluation study focused on the adequacy of the performance statements, the reliability and validity of

Legitimizing Subjectivity
the documentation process and the rating scales used as well as the feasibility of using PEP for merit pay or PIP (see Berry and Ginsberg, 1987b for more details).

In addition, we have reviewed the findings of the third-party evaluation of the PIP, which relied on the survey of 60 percent of the principals (both participating and non-participating) in the nine 1986-87 pilot districts (see MGT., Inc., 1987b). Several of the survey questions speak to principal beliefs regarding the efficacy of their evaluation systems (see MGT., Inc., 1987b).

WHAT TEACHERS SAY ABOUT MERITORIOUS PERFORMANCE AND EVALUATION

Our review of data suggest that teachers are somewhat ambivalent about whether or not their school districts can implement an unbiased process to identify and reward superior teachers (see Table A). In large measure, teachers' ambivalence is a result of their lack of faith in performance evaluations distinguishing among poor, average, and superior teaching.

As one secondary school teacher asserts:

Being evaluated by the APT is like being on stage and the teacher is the actor. You put on your best performance two or three times over the course of the year....The problem is that some of your worst teachers are your best actors and therefore pass the evaluation. These teachers will warn their kids, "the principal will be coming tomorrow and you better be good." I know some teachers will give their students the answers to the questions that will be asked the next day so they will look good....The APT is useless.

Another teacher voices a similar concern:

The APT can only show you if a teacher is capable of conducting a good lesson....It does not mean that you always conduct good

Legitimizing Subjectivity

144
lessons, much less great lessons....How can you identify superior teaching in two observations a year....This is an insult. Thus, it is not surprising that in the survey of teachers in South Carolina's "merit pay" pilots, 67 percent of the participants and 68 percent of the non-participants do not believe the program's evaluation procedures ensure that superior performance will be awarded. In turn, 79 percent of the participants and 85 percent of the non-participants do not believe the majority of the district's superior teachers are choosing to compete for merit awards. Similarly, 64 percent of the participants and 62 percent of the non-participants have conceptions of superior teaching which are not consistent with the award requirements of the program.

Our study of the South Carolina school district implementing the lead teacher concept reveals similar teacher beliefs. Specifically, only one-third of the district's teachers believe their current evaluation process can be used to designate and recognize outstanding teachers. On the other hand, two-thirds believe a new evaluation system should be designed to designate and recognize outstanding teachers (see Table B). As one teacher notes:

We need more observations, but not anything like the APT....We need different type of instruments. No, I take that back, we need different types of people....Peer teachers know something--we have to get them into classrooms.

In fact, 88 percent of the district's teachers believe that identifying superior teaching requires multiple classroom observations over an extended period of time (see Table B). As an elementary school teachers asserts:

Legitimizing Subjectivity

-  145
We need to have these observations conducted over a period of several weeks so that you can see how well she teaches different lessons at different times of the day.

Teachers believe that their peers are better judges of teaching performance than school administrators. For teachers, administrators just do not have the time or the specific expertise to provide the diverse types of instructional leadership required in their schools. The physics teacher and the special education teachers lament that principal just cannot "truly understand" what they do in their classrooms. The composite survey data reflect their beliefs. On the one hand, 73 percent of the teachers believe that expert teachers can establish criteria and standards for the identification of superior teachers. On the other hand, only 45 percent of the teachers believe that expert administrators can establish criteria and standards for the identification of superior teachers. Nonetheless, teachers believe strongly in a partnership with administrators in identifying the best teachers. As an elementary school teacher claims:

We need a team of teachers and the principal to come in at any time of the day and see what is going on. But the team needs to understand what the teacher has been doing and what she intends to be doing later....To determine if she is an effective teacher requires the team to determine if the teacher is effective consistently.

Our interviews with teachers (both from the 17 "merit pay" pilots and the "soon-to-be" lead teacher district) reveal that they are ready to legitimize subjectivity in their performance evaluations and move away from the highly mechanistic, low-inference instruments which currently prevail. However, a caveat is in order. Teachers are not

Legitimizing Subjectivity
ready to turn over "subjective" evaluations to principals and peers who are not adequately trained. One reason that teachers are not totally opposed to "champagne" evaluation is that it has protected them from the capriciousness of the generalist evaluator who "knows nothing about their subject" and "little about their students." Teachers recognize that principals, other administrators, and their peers do not necessarily have the skills, the training, and the time to conduct performance evaluations which can adequately assess their teaching more systemically; or in other words, assess their teaching more subjectivity. There appears to be a latent tendency of teachers to call for an evaluation which assess how well they plan, how well they impart the structure of knowledge in their discipline, and how well they account for students' levels of development and prior learning. However, teachers are unsure if their district will allow them the time and resources to develop such an evaluation system. To judge relative degrees of competence in teaching, the format of the evaluation "must reach beyond observed teaching behaviors on a given day or days...(and) the quality of ongoing classroom activities depends on how what happens today relates to what happened yesterday and last week, as well as what will happen tomorrow and thereafter (Wise, et. al., 1984, p. 55). Unfortunately, the development of such a system does not have support and commitment in many South Carolina school districts, and teachers know it.

Legitimizing Subjectivity
WHAT PRINCIPALS SAY ABOUT EVALUATION AND MERITORIOUS PERFORMANCE

Our review of data suggest that principals are also somewhat ambivalent about whether or not their school districts can implement an valid process to identify and reward superior principals (see Table C). In large measure, principals' ambivalence is a result of their lack of faith in performance evaluations distinguishing among poor, average, and superior teaching. As one principal asserts:

Being evaluated by the PEP (Principal Evaluation Instrument) is like the APT for teachers....A bad principal can effectively pass this instrument.

Another principal echoes this belief about the inadequacies of PEP for differentiating performance.

The process may be able to point a real bad principal....I am not sure it is accurate enough to point out higher levels of performance....The PEP process sets out minimum standards.

Thus, it is not surprising that in the survey of principals in South Carolina's "merit pay" pilots, only 26 percent of the principals believe the program's procedures ensure that superior performance will be awarded. Similarly, only 29 percent of the principals believe that PEP differentiates properly between acceptable and superior performance (see Table C). In the survey of the PEP pilots, principal responses reveal relatively low mean scores and high variability on the two items related to utility of PEP for differentiating performance and for the Principal Incentive Program (see Table D).

Our third party evaluation of the implementation of the 1986-87 Principal Evaluation Program revealed solid support among principals

Legitimizing Subjectivity
for the instrument's 24 performance statement (Ginsberg, Berry, and Cook, 1988). In fact, in the survey of principals in South Carolina's "merit pay" pilots, it was found that 79 percent of the principals felt that the PEP instrument should be an integral part of the incentive program (see Table D). However, only 52 percent of these same principals believe that PEP is appropriate for evaluating them (see Table D). Other findings suggest why principals are so concerned about using the present evaluation system for merit pay, or using "PEP for PIP" (see Ginsberg, Berry, and Cook, 1988). Our PEP evaluation revealed that certain critical performance statements--such as those related to conflict management, sensitivity, and making timely decisions--are too general, and thus, very difficult to document as required by the evaluation process. In the interviews, we learned about: (1) the problems related to the difficulty of making judgments solely from the portfolio, (2) how easy it might be for one to falsify documentation, and (3) how the whole process as conceived fails to point some very important key elements of principaling. Two comments from the interview data set reveal how easy it is for principals to meet the letter and avoid their spirit of the PEP process,(what we call "fudge") and, how the portfolio method does not capture what principals need to do (what we call "misses point"): FUDGE: It is easy to make what you do look like what they want; If you are organized, then you can make the portfolio look good, even if you are not an excellent principal; I will pull good letters and leave poor ones in the file; Much of the documentation I might do would be contrived; It is easy to beat the system.

Legitimizing Subjectivity
MISSES POINT: I do much more than can be gleaned from the data sources; How do you document, for example, that I resolve conflict? It falls short of giving a complete picture of your performance and what you do as a principal; So many things you do informally do not lend themselves to documentation; Attitudes cannot be documented; It calls for dotting "i's" and crossing "t’s."

The interviews conducted during the PEP evaluation revealed that to make accurate assessments, "more time needs to be spent looking at the whole school environment." Potential sources for better evaluation would come from actual observations, shadowing of principals, and surveys of teachers, parents, and students. In fact, in the survey of the "merit pay" pilots, 44 percent of the principals agreed that surveys of teachers, surveys, and parents should be a major component in documenting their performance (see Table C). Principals believe strongly that if meritorious performance is to be identified and rewarded, then more subjective measures need to be incorporated into the process.

Much like teachers, principals express reservations about "subjectifying" the evaluation process without adequate training for and commitment of the evaluators. In fact, our PEP evaluation revealed that both superintendents and evaluators generally rated their training for the administration of the process as "poor" or "limited" at best (Berry and Ginsberg, 1987). If subjectivity in performance evaluation is to be legitimated, then those who are evaluated must entrust those who will do the evaluation. Unfortunately, for principals and teachers alike, auspicious conditions do not currently exist for training and empowering evaluators in making subjective evaluation judgments.

Legitimizing Subjectivity
IMPLIEDATIONS

The movement to professionalize teaching (and to some extent the principalship) through offering schemes for incentive and merit pay have intensified the importance of sound evaluation mechanisms. Much like other state-mandated evaluation systems, data from the South Carolina teacher and principal reforms reflect quantitative and objective approaches which provide an "air" of science and a haven from legal challenges—especially if tenure or merit pay decisions are involved. Public bureaucracies—like schools—are compelled to at least treat all employees alike. So educators go to great lengths to evaluate each teacher or principal exactly the same way. Evaluation resources are diluted to meet mandates and the work of teachers and principals which are not easily measured with objective indicators may go unrecognized and unrewarded.

Thus, teachers are not evaluated by such critical criteria as capacity to communicate or relate to children, knowledge of subject matter, the appropriateness of technique, and the integrity of curriculum (Darling-Hammond, 1985). Similarly, principals may not be able to document the critical management functions of treating personnel with sensitivity, making timely decisions, and dealing with conflict.

Effective teaching and principaling are context-specific. Those who must utilize a repertoire of complex behaviors in the complex and dynamic setting of public schools are likely to resist highly, mechanistic evaluation systems. These systems cannot necessarily

Legitimizing Subjectivity
reflect standards which must be based upon the diverse needs of students and schools as well as varying degrees of resources.

The evaluation systems currently in vogue treat teaching and principaling as processes which can be engineered and dissected by an exact and objective formula. However, teaching and principaling are knowledge-based work that is "indeterminate" and cannot be reduced to prescriptions for practice (Darling-Hammond, 1985). Quite simply, teaching and principaling are nonroutine work which requires practitioners to make innumerable judgments which best serve their clients. Thus, knowledge-based work--the work of professionals--"is by its very nature not amenable to mechanization and rationalization" (Friedson, 1973, p. 55). Standards cannot be codified and mandated from the outside; and only the authority of institutionalized expertise can develop and enforce professional norms and standards (Darling-Hammond, 1985). Institutionalized expertise is legitimated through the acceptance of subjective judgments by those who know and understand goals, processes, and outcomes.

Those who evaluate teachers and principals must be highly knowledgeable, experienced, and be able to translate theory into situation-specific actions. These evaluators must not only "know," but they must "know how to know." Validity and reliability are accounted for through legitimated selection procedures, intensive training, and multiple judgments offered over time by multiple experts who utilize a wide range of data sources. Classrooms and schools must be opened up and those who know best about teaching a certain type (or age) of

Legitimizing Subjectivity

152
child, or who knows best about teaching a particular subject, or who
knows best about managing a certain kind of school, get involved in
evaluation. But, there is more. Informed judgments can be derived
from teachers (other than peer evaluators), parents, students, and the
community. Individuals who know--experience and feel--the actual
effects of teacher and principal work are a part of a more
professionally-oriented evaluation.

The South Carolina data show that teachers and principals
recognize that the complexities of their job tasks demand multiple data
sources and highly trained, professional evaluators who have the time,
resources, and expertise to judge their performance along high-
inference variables. Teachers and principals want well trained
experts to judge their performance. Even in performance-based pay
systems--TIP and PIP--expert judgment is seen as a requisite part of
assessing superior teaching and principaling.

The work of teachers and principals is such that the
professionalization of each role requires expert evaluators to be
entrusted to judge performance. As more demands for merit pay schemes
surface, policymakers would be wise to appreciate the pleas of
practitioners in states like South Carolina, and devise programs which
respect the need for subjectivity in evaluation. Both adequate
training for evaluators, and commitment to collecting an array of
sources of information will be necessary to legitimize subjectivity in
teacher and principal evaluation. Given the context-specificness of
effective teaching and principal, the movement to remove subjective
judgment of experts is tantamount to judging wine by the shape of the bottle and the graphic on the label without any regard for its color, bouquet, and taste.
REFERENCES


\[\text{Legitimizing Subjectivity}\]


Legitimizing Subjectivity


Olsen, L. (1987). Performance-based pay systems for teachers are being re-examined. Education Week, April 15, 1, 16-17.


Legitimizing Subjectivity
<table>
<thead>
<tr>
<th>SURVEY ITEM</th>
<th>TIP APPLICANTS</th>
<th>TIP NON-APPLICANTS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Agree</td>
<td>No Opinion</td>
</tr>
<tr>
<td>It is possible to implement an unbiased process to identify and reward superior teaching in my district.</td>
<td>47</td>
<td>8</td>
</tr>
<tr>
<td>Present TIP procedures ensure that most teachers who demonstrate superior performance will be rewarded.</td>
<td>27</td>
<td>6</td>
</tr>
<tr>
<td>TIP provides an incentive for keeping South Carolina's superior teachers in the classroom.</td>
<td>31</td>
<td>9</td>
</tr>
<tr>
<td>The majority of superior teachers in my school are participating in TIP.</td>
<td>18</td>
<td>3</td>
</tr>
<tr>
<td>Superior teachers should receive different amounts of bonus pay based on different levels of performance and productivity.</td>
<td>44</td>
<td>18</td>
</tr>
<tr>
<td>Teachers should compete against each other for bonus pay.</td>
<td>12</td>
<td>6</td>
</tr>
<tr>
<td>Teachers should receive differential pay each year based on their level of performance.</td>
<td>45</td>
<td>15</td>
</tr>
<tr>
<td>My concept of a superior teacher is consistent with the awards requirements of TIP.</td>
<td>24</td>
<td>12</td>
</tr>
</tbody>
</table>

NOTES:

* Teachers responded to a 5-point scale. Above data were collapsed into a 3-point response set.

* Above respondents represent a 10% random sampling of teachers in 17 pilot "merit pay" school districts in South Carolina. See MGT, Inc., 1987, for more details.
### Table B
Teacher Survey Questions Related To Performance Evaluation
South Carolina School District Implementing Lead Teacher Concept
Percentage Responding
(n=368)

<table>
<thead>
<tr>
<th>SURVEY ITEM</th>
<th>Agree</th>
<th>No Opinion</th>
<th>Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>49. The current teacher evaluation provides sufficient support to teachers who need assistance.</td>
<td>50</td>
<td>19</td>
<td>31</td>
</tr>
<tr>
<td>50. The current teacher evaluation process can be used to designate and recognize outstanding teachers.</td>
<td>34</td>
<td>15</td>
<td>51</td>
</tr>
<tr>
<td>51. A teacher evaluation process should be designed to identify and recognize outstanding teachers.</td>
<td>66</td>
<td>18</td>
<td>16</td>
</tr>
<tr>
<td>52. Your school principal is the best judge of your teaching performance.</td>
<td>33</td>
<td>20</td>
<td>47</td>
</tr>
<tr>
<td>53. Your assistant principal is the best judge of your teaching performance.</td>
<td>15</td>
<td>35</td>
<td>60</td>
</tr>
<tr>
<td>54. Curriculum coordinators are the best judge of your teaching performance.</td>
<td>19</td>
<td>25</td>
<td>56</td>
</tr>
<tr>
<td>55. Expert teachers from your subject area are the best judge of your teaching performance.</td>
<td>60</td>
<td>18</td>
<td>22</td>
</tr>
<tr>
<td>56. Identifying superior teaching performance requires multiple classroom observations over an extended period of time.</td>
<td>88</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>57. Expert teachers can establish criteria and standards for the identification of superior teachers.</td>
<td>73</td>
<td>17</td>
<td>10</td>
</tr>
<tr>
<td>58. Expert administrators can establish criteria and standards for the identification of superior teachers.</td>
<td>45</td>
<td>25</td>
<td>30</td>
</tr>
<tr>
<td>59. The identification of superior teachers requires the use of student performance data.</td>
<td>65</td>
<td>11</td>
<td>24</td>
</tr>
</tbody>
</table>

The following criterion should be used to select lead teachers:

<table>
<thead>
<tr>
<th>SURVEY ITEM</th>
<th>Agree</th>
<th>No Opinion</th>
<th>Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>79. The current teacher evaluation rating</td>
<td>61</td>
<td>22</td>
<td>17</td>
</tr>
<tr>
<td>80. Recommendations from other teachers</td>
<td>76</td>
<td>16</td>
<td>8</td>
</tr>
<tr>
<td>81. Recommendations from principal</td>
<td>81</td>
<td>14</td>
<td>5</td>
</tr>
<tr>
<td>86. Results from a specially designed teacher evaluation process</td>
<td>65</td>
<td>29</td>
<td>6</td>
</tr>
<tr>
<td>87. Special observations of classroom performance</td>
<td>87</td>
<td>10</td>
<td>3</td>
</tr>
</tbody>
</table>

**NOTES:**
* The above items were selected from a 128 item survey set soliciting teachers beliefs regarding lead teacher role and selection as well as performance evaluation.
* Teachers responded to a 5-point scale. Above data were collapsed into a 3-point response set.
* Above respondents represent 90 percent of teachers in the school district. See Berry & Ginsberg, 1988, for more details regarding methodology, survey development and analysis.
Table C
Principal Survey Questions Related to Merit Pay and Performance Evaluation
South Carolina Principal Incentive Pilot (9 School Districts)
1986-87 Program Year
Percentage Responding
(N=81)′

<table>
<thead>
<tr>
<th>SURVEY ITEM</th>
<th>Agree</th>
<th>No Opinion</th>
<th>Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. An unbiased incentive program can be implemented to identify and reward</td>
<td>46</td>
<td>18</td>
<td>36</td>
</tr>
<tr>
<td>superior principals using PIP.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. The PIP procedures used in the model pilot-tested in your district ensures</td>
<td>26</td>
<td>24</td>
<td>50</td>
</tr>
<tr>
<td>that principals demonstrating superior performance/productivity will be</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rewarded.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. The Principal Evaluation Instrument should be an integral part of any</td>
<td>79</td>
<td>12</td>
<td>9</td>
</tr>
<tr>
<td>PIP model.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. The Principal Evaluation Instrument is an appropriate instrument for</td>
<td>52</td>
<td>19</td>
<td>29</td>
</tr>
<tr>
<td>evaluating me.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. The Principal Evaluation Instrument differentiates properly between</td>
<td>29</td>
<td>37</td>
<td>34</td>
</tr>
<tr>
<td>acceptable and superior performance.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. Surveys of teachers, students and parents should be a major component</td>
<td>44</td>
<td>17</td>
<td>39</td>
</tr>
<tr>
<td>in documenting my performance.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: MDT (1987b)

′ 60% random sample of principals in pilot school districts.
Table D
Principal Survey Questions Related to Principal Evaluation Program and Principal Incentive Program
South Carolina Principal Evaluation Program Field Test (14 School Districts)
1986-87 Program Year
(N=186)\(^a\)

<table>
<thead>
<tr>
<th>Survey Item</th>
<th>Level of agreement</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>This principal evaluation instrument adequately differentiates levels of performance.</td>
<td>Highest 4</td>
<td>2.8</td>
<td>.7</td>
</tr>
<tr>
<td>This principal evaluation instrument is sufficient for utilization in the Principal Incentive Program.</td>
<td>Lowest 1</td>
<td>2.7</td>
<td>.9</td>
</tr>
<tr>
<td>There are priority tasks of a school principal which are not included in this instrument.</td>
<td>Highest 4</td>
<td>2.6</td>
<td>1.4</td>
</tr>
</tbody>
</table>

Source: Berry, et al. (1987)

\(^a\) 95% of principals in field test districts included.
The two following articles continue the Journal's presentation of research results of the Graduated Work Incentive Experiment. The Spring 1974 issue of the Journal included six papers from the analyses of the experiment, four of which dealt with the central question of the labor-supply response to the experiment plans. Briefly discussed in each of these papers were various complications and shortcomings that pose challenges to the analysis. Perhaps the most serious challenges are the two which are expressed in the titles of the two papers included here: "The Effects of Welfare Programs on Experimental Responses" and "Predicting the Effects of Permanent Programs from a Limited Duration Experiment."

The papers are important for gaining a fuller understanding of the analyses and for interpreting the conclusions of the previous labor-supply papers; moreover, the papers stand on their own merits in the exposition of important methodological issues concerning economic analysis of controlled experiment...