DOCUMENT RESUME

ED 270 888 EA 018 575

AUTHOR	Barro, Stephen M.
TITLE	The Logic of Teacher Incentives.
INSTITUTION	National Association of State Boards of Education, Alexandria, VA.
SPONS AGENCY	National Inst. of Education (ED), Washington, DC.
PUB DATE	Apr 85
NOTE	109p.
PUB TYPE	Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC05 Plus Postage.
DESCRIPTORS	*Career Ladders; *Educational Change; *Educational Improvement; Elementary Secondary Education; Evaluation Methods; *Merit Rating; Performance Contracts; Performance Factors; Promotion (Occupational); Teacher Effectiveness; Teacher Employment Benefits; *Teacher Evaluation; Teacher Improvement; *Teacher Promotion; Teacher Salaries
IDENTIFIERS	National Association of State Boards of Education; National Commission on Excellence in Education

ABSTRACT

Widely endorsed national reports on educational reform have proposed career ladders and merit pay to raise the quality of the teaching force, and hence contribute to educational excellence. This report contends that careful analysis of proposed changes of teacher reward systems has been omitted. The issues requiring attention involve incentive rationale and incentive system design. Chapter 2 considers rationales on which recent proposals for performance-based teacher pay and promotion are founded, including behavioral assumptions, mechanisms to enhance educational quality, and conditions under which mechanisms are likely to work. Chapter 3 focuses on questions concerning the teacher evaluation component of an incentive system, putting forth criteria that evaluation methods must meet, applying them to evaluation approaches and assessing adequacy of prospective performance measures. Chapter 4 deals with such design issues as structuring of rewards. A final chapter consolidates the report's major findings; recommended are performance-contingent pay increases, universal coverage, differentiation among multiple performance levels, and predetermined reward criteria. (CJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U S DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it
Minor changes have been made to improve reproduction quality
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

ED270888

SMB ECONOMIC RESEARCH INC.
6315 29th Place, N.W.
Washington, D.C. 20015
(202) 362-0316

THE LOGIC OF TEACHER INCENTIVES

Stephen M. Barro

April 1985

Prepared for

THE NATIONAL ASSOCIATION OF STATE BOARDS OF EDUCATION

This paper was written pursuant to a grant from the National Institute of Education, U.S. Department of Education. However, the opinions expressed herein do not necessarily reflect the position or policy of either the National Institute of Education or the National Association of State Boards of Education (NASBE), and no official endorsement by the National Institute of Education or NASBE should be inferred.

TM 018 575

I. INTRODUCTION

Since the appearance, beginning in 1983, of a series of highly publicized national reports on educational reform, the method of rewarding teachers has become a salient issue of American education policy.[1] Traditional pay and promotion systems, according to these reports, are inimical to educational excellence: they deter capable people from becoming teachers, discourage good teachers from staying in the profession, produce shortages of teachers with special skills, and offer little motivation for more-than-mediocre performance. The two most frequently recommended reforms are (1) increasing the level of pay to make teaching competitive with other occupations for the services of talented people, and (2) linking pay and promotion to performance, specifically through merit pay and career ladder plans.[2] Taking these steps, the reports contend, will raise the quality of the teaching force, stimulate better teaching, and hence contribute significantly to the pursuit of excellence in American schools.

The proposals for teacher incentives—career ladders and merit pay—have received an extraordinarily warm reception. They have been endorsed by the Reagan Administration and embraced by state and local political

---

[1]The reports that deal with the teacher reward system are those of the National Commission on Excellence in Education (1983), the Twentieth Century Fund (1983), the Task Force on Education for Economic Growth (1983), the Carnegie Foundation for the Advancement of Teaching (1983), John Goodlad (1984), and Theodore Sizer (1984).

[2]Other related proposals include market-sensitive pay differentials for teachers with special skills, such as proficiency in mathematics or science; freer entry into teaching for persons with needed skills but without traditional teacher training and certification; and subsidies to induce students to enter and complete teacher training programs. These are discussed only in passing in this report.

figures and education officials around the country. Even the American Federation of Teachers (but not its larger rival, the National Education Association) has voiced its qualified approval. Several states, including Tennessee, California, and Florida, and some local school districts have already adopted such systems and are now implementing them in their schools (U.S. Department of Education, 1984; Education Week, 1985). Approximately 30 other states, as of late 1984, were developing or considering, or had given preliminary approval to similar plans (ibid.) In some of these states, pilot projects have commenced or will soon be getting under way.[3] Within a few years, if these developments continue, the salaries and ranks of significant numbers of elementary and secondary teachers around the country could depend, in part, on assessments of their teaching performance. This would be a major break with the status quo in American education (although not necessarily with historical precedent) and one with potentially far-reaching consequences for the schools and the teaching profession.[4]

The rapidity with which the teacher incentive movement has developed is at once impressive and disturbing. On one hand, it is a refreshing departure from the normally hesitant pace of educational innovation. Less than two years has elapsed since publication of the report of the National Commission on Excellence in Education, yet already parts of

---

[3]Among the states that have initiated or approved pilot projects, according to the survey in Education Week (1985), are Arizona, Idaho, New Jersey, South Carolina, and Virginia.

[4]According to Johnson (1984) and Cohen and Murnane (1985), there was great interest in merit pay and actual use of merit pay systems in large numbers of school systems in the 1920's and, to a lesser extent, in the 1950's. It is unclear to what degree those systems resembled those now being implemented or considered in the states.

BEST COPY AVAILABLE

its prescription for "professionally competitive, market-sensitive, and performance-based pay"--are close to being put into practice in several states. On the other hand, this quick leap from idea to action is troubling because of the omitted steps in between: analysis of the implications of the proposed changes and careful design of the new teacher reward systems. Thus far, little of either has taken place. The reform commissions themselves undertook no policy analyses and offered little guidance about system design (Peterson, 1983). Their endorsements of incentive approaches rest only on the most general rationales and on strong, untested, and unstated assumptions about how teachers behave and respond. Many of the recommendations of individual state task forces have been similarly vague, and some of the plans placed before state legislatures show signs of having been assembled hastily, with only minimal consideration to how their parts fit together and how the stated goals would be achieved. A new literature on teacher incentives is emerging, but most of what has been written thus far is more polemical than analytical.[5] Thus, there is a significant information gap. State policymakers, caught up in the enthusiasm over performance-based rewards, appear to be rushing to adopt and install the new systems, even though many of the underlying issues have barely been addressed, much less resolved.

What are the issues that seem to require attention? These fall into two broad categories, which, broadly speaking, can be said to embrace the "why" and the "how" of performance-based rewards:

---

[5]Among the more analytical reports that have appeared to date are Hatry and Greiner (1984) and Cresap, McCormick and Paget (1984).

BEST COPY AVAILABLE

First, the rationale for teacher incentives needs to be explored. Amidst all the excitement over merit pay and career ladders, there is considerable vagueness about what is to be accomplished (who is to be induced to do what by linking rewards to performance?), the reasons for believing teachers will respond as intended, and the relationship between teacher responses and educational results. There is also relatively little information about the the advantages and disadvantages of different incentive strategies, the conditions under which each is likely to work, and the full range of implications for the schools. All these matters converge on a question of central importance to policymakers: are there reasonable grounds for believing that performance-based teacher reward systems, properly designed, can raise the quality of teachers and teaching (while avoiding major adverse side effects), as the reform commissions and other advocates contend?

Second, assuming that the answer to the preceding question is at least a qualified "yes," one must next confront certain generic issues of incentive system design, which any state or local district considering an incentive system would somehow have to resolve. These design issues are numerous, but the following nonexhaustive list illustrates the range of relevant concerns:

- What form(s) should rewards take: permanent pay increases, one-time performance bonuses, promotions, special recognition, nonmonetary benefits?
- What dimensions of teacher performance should be evaluated, using what evaluation methods, and by whom?
- What is the appropriate size of performance-based increments in pay?

- o How many different levels of rewards should there be, and what should be the performance criteria to qualify for each?
- o Who should be eligible for performance-based rewards: all teachers or only those with those with certain minimum levels of seniority?
- o Should participation in merit pay or career ladder plans be mandatory or voluntary?
- o Against which other teachers should any particular teacher be compared?
- o What level of performance (sustained over what period) should be required to qualify for each type or level of reward?
- o If rewards take the form of promotions (e.g., to master teacher), what functions should be assigned to teachers who attain the higher ranks?
- o Should rewards be rationed and, if so, according to what rules?

This paper, an exploration of the logic underlying teacher incentives, deals with both sets of issues. It considers first (in Chapter II) the rationales on which the recent proposals for performance-based teacher pay and promotion appear to be founded, including the behavioral assumptions, the various mechanisms by which incentives might enhance educational quality, and the conditions under which those mechanisms are likely to work as proponents of incentives intend. It then turns, in the following two chapters, to the major issues of incentive system design. Chapter III focuses exclusively on questions concerning the teacher evaluation component of a teacher incentive system. It sets forth the criteria that teacher evaluation methods must meet to support systems of performance-based pay or promotion, applies them to the major evaluation approaches, and assesses the adequacy of current and prospective performance measures as the basis for distributing rewards. Chapter IV deals with all design

BEST COPY AVAILABLE

issues other than issues of teacher evaluation--which is to say, with the issues of how rewards should be structured, how they should be apportioned among teachers, and how the reward system should be installed and operated in the schools. Along the way, it touches on all the specific design question listed above. A short final chapter (Chapter V) brings together the major findings and conclusions of the paper, highlights the major uncertainties, and suggests how experience with the incentive systems now being installed in several states might be used to resolve some key empirical questions.

II. THE RATIONALE FOR INCENTIVES

The rationale for performance-based rewards for teachers rests on assumptions that are rarely stated explicitly or examined in detail. Proponents of incentives (including the reform commissions) have often done little more than to assert that we can have better teaching if we are willing to pay for it--that if explicit rewards are offered for good teachers and teaching, more of each will be forthcoming; and if rewards for poor teaching are reduced, less poor teaching will be supplied. This assertion may be correct, but without further justification only true believers in the market are likely to be convinced. To make the case more persuasively, either for or against incentives, one must establish whether reasonable assumptions about teacher behavior and its connection to educational quality suggest that incentives will work. Before discussing specific incentive plans, therefore, I review in this chapter the premises on which the major approaches are founded and the reasons for believing or disbelieving that the proposed incentive systems will work.

BASIC PREMISES ABOUT TEACHER BEHAVIOR

The central premise on which the recent incentive proposals depend is one concerning teachers' behavior. It is that teachers and prospective teachers, like other human beings, take the consequences into account when they decide what careers to pursue and how to behave on the job, and that those consequences include the tangible rewards (compensation, economic security) and the intangible rewards (job satisfaction, professional status) associated with different courses of action. Specifically, to believe that teachers may respond to the performance-contingent pay and

BEST COPY AVAILABLE

promotion plans now being discussed, one must assume that teachers care, how well they can support themselves and their families and value recognition of work well done. From that assumption, it follows that if economic benefits or professional recognition could be earned by teaching well, and if benefits could be lost by teaching poorly, some teachers, at least, might behave differently than under the existing system. Some might take steps to improve their teaching that they would not have taken otherwise (incurring costs, if necessary, to do so); some might try more strenuously to avoid poor teaching; some already-employed good teachers might remain in teaching longer; some low performers might leave sooner; and some potential good teachers might be recruited who would otherwise have chosen other fields. All these possibilities arise because under a system of performance-based rewards the optimal ("utility maximizing") pattern of behavior would be different for many current and prospective teachers than under the present regime of rewards unrelated to teaching performance.

Several points are worth noting about the foregoing premise regarding teachers' economic behavior:

First, to believe that teachers would respond to performance-contingent rewards, one does not have to make strong assumptions about the degree to which teachers and prospective teachers are economically motivated. Contrary to what has been said by some who find incentives distasteful, it is not necessary to assume that financial reward is the "primary" motive of teachers; that teachers are as interested in financial rewards and status as persons in other fields; or that such other motives as the desire to serve or to gain satisfaction from helping children learn are unimportant. It suffices that teachers assign some value (are not

BEST COPY AVAILABLE

indifferent) to material and other extrinsic rewards as well as to the intrinsic rewards of teaching. Nor is it necessary to assume that every teacher will pursue extrinsic rewards to believe that incentives can raise the average quality of teaching. All that is required for incentives to work is that some significant number of teachers respond to some significant degree.[1]

Second, the assumption of responsiveness to rewards is sufficiently general to allow for a wide variety of behavioral changes in pursuit of higher pay or status. Rewards might induce some teachers to work harder or longer, some to invest more time and effort in improving their skills, others to reallocate their efforts among types of students or areas of the curriculum, and still others to abandon their accustomed teaching methods for less comfortable but more effective alternatives. In addition, incentives might exert quality-enhancing effects on patterns of retention and entry into teaching. Thus, the efficacy of rewards is not contingent on any one form of response. There are multiple modes of constructive response and multiple channels by which incentives might affect the quality and performance of the teaching force.

Third and most important, teachers' responsiveness to extrinsic rewards is a necessary condition but not a sufficent condition for incentives

---

[1]Of course, the larger the number of teachers who respond and the greater the weight they assign to the proferred rewards, the greater will be the affect of incentives on teacher behavior. Conceivably, the degree of response could be so low (or, equivalently, the magnitude of rewards required to produce the desired effects could be so high) that incentives will prove uneconomical. But response rates cannot be inferred from theoretical arguments. Only empirical analysis, based on actual trials of performance-based reward systems, can establish whether teachers are responsive enough to make the incentive approach worthwhile.

BEST COPY AVAILABLE

to enhance the quality of teaching. Teschers and prospective teachers must have not only the motivation but al-o the capacity and freedom to change their behavior. Depending on the context, this may mean the capacity and freedom to alter behavior in the classroom, to enter teaching, or to switch from teaching to another occupation. Moreover, the objective circumstances must be such that responses can be effective. It would make no difference how responsive teachers were if there were no actions they could take that would enhance student learning. Thus, to assess the prospects for incentives, one must consider not only whether teachers can be motivated but also whether it is feasible for the hoped-for educational improvements to occur.

Because there are different types of incentives, alternative modes of response, and multiple determinants of whether incentives will work, one can say relatively little about the effects of teacher incentives in general. To go further, one must particularize the discussion, focusing on particular incentive and response mechanisms and the conditions under which each will lead toward the desired educational results.

INCENTIVES FOR WHOM TO DO WHAT?

Athough all the recent proposals for higher salaries, merit pay, career laddera, and the rest are intended to produce better teaching, they would do so by a variety of means. Some are aimed at behavior in the labor market and some at behavior in the classroom, some at already-employed teachers and some at prospective recruits. Corresponding to these different targets are different incentive strategies, and underlying each strategy is a particular theory of how teschers respond. Sorting

out these strategies is important because not all are ly c isistent or compatible. A system that creates an incentive f one gro p, or one desired type of behavior, may creat no incentive, r eve a disincentive, for another To clarify the possibilities, I enumerate e fferent target groups and incentive mechanisms, following whi e amine in detail the logic underlying each major incentive ap

Probably the easiest way to differentiate a ncentive stra te is to consider the multiple ways in which educational quality could conceivably be affected by incentives directed at the teaching force. Incentives could work, first, by eliciting better performance from existing teachers or, second, by altering the membership of the teaching force so that average quality rises. To accomplish the former, the incentives must induce already employed teachers either to upgrade their skills or to apply more intensively or effectively the skills they already have. To alter the make-up of the teaching force, incentives must influence either teacher turnover or recruitment. Specifically, to enhance quality via the turnover mechanism, incentives must reduce the turnover rate of above-average teachers, increase the turnover rate of below-average teachers' or both; and to succeed via the recruitment mechanism, incentives must attract more capable people into teaching (either from the ranks of new college graduates or from those already employed) than would have entered the profession under the existing reward system.

BEST COPY AVAILABLE

The main incentive strategies, then, can conveniently be categorized as follows:

1. Raising the average performance of already-employed teachers
    a. by inducing teachers to utilize their existing capabilities more intensively or effectively.
    b. by inducing existing teachers to upgrade their capabilities.
2. Influencing teacher turnover rates so that
    a. good teachers remain in the profession longer.
    b. poor teachers leave teaching sooner.
3. Attracting higher-quality entrants into teaching
    a. from the ranks of talented new college graduates.[3]
    b. from the pool of talented persons already in the labor force but not employed as teachers.

I consider below the conditions under which each of these strategies is likely to produce the desired effects.

INCENTIVES TO IMPROVE THE PERFORMANCE OF ALREADY-EMPLOYED TEACHERS

I have already alluded to some of the reasoning underlying the belief that performance-contingent rewards, such as those embodied in merit pay and career ladder plans, can raise the performance of existing teachers. The complete argument can be compressed into a set of three propositions,

[3] A longer-run strategy aimed at improving the quality of new college graduates entering teaching is to induce promising college students to enter and complete teacher training programs. However, incentives aimed at teacher training are outside the scope of this report.

BEST COPY AVAILABLE

all of which must be true if incentives are to have the intended quality-enhancing effects:[4]

1. There is room for improvement--currently employed teachers are capable of teaching better than they teach now.
2. Teachers have the capacity and freedom to improve--that is, educational quality can be raised by changes that individual teachers can make.
3. Teachers can be induced to make these performance-enhancing changes by the offer of performance-contingent rewards.

I consider each proposition in detail, commenting on arguments for and against its validity.

Room for Improvement

Rewards for performance can raise performance only if significant numbers of teachers are not already doing the best teaching of which they are ultimately capable. Recognizing this, some have detected in the proposals for incentives an implicit slur on teachers--namely, the implication that teachers are not working as diligently as they should be, or even that they are deliberately "withholding" services from their pupils. This is not a well-conceived reaction, however, because it neglects the variability and multiplicity of the determinants of teacher performance.

The performance that an individual teacher delivers can be thought of as jointly determined by the amount and intensity of the teacher's effort, the teacher's capabilities (knowledge and skills), and such other factors as the teacher's time allocations and choices of instructional

[4]A similarly structured set of propositions is analyzed by Rosenholtz (1985), but both my formulations of the propositions and my conclusions differ sharply from hers.

BEST COPY AVAILABLE

methods.[5] Each factor is potentially variable, does in fact vary widely among teachers, and is at least partially determined at each individual teacher's discretion.

Consider the amount and intensity of effort that a teacher puts into his or her job. It takes no more than casual empiricism to convince oneself that some teachers work harder and longer than others. Some teachers expend much energy planning lessons, grading papers, working with individual pupils and parents, etc.; others expend hardly any. Some devote many more classroom hours than others to direct instruction, as opposed to more passive (on the part of the teacher) forms of instructional activity. Less tangibly, some deal actively with children's situations and learning problems, while others engage in less-demanding, more routine modes of classroom teaching. Variability alone constitutes prima facie evidence that there is room to improve. That is, all teachers not near the upper end of the effort distribution presumably have the opportunity to emulate their more-intensively working peers.

What seems to obscure the issue of whether incentives can elicit increased effort and to provoke defensive reactions to the suggestion that they can is a tendency to think in either/or terms: either a teacher is making an acceptable effort or not; either a teacher is working at full potential or withholding effort. But effort (especially its intensity dimension) cannot reasonably be looked at in this way. Intensity of

---

[5] I distinguish here between the teacher's performance and teaching effectiveness. The latter depends not only on th teacher's performance but also on many factors outside the teacher's control, including the students' backgrounds and prior educational experiences, the resources with which the teacher has to work, the prescribed curriculum, and the quality of school leadership, among many others.

BEST COPY AVAILABLE

effort has no clear-cut upper bound (a teacher exerting maximum effort, taken literally, would have no time or energy for any other aspect of life). For all practical purposes, the effort continuum is open-ended, and most teachers always have the option of moving upward or downward along it. The proposition that there is room for increased effort does not depend in any way on the belief that some teachers now do unacceptably little work (although it is surely true that some have "burnt out" or "retired on the job"). Even if every member of the teaching force already exceeded minimum standards of effort, the possibility that incentives would motivate some teachers to do more could not be ruled out.

Apart from effort, a teacher's performance depends on the capabilities--knowledge, skills, "competencies," etc.--that he or she brings to the job. To say that a teacher is "working at maximum potential" implies, therefore, not only that the teacher is making the best use of the capabilites he or she already possesses but also that those capabilities cannot expand. But a teacher's capabilities are largely determined by what he or she has learned about teaching--which is to say, by the cumulative effects of preservice and in-service training, experience, on-the-job learning, and informal self-improvement activity. Only if one believes, therefore, that these opportunities for learning have been exhausted--that most teachers already know nearly all there is to learn about teaching children or are incapable of learning more--can one reasonably maintain that there is no room for individual teachers to improve.

The assumption that there is room to improve is embedded in the current practices and institutions of the educational system. It is implicit in the emphasis placed on in-service training in many school

BEST COPY AVAILABLE

systems; in the widespread practice of formative evaluation, wherein principals or supervisors observe teaching behavior, identify problems, and help teachers upgrade their techniques; and in the financial incentives for post-graduate education built into nearly every school district's teacher salary schedule. None of these devices would make sense if it were believed that most teachers have already maximized their skills.

Finally, many teachers could improve their performance (in the eyes of district authorities) by matching their allocations of time and effort more closely to district priorities. In the decentralized, loosely supervised instructional settings of most American schools, individual teachers have considerable control over the amounts of time allocated to different areas of the curriculum and the effort devoted to different categories of children (Goodlad, 1984). The actual pattern of time allocations in any classroom is likely to reflect some compromise between the official priorities of the school system and the preferences and interests of the classroom teacher. Here, once again, variability—in this case, of time allocations—is prima facie evidence that there is room for improvement. That is, teachers whose allocations do not correspond fully with school system goals can raise their performance by reducing the disparities. Neither increases in aggregate effort nor increases in teachers' skills are necessarily required for this form of improvement to occur.

In sum, to deny that there is room for improvement, one must make a series of drastic and implausible assumptions: that few teachers can work harder or longer than they do now (despite the fact that some work much harder and longer than others); that most teachers have exhausted their capacities for learning or that there is nothing useful about teaching

BEST COPY AVAILABLE

to be learned; and that most teachers have already optimized their time and effort allocations and teaching methods. It is difficult to take seriously that any one of these reflects reality and hence difficult to deny that most teachers have room to improve.

Capacity and Freedom to Improve

Incentives can make a difference only if there are actions that individual teachers can take to improve the effectiveness of their teaching. The possibility of action depends, first, on capacity—the ability of teachers to learn what will raise performance and then to translate that knowledge into practice; and second, on teachers' freedom to make the necessary changes in instructional behavior. Thus, incentives will fail if (a) teachers are too limited, intellectually or otherwise, to learn to teach better, (b) there is nothing useful about better teaching for teachers to learn, or (c) teachers are prevented from upgrading their teaching by forces outside their control.

To maintain that teachers have no capacity at all to improve is far-fetched, since some steps available to teachers require nothing new but merely more of the same. Many teachers have the options, already noted, of increasing the amounts and intensities of their inputs into instruction or reallocating their efforts among categories of pupils and areas of the curriculum. These avenues of improvement do not depend on either improved skills or the availability of better instructional techniques.

Whether there is a widespread capacity for qualitative upgrading is a more interesting question. To conclude that there is not, one must

BEST COPY AVAILABLE

believe that learning to do better is precluded for most teachers by some immutable or hard-to-change characteristic, such as limited innate ability or inadequate college preparation. Or, alternatively, one might believe that teachers are capable of learning but there is nothing worthwhile to learn--i.e., that there is no valid or useful body of transferable information regarding what constitutes good teaching. Either belief flies in the face of the aforementioned faith, widely shared by teachers, administrators, and teacher educators, in the powers of formative evaluation and in-service training. It is possible, of course, that the conventional wisdom of the education profession is wrong and that leeway for improvement is minimal, but there is hardly evidence to support so nihilistic a position.

In this respect, the teacher incentive movement is an optimistic one. A belief in the power of incentives is incompatible with the view that the present inadequacies of teaching are mainly attributable to the low caliber of teachers. It rests instead on the premise that the reward system rather than the human material is deficient--that there is unrealized potential in the present teaching force, which could be realized if the appropriate motivators were supplied.

Whether teachers are free to respond constructively to incentives is a more complex question. Clearly, there are bounds on the changes that individual teachers can make. Teachers are often not at liberty, for instance, to formulate their own curricula, choose their own textbooks, or decide, without restriction, what instructional strategies or methods to employ. Neither, on the other hand, are teachers typically so tightly constrained that they must follow rigid, preprogrammed routines in their classrooms. In most school districts, curricular and methodological

BEST COPY AVAILABLE

guidelines are sufficiently general to embrace significant variations in teaching techniques and styles, and the lack of detailed supervision of instruction affords teachers wide discretion to modify their methods and allocate their efforts as they see fit. Thus, most teachers would be free within limits to put newly developed skills to use in the classroom.

It should be recognized, however, that just because teachers have some freedom to respond does not imply that the scope of their autonomy is adequate or optimal. Freedom to experiment and innovate—to try, fail, and try again—is so essential a counterpart of performance incentives that it would almost surely need to be broadened and institutionalized to obtain the best results under an incentive system. Autonomy due to loose supervision is not the same as the autonomy that could be officially conferred as part of a system of performance-based rewards. I believe that the issue of teachers' freedom is a very real one, therefore; one whose implications have not been adequately considered in the context of the incentives debate; and one that should be dealt with more explicitly when incentive systems are proposed.

Pay and Promotion as Motivators

Incentives would work, if at all, by strengthening teachers' motivation to perform, and so to believe that incentives can be useful one must also believe that (a) under the existing reward system, adequate motivation is lacking, and (b) performance-contingent pay and promotion are potential motivators. For instance, if most teachers were already driven to their limits by a passion for educating children, there would be little reason to offer them extrinsic rewards. Once one has conceded that there is

BEST COPY AVAILABLE

room for improvement and that teachers have the capacity and freedom (within limits) to improve, however, the question of whether motivation is the missing factor becomes one of semantics rather than substance. The essential point is that improvement is a matter of the individual teacher's choice. Whether one attributes the failure to make performance-enhancing choices (absent incentives) to "lack of motivation," "low morale," "frustration," or other attitudes or mental states is immaterial. What counts is that there are steps that teachers can take that many teachers have not taken in the absence of performance-contingent rewards but that they might conceivably take when such rewards are introduced.

I discussed at the beginning of this chapter the behavioral assumptions underlying the belief that performance-based pay and promotion may serve as the missing motivators for teachers. To an economist, these assumptions seem so thoroughly unremarkable that it is surprising to find them seriously contested. After all, to believe that most teachers would not even try to alter their teaching behavior if more pay were offered for good performance one must assume some rather extraordinary things: that teachers assign no value whatever to higher standards of living for themselves or their families (their marginal utility of income is zero); or that they assign essentially infinite value to any cost, exertion, or disruption of habit required to earn rewards. To assume either is to make teachers a species apart, outside the economic sphere, and without the wants and needs that cause other people to work.

But that teachers do care how much they are paid and are willing to exert themselves to obtain higher pay is amply confirmed by multiple forms of behavior: teachers bargain collectively, go out on strike, and

BEST COPY AVAILABLE

initiate other job actions to win higher salaries; lobby vigorously for higher pay before state legislatures and local school boards; apply in larger numbers to teach in districts where pay levels are higher; and respond to the one clear-cut financial incentive in traditional salary schedules by accumulating pay-enhancing course credits and advanced degrees. There is also evidence that the decline in the numbers of talented college graduates seeking to enter teaching is due in part to low relative salaries in the profession and that the flow of good teachers out of teaching is attributable partly to the greater economic rewards in other fields. None of this would be true if teachers were indifferent to economic rewards.[6]

The above notwithstanding, some observers seem to find troubling the suggestion that teachers might respond to economic rewards, and some have gone through intellectual contortions to deny it. It has been asserted, for example, that teachers value "only" such intrinsic rewards of teaching as the satisfaction of helping children learn and consequently cannot be induced to change their teaching behavior by offers of extrinsic rewards. Setting aside the implausible image of 2,500,000 selfless, ascetic teachers who care about children so much that they are indifferent to their own children's welfare, one can only note that the changes incentives are supposed to bring forth are precisely those that would enhance the learning teachers supposedly cherish. Earning performance-contingent pay does not mean sacrificing intrinsic rewards; if anything, children would learn more, and whatever pleasure that brings teachers would presumably be enhanced.

---

[6]Note also an implication that those who portray teachers as indifferent to economic rewards rarely confront: such indifference implies that salaries could drastically be cut without adversely affecting teaching performance.

BEST COPY AVAILABLE

In a classic non sequitur, studies reporting that teachers do not rank salaries high among the rewards of teaching (the source of this finding: teachers said so!) have been cited as evidence that performance-contingent pay will not induce teachers to perform. There is evident confusion here between the rewards that induce persons to become and remain teachers and those that influence on-the-job performance. That teachers may have traded off opportunities for higher salaries in other occupations to obtain the nonsalary benefits of teaching (which, by the way, include such economic benefits as job security and the 180-day work year) does not imply that once employed in teaching, with those other benefits secured, teachers would turn down the opportunity to earn more by teaching more effectively. The trade-offs in the two situations are different. The "price" of becoming a teacher consists of the foregone opportunities in other fields, including (for some teachers) the prospects of higher pay.[7] The price of earning performance-contingent rewards includes the extra time, energy, and effort that teachers would have to devote to their jobs to meet the performance criteria. That teachers have been willing to trade off pay for other benefits in the former case in no way implies that they would be unwilling to work for pay in the latter.

I conclude, therefore, that one cannot plausibly argue away the proposition that performance-contingent rewards will elicit higher performance by citing alleged peculiarities (or virtues) that set teachers apart

---

[7]It should not be assumed that all teachers have foregone higher paying opportunities elsewhere. For some, the most likely alternative may be clerical work rather than other professional employment. Moreover, it appears that for women, at least, the salaries paid in teaching are in the mid-range of those earned by employed college graduates.

BEST COPY AVAILABLE

from the rest of the labor force. It may well be that such rewards would not affect performance enough to justify their costs, but to argue that is quite different from claiming teachers are indifferent to rewards. The real issue--not whether but to what degree incentives can motivate teachers to perform--will not be resolvable until incentives (of various sizes and shapes) are established and tried.

INCENTIVES AND TEACHER TURNOVER

There is much broader agreement that economic incentives can affect teachers' behavior in the labor market than that they can affect behavior in the classroom. Even those who believe that merit pay will not raise the performance of existing teachers concede that salaries help determine who enters and who remains in teaching (e.g., Goodlad, 1984; Rosenholtz, 1985). Usually, however, the effects on entry and retention of only the level of teacher pay are considered. For the purpose of this paper, although the effects of changes in pay levels are of some interest, the potential impact of performance-contingent pay is more important. In this section, I consider how changing the structure of rewards, including making them performance-contingent, is likely to affect teacher quality via altered turnover and retention; in the following section, I consider how the same changes are likely to affect quality by altering the characteristics of entering teachers.

The economic argument that incentives can affect turnover depends on two propositions: (1) that teachers' desires to remain in or leave teaching are determined, in part, by the relative rewards available to them in teaching and other occupations, and (2) that significant numbers

BEST COPY AVAILABLE

of teachers have viable options (career or noncareer) outside the teaching field. Both must be true for performance-contingent rewards to have the intended quality-enhancing effects on teacher turnover rates.

That relative rewards influence retention is merely an extension of the earlier premise that teachers are not indifferent to extrinsic rewards. To avoid any misunderstanding, however, I spell out in a little more detail the applicable economic model. Each currently employed teacher confronting the decision whether to stay or to leave can be viewed as facing a "package" of benefits in teaching, consisting of both intrinsic and extrinsic rewards (including salaries), and a similar package of benefits in alternative occupations.[8] The decision hinges on which total package is preferable. In general, neither salary nor any other single benefit unilaterally dominates, so the availability of higher salaries outside teaching does not, in itself, necessarily tip the balance against remaining a teacher. Not all teachers assign the same relative weights to extrinsic and intrinsic rewards, however. Some would make large financial sacrifices to enjoy the satisfactions of teaching; others would make only smaller sacrifices. One can think of teachers as ranged along a continuum with respect to the degree to which they would trade off opportunities for greater economic rewards elsewhere to continue as teachers. Doubtless there are some to whom the intrinsic benefits of teaching are so great that they would not consider other jobs even in response to major declines in relative pay. But at the same time, there are teachers

[8]To simplify the discussion, it is convenient to think of the liabilities associated with teaching and other occupations (e.g., unpleasant working conditions) as negative benefits, or "disbenefits."

BEST COPY AVAILABLE

nearer to the margin for whom more moderate pay changes, upward or downward, could swing the decision for or against departure. That some teachers do leave teaching for other kinds of work, sometimes stating explicitly that inadequate pay made continuation in teaching unattractive, is an indication that there are members of the teaching force for whom fractional increases or reductions in pay would be decisive.

Little if anything seems to be known about the alternative opportunities available to teachers. That some people do switch from teaching without joining the ranks of the permanently unemployed indicates that alternatives exist, but how they are distributed among other professional fields, clerical occupations, service work, or other job categories is a mystery. The existence of such opportunities is obviously critical to the efficacy of any incentives aimed at teacher retention. If teachers had few nonteaching alternatives, financial incentives would be of little use either for driving poor teachers away or inducing good teachers to stay. That teachers do in fact leave the profession in considerable numbers, expecially at early stages of their careers, suggests that a general shortage of alternative opportunities is not the problem. What is important, however, is how alternative opportunities vary between teachers ranked high and low on the teaching performance scale.

Would an across-the-board change in the relative level of teachers' salaries raise, lower, or leave unchanged the average quality of retained teachers?[9] The answer depends, obviously, on the relative responsiveness

[9]For the moment, I speak only of retained teachers, setting aside the question of how changes in pay affect the quality of new entrants into teaching and hence the average quality of the whole teaching force. Those questions are discussed separately below.

BEST COPY AVAILABLE

to salary changes of teachers with above-average and below-average performance. If good teachers and poor teachers responded similarly, and if their opportunities outside teaching were the same (i.e., uncorrelated with teacher performance), a pay increase would reduce turnover rates but average quality would not change. On the other hand, if response rates or opportunities were unequal for low-performing and high-performing teachers, a change in the level of pay would mean a change in the average quality of retainees.

In the absence of empirical findings on the turnover rates of more and less proficient teachers, one cannot be sure of the net effects of changes in turnover on quality. There is some reason, however, to suspect that they might not be neutral. A reasonable assumption is that some of the traits associated with good teaching (general intelligence, verbal ability, subject matter mastery, etc.) are associated with proficiency in nonteaching jobs as well and hence with expected economic rewards in other fields. If so, this implies that good teachers must be less willing, at the margin, to trade off the benefits of teaching for higher pay, which in turn suggests that the turnover rate of good teachers may be less responsive to changes in teachers' salaries than the turnover rate of poor teachers.[10] There is some danger, therefore, that general pay increases could reduce the turnover rates of poor teachers more than those of good teachers, thereby decreasing the average quality of the

[10]That is, if good teachers have better job prospects outside teaching than do poor teachers, they are sacrificing more ("paying a higher price") to remain teachers, which suggests that they must value especially highly the particular benefits of teaching that are not obtainable in other occupations.

teaching force. This is admittedly speculative, and the opposite could turn out to be true. It is by no means certain, for instance, that talent in teaching is positively correlated with talent or potential earnings in other lines of work. Nevertheless, even a speculative argument suffices to make the point: one should not take for granted that reducing turnover would be beneficial, or even that it would not detract from teacher quality.

Now, consider the effect on teacher turnover, and the consequent effect on quality, of introducing performance-contingent rewards. A pay-for-performance system, by definition, would raise the relative salaries of good teachers and lower the relative salaries of poor teachers. Continuing to teach would become more attractive to the former and less attractive to the latter; better teachers would tend to remain in teaching longer and poorer teachers to leave sooner; and average quality would rise. The effects on the retention rates of good and poor teachers might not be symmetrical, however. If lower-quality teachers have limited opportunities outside teaching (as one would expect if productivity in other jobs is correlated with proficiency in teaching), even relatively sharp declines in their relative pay as teachers might not induce them to depart at significantly higher rates. The effects of performance-contingency, then, would be that the average quality of retainees would rise, due mainly to increased retention of good teachers, but the overall turnover rate would fall. The concommitant reduction in the number of openings for new entrants could have adverse implications for quality, possibly even offsetting the positive effects of the improvement in the quality of retainees. This is an important and generally unappreciated point, and one that I will pursue more fully below.

INCENTIVES AND THE QUALITY OF NEW ENTRANTS

The idea that higher salaries can attract talented new entrants into teaching has been seized upon enthusiastically and used to justify proposals for across-the-board increases in teachers' pay. In comparison, the effects of performance-based rewards on new entrants have largely been ignored. In this section, I consider higher levels of rewards and performance-based rewards together, asking how either or both might affect the quality of new recruits.

There is little doubt that raising salaries would attract more high-quality applicants for teaching positions. This conclusion depends on nothing more than the weak assumption that there are some new college graduates (and perhaps already-employed persons considering switching to teaching) for whom the level of pay can tip the balance for or against applying to teach. Essentially the same model as outlined in the foregoing discussion of retention applies here. Talented new graduates (and other prospective recruits) can be thought of as distributed along a continuum with respect to their willingness to trade off other opportunities in order to teach. Some may want to teach so much that they would apply at almost any level of pay; others would not consider teaching under almost any conditions; but those in between might or might not apply, depending on the relative rewards. Any improvement in rewards would draw some of the latter (those "sitting on the fence") into the applicant pool. Putting the issue in more practical terms, improved opportunites elsewhere (especially for college-educated women) have made teaching relatively less attractive, reducing the numbers of talented people who

apply. Higher teacher pay scales can help to offset this trend, inducing more of the able once again to consider teaching careers.

What is often taken for granted in discussions of teacher quality—in my view, mistakenly so—is that attracting more high-quality applicants necessarily leads to hiring more high-quality recruits. Whether the best candidates would be hired depends, however, on how good school districts are at predicting applicants' teaching performance. Salary increases may well make the selection process more difficult. Along with the more-talented applicants drawn by higher pay would come larger numbers of mediocre and average applicants than currently apply. Quite possibly, the average quality of candidates would fall, even though there would be more high-quality individuals in the applicant pool. Thus, the ability of employers to discriminate is critical. I expect that school districts, by and large, would discriminate successfully, based mainly on the belief that certain relatively easy-to-assess characteristics, such as verbal ability, academic performance in college, and proficiency in student teaching, correlate with ability to teach. Proposals for tightening certification requirements, already enacted in some states and pending in others, may facilitate the screening process.[11] But there is nothing automatic or guaranteed about the outcome. That talented persons respond to economic rewards by applying to teach is only a necessary condition, not a sufficient condition, to ensure that some of them actually become members of the teaching force.

---

[11]See U.S. Department of Education (1984) and Education Week (1985) for state-by-state tabulations of steps being taken to upgrade the requirements for teacher certification.

I turn now to the less-explored aspect of incentives for new entrants, the prospective effect on the quality of entering teachers of introducing performance-based rewards. Ideally, in a world of perfect information, such rewards would contribute to the quality of entrants in the following manner. Each potential applicant, aware of the performance-contingent nature of pay and promotion, would evaluate the economic attractiveness of teaching in light of his or her assessment of his or her own potential as a teacher. Those potential applicants expecting to do well as teachers would find teaching more rewarding under a regime of performance-based pay, while those expecting to be average-or-below teachers would find it less inviting. Thus, the former would tend to apply in larger numbers and the latter in smaller numbers, and the average quality of applicants would rise.[12]

But assuming that potential applicants can predict their own performance as teachers is hardly realistic. Many prospective teachers have no way of knowing before actually starting to teach (and possibly not even for some time thereafter) whether they will turn out to be above-average or below-average performers. What they do know, however, is that under a performance-contingent system, their future level of pay is uncertain. Other things being equal, uncertainty is likely to act as a deterrent, discouraging some prospective teachers, including some prospective good teachers, from seeking teaching positions. To illustrate, consider how prospective applicants might feel about applying to a district that pays, say, $25,000 per year with certainty, as opposed to one that pays $25,000

[12]The implicit comparison here is between systems with and without performance-contingent pay but with the same average salary levels.

per year on average, but with the actual salary of each teacher varying from $20,000 to $35,000 according to performance.[13] An applicant confident of being a top performer would presumably prefer the latter, but an uncertain, risk averse (and presumably more typical) prospective teacher might well prefer the certain $25,000 to the probabilistic alternative. Thus, even though both districts pay the same average salary, applicants might treat the district with performance-contingent salaries as if it paid less. In other words, the uncertainty inherent in performance-based pay detracts from the value of the salary package.

Of course, raising the level of pay can overcome the negative effects of uncertainty. Referring to the foregoing example, it may be that raising the average pay level under the performance-contingent pay scheme to, say, $27,000 would offset the negative effects of uncertainty in the mind of the typical risk-averse potential teacher. But to say that a certain $25,000 salary is equivalent to a $27,000 uncertain, performance-contingent salary merely underscores the point I am trying to make, which is that performance-contingency per se may be a liability in terms of attracting new teachers.

In practice, it seems that virtually every recent state proposal for a merit pay or career ladder system would combine an across-the-board pay increase with the performance-contingent rewards. Whatever the motive for such combinations, the effect should be to offset the disincentive to some applicants that would otherwise be created by performance-contingency alone. The resulting combination of higher and performance-contingent

---

[13]To simplify the example, I leave out the pay increases that accrue to teachers as a function of seniority.

pay should offer inducements to both the more confident and the risk-averse applicants, thereby adding to the chances of hiring higher-quality recruits.

HIGHER SALARIES, MERIT PAY, AND CAREER LADDERS AS INCENTIVES FOR QUALITY TEACHING

Having considered the likely effects of incentives on existing teachers' performance, retention, and recruitment, I now draw on the findings to assess the main incentive mechanisms on the current policy agenda. Foremost among these are career ladder plans, especially those featuring master teachers, and, to a lesser extent, proposals for merit pay. But before turning to these, I comment first on the straightforward proposal that quality can be improved by raising the general level of teachers' pay.

Higher Salaries

All the recent education reform reports that deal with the teacher reward system call for increases in the level of pay, either as a desirable reform in its own right or in conjunction with merit pay or career ladder plans. Most states now considering performance-based rewards are also considering, or have already approved, general pay increases. In addition, many states that have not yet acted on incentives have recently raised teachers' pay substantially or are contemplating doing so. In light of the foregoing discussion of teachers' responses, what can be said about the probable effects of such pay increases on teaching performance?

One important point is that a general pay increase creates no direct incentive for existing teachers to improve their performance. A general increase, by definition, raises everyone's pay without regard to any change in behavior. There are some indirect channels through which such

an increase might conceivably contribute to performance—for instance, by reducing teachers' needs to "moonlight" on other jobs to support their families and relieving them of function-impairing anxieties associated with low earnings—but there is no reason to think these effects would be significant. Most of the money devoted to an across-the-board increase would go to already-employed teachers, and most, therefore, would have no short-term quality-enhancing effects.

The effects of a general pay increase on entry and retention need to be considered together because the two interact. If the quality of entrants were the only consideration, a pay increase would almost surely be a quality-enhancing force. Apart from the reservation expressed above about the ability of school districts to select good applicants, it is reasonable to believe that better people will enter teaching if teaching pays more. The effect of higher pay on the quality of retainees is ambiguous for reasons discussed above. If all retention rates rose uniformly, the average quality of retainees would remain the same; if the retention rates of below-average teachers rose more, quality would decline. The main significance of the effect of higher pay on retention, however, is not that a skewing of retention rates would affect quality directly (although such an effect is possible) but rather that increased retention reduces the number of openings for new entrants. In this respect, an across-the-board pay increase works against itself. On one hand, it attracts higher-quality applicants; on the other, it reduces the number of spaces they can fill. Every poor or mediocre teacher who remains in the system because of the overall increase in pay occupies a position that might otherwise be filled by a promising new recruit. Thus, much

of the potential gain from attracting talented persons into teaching might be lost because of the retention effect.[14]

What, then, is the likely net effect of higher pay on quality? In the long run, it will almost surely be positive, despite the negative factors mentioned above. Gradually, existing teachers will be replaced by teachers hired under the higher-salary regime, and those teachers, by and large, will be of higher quality than would have been recruitable if pay had remained the same. In the short-run, however, any improvement in quality is likely to be small. Most of the expenditure for higher pay will go to already-employed teachers in a form that creates no incentive for better teaching. At the same time, the aforementioned retention effect will impede the inflow of higher-quality recruits. In the worst case, average quality in the short run could actually decline.[15] The key problem is that, by definition, a general pay increase is nondiscriminatory. It rewards both the good teachers the school system would like to retain and the poor teachers it would rather see leave. In consequence,

---

[14]It has sometimes been suggested that this problem could be avoided by increasing starting salaries only rather than salaries in general, but there are two reasons why this solution does not work. First, there is very little scope for raising the pay of new entrants without also raising the pay of those above them on the seniority ladder (it is clearly not feasible to pay teachers more in their first year than in their second). Second, it is unreasonable to suppose that prospective new teachers consider only how well they would be paid in their first year. What counts is the prospective stream of earnings during a teaching career, and raising the value of that stream implies raising salaries across the board.

[15]The possibility of a decline in overall quality would arise if, prior to the across-the-board increase in pay, the average quality of new recruits already exceeded the average quality of teachers departing from the system. Under that condition, the increased retention of relatively low-quality teachers and the reduction in openings for new entrants could more than offset the positive effect of improved entrant quality.

an across-the-board increase is a costly, low-leverage, slow-acting method of improving teacher quality, compared with alternatives that target resources more precisely on rewarding good performance.

Performance-Contingent (Merit) Pay

In contrast to general pay increases, which would increase quality primarily by attracting better new entrants into teaching, performance-contingent pay would affect mainly teachers already employed. It is plausible, for reasons discussed at length above, that two positive effects would be obtained: (1) the classroom performance of some significant number of teachers would improve, and (2) retention rates would be altered favorably, in that high performers would have a greater propensity to remain in teaching and low-performers a greater propensity to depart. However, certain major qualifications must be noted. First, the magnitudes of teachers' responses, both with respect to the numbers of teachers responding and the degree to which they alter their behavior is unknown and unknowable until direct evidence is accumulated. Second, the effectiveness of a merit pay system is certain to depend on many specific features of the system's design, including the methods of evaluating performance and structuring and apportioning rewards. These design aspects are covered in the following two chapters. Pending those discussions, I confine myself to the generalizations that merit pay plans, properly designed, have the potential to stimulate better teaching performance and to raise the average quality of members of the teaching force.

The most problematic aspect of performance-based pay is its effect on prospective new entrants. As explained above, the uncertainty inherent

in merit pay is especially great for those who have not yet begun to teach, and it could offset, wholly or in part, the attractiveness to promising potential teachers of performance-based rewards. Thus, a "pure" merit pay plan--one that leaves the average level of pay unchanged while linking pay to performance--might produce negative effects on recruitment that run counter to the positive effects on those already employed.

The kinds of merit pay plans we are most likely to see in practice, judging by current proposals in the states, are those that combine higher base salaries with performance-contingent rewards. Their effects are likely to fall in between those of conventional salary increases and pure merit pay. Higher base pay can offset uncertainty, averting possible adverse effects of performance-contingency on recruitment. At the same time, higher pay can be expected to weaken the beneficial effects of merit pay on teacher retention. Choosing the right balance is one of the more difficult problems of program design.

Career Ladders

To assess the newly popular career ladder schemes, one must first clarify the similarities and differences between a career ladder and merit pay. A career ladder subsumes merit pay, but its essence is merit promotion. Teachers rise from one rank to another by virtue of performance, and those who earn promotions receive higher salaries as well. But ranks and promotions have significance, other than honorific, only if they are tied to differentiated roles. In some "master teacher" or "mentor teacher" plans, those accorded such titles would have special out-of-classroom responsibilities for developing, supervising, and evaluating

other teachers and participating in curriculum development. In other plans, role differentiation seems more nominal than real. The "career ladder" rubric spans the whole range.

Without substantially differentiated roles, a career ladder amounts to merit pay plus nonmonetary recognition. Performance-contingent status reinforces the incentive of performance-contingent pay. I can offer no insights into the value of status, per se, as a stimulant to good teaching, except to observe that its potential should not be underestimated. Too many honors, awards, medals, certificates of merit, etc., are bestowed in the world, both in and out of education, to let one believe they do not count. Also, benefits more tangible than status may be associated with steps up the ladder. The first step means, in some plans, a transition from probationary to permanent teacher; higher steps may bring a greater voice in the affairs of the school. Though short of differentiated staffing, these are by no means insignificant rewards.

Where differentiated roles are substantial, as in the aforementioned master teacher plans, further considerations come into play. First, there is the effect of differentiated roles on the incentive to teach. The roles most often mentioned--staff development, teacher evaluation, and the like--are likely not to have universal appeal. To some teachers they would be rewards to pursue, to others burdens to avoid. The incentive effects would range from substantial to nil. Second, there is a criterion conflict. The teachers who perform best in the classroom are not necessarily those who would best carry out the new nonteaching roles. Hence, if teaching performance is the criterion, the wrong people may be selected for the master/mentor role, while if suitability for the role is the

criterion, teachers are unlikely to be rewarded for how well they teach. Third, there is the effect of nonteaching roles on teaching performance. If the best teachers are rewarded with nonteaching assignments, their time in the classroom will be diminished and their direct contributions to learning reduced. On the other hand, their indirect contributions, especially to other teachers' proficiency, might have an offsetting effect. In any event, it is clear that such career ladder plans must be assessed not only as systems of rewards but also as methods of reorganizing the instructional staff and the delivery of services within schools. All these considerations are closely linked to aspects of system design, such as the specifics of role differentiation and the criteria for teacher promotion, and I return to them in the discussions of design issues, below.

III. TEACHER INCENTIVES AND TEACHER EVALUATION

Any merit pay or career ladder system consists of two main components: a method of evaluating teaching performance and a method of linking the performance ratings to rewards.[1] There is a massive literature on teacher evaluation, but relatively little of it deals explicitly with the relationship between performance measurement and incentives.[2] That specific aspect of evaluation is the topic of this chapter. The equally important but much less analyzed issue of how performance ratings, once obtained, should be used to apportion rewards is discussed separately in Chapter IV.

Raising performance by rewarding performance presupposes an ability to measure performance correctly. The quality of performance measurement impinges on the effectiveness of incentives in a variety of ways:

- Valid, reliable, and fair performance measures are needed to guarantee that the "right" teachers are rewarded.
- Accurate performance measurement is essential to ensure that effective teaching behaviors will be encouraged, and undesirable behavior discouraged, by the incentive system.
- The quality of the performance measurement method determines, in many respects (to be spelled out in Chapter IV) how the rest of the incentive system can be structured.

In this discusssion, I focus on three issues concerning the relationship between performance evaluation and incentives: (1) the suitability of

---

[1]There may be other components also, such as staff development activities and differentiated staffing arrangements, but the two components mentioned are the ones that make up the "incentive" part of the system.

[2]Recent contributions that do relate performance evaluation to incentives include Darling-Hammond, Wise, and Pease, 1983; Wise et al., 1984; Hatry and Greiner, 1984; and Cresap, McCormick and Paget, 1984a).

existing teacher evaluation methods for use in a system of performance-based rewards, (2) the implications for incentives of current shortcomings in the art of measuring teaching performance, and (3) the prospects for better performance measurement in the future.

RELEVANT EVALUATION METHODS

The issue of how teaching should be evaluated is, of course, one that concerned educators long before the current furor over career ladders and merit pay. States and school systems have many reasons unrelated to incentives to want to know how well teachers teach, including the needs to monitor the quality of instruction, to certify probationary teachers for permanent status, to diagnose teaching problems, to help teachers improve their skills, to develop teacher certification requirements, and to upgrade teacher training. Numerous teacher evaluation methods and instruments have been developed, and some are used routinely by states and school systems around the country. There is a substantial menu from which to select measurement techniques.

As evaluation methods have proliferated, whole taxonomies have had to be created to sort out the various approaches. Scholars typically classify teacher assessment methods according to the dimensions of performance to be measured and the sources of evidence for judgments about teacher proficiency, and then, in greater detail, by the specific evaluation techniques, instruments, and/or estimation procedures used to arrive at performance ratings. For instance, Darling-Hammond, Wise, and Pease (1983), following Medley (1982), distinguish among evaluations of specific elements of the teacher's knowledge and skill (competencies), the overall

quality of the teacher (competence), th quality of teaching (performance), and the quality of student outcomes (effectiveness); while Millman (1981) classifies methods according to whether they relv on teacher interviews, competency tests, classroom observation, scudent ratings, peer review, student achievement, teacher out-of-class activities, and faculty self evaluation. Much has been written about the strengths, weaknesses, uses, and limitations of each evaluation mode.

Fortunately, many options can be eliminated at the outset as inherently unsuitable for use in a system of performance-based rewards. In the context of incentives, we are concerned, first of all, with summative rather than formative evaluation--that is, with determining how well teachers are performing rather than with helping teachers improve. This substantially narrows the range of acceptable evaluation methods. As pointed out by Wise et al. (1984),

> For purposes of accountability [summative evaluation], teacher evaluation processes must be capable of yielding objective, standardized, and externally defensible information about teacher performance. For improvement objectives [formative evaluation], evaluation processes must yield rich, descriptive information that illuminates sources of difficulty as well as viable courses for change. (p. 12)

Many evaluation systems currently used in the schools fit the latter description more closely than the former and consequently are of little value for dispensing performance-based rewards (Darling-Hammond, Wise, and Pease, 1983).

Further, within the category of summative evaluation, we require methods for differentiating superior or excellent teaching from average

performance, not merely for determining whether teachers are minimally competent. Since some of the more commonly used evaluation devices serve only the latter purpose (e.g., minimum-competency tests for teachers), the range of relevant assessment methods is further reduced.

Among the methods that remain, one fundamental distinction overshadows all others: that between methods of measuring the products or outcomes of teaching--how much students learn--and methods of assessing the process of teaching, or what a teacher does and knows. The significance of this dichotomy is twofold. First, the products of education--the child's learning and development--are valued as ends in themselves, whereas such input or process indicators as the teacher's subject-matter knowledge, pedagogical expertise, and classroom "competencies" are valued only as means to an end--i.e., only insofar as they are believed to contribute to educational results. Second, the methods of process and product evaluation differ dramatically. Process evaluation is direct, observational, and judgmental; product evaluation indirect, impersonal, and inferential. The former judges teaching in the classroom; the latter rates teachers by what happens to their students, without looking at teaching at all.[3]

Whether to evaluate teachers by student outcomes, teaching processes, or a mixture of the two is the main design issue concerning the teacher evaluation component. The specifics of evaluation methods, standards, instruments, schedules, etc. are all secondary in comparison. Accordingly,

---

[3]Strictly speaking, it is incorrect to equate process-oriented evaluation with evaluation based on classroom observation, since such other data sources as student ratings, parent rating, and interviews may also be part of the process approach. In practice, however, the process-oriented methods proposed for incentive systems all center around observation of classroom teaching by supervisors, experts, or peers.

after spelling out the applicable selection criteria, below, I focus in detail on the product vs. product choice.

CRITERIA FOR EVALUATING EVALUATION

The standard criteria for judging summative evaluation systems--validity, reliability, and unbiasedness--are relevant to the evaluation components of teacher incentive plans, although some need reformulation to fully apply. Also, certain more specialized criteria become germane when the evaluation results are to be used for determining pay and promotion. Before turning to the specific evaluation approaches, therefore, I review the main criteria for deciding whether a method is usable in a system of performance-based rewards.

Validity

Performance ratings used for apportioning rewards must, above all, be valid indicators of each teacher's contribution to the school system's educational goals. Assuming that student learning is the overriding goal, a teacher should be rated "superior" if and only if there is reason to believe that his or her behavior contributes to student learning to an above-average degree. The test of validity is predictive power. If performance ratings predict, or correlate with, teachers' contributions to educational outcomes for children, the rating method is valid; if they do not, no other attribute of the evaluation system can compensate for this fundamental flaw.

Implicit in this general definition of validity is the subcriterion of content validity. A teacher can reasonably be judged only for what students learn within that teacher's sphere of responsibility. Thus,

the performance standards applied to a particular teacher should pertain to the subject(s) and grade levels he or she teaches and the categories of learning specified in the curriculum. Moreover, where a teacher is responsible for multiple subject areas, as in most elementary teaching, ratings should reflect performance across the whole spectrum of subject areas and should weight the different areas appropriately (which is to say, in proportion to their importance to the school system's educational goals).

Reliability

Reliability refers to the consistency of the results when teachers are rated at different times, by different observers, with different instruments, or when working with different sets of pupils. The lower the reliability, the greater the uncertainty over how teachers rank and who should receive rewards. The reliability of a performance rating is a function, other things being equal, of the number of observations of process or outcome on which the rating is based. Thus, reliability can be enhanced by repetition, and an otherwise less-than-reliable method, such as classroom observation, can be made more reliable if used enough. Reliability is a major determinant of acceptability. A system is unlikely to win support unless it inspires confidence that ratings are nonrandom, reasonably stable, and truly reflective of each teacher's position on the specified performance scale.

Unbiasedness

To be unbiased, performance ratings should be unaffected, for better or worse, by relationships between rater and ratee and minimally dependent

on the subjective judgment of any individual. This calls into question, for example, any evaluation system that depends heavily on ratings of a teacher by his or her own building principal. The unbiasedness criterion also implies (as does the reliability criterion) that where subjective judgment is essential, as in most methods based on classroom observation, performance should be judged by multiple evaluators to minimize the impact of any evaluator's prejudices regarding either the individual being evaluated or that individual's teaching style. Unbiasedness also interacts with validity in that it depends on the selection of evaluation criteria and instruments that reflect the school system's educational goals but give no undue advantage to teachers by virtue of sex, age, race, ethnicity, or other personal characteristics.

Discriminating Power

A more specialized criterion of a performance rating system is the degree to which it is able to distinguish gradations of performance. This ability to discriminate can be thought of as an aspect of reliability, since it depends on the consistency with which distinctions can be made between teachers relatively close together on the performance spectrum. An example of a system with too low discriminating power to be useful for apportioning rewards is one that can distinguish reliably only between "unsatisfactory," and "satisfactory" teachers. A system with higher power might classify teachers into four, five, or more performance strata, each of which could then be associated with a different level of reward. As will be seen, the design of an incentive system can be significantly constrained by the fineness with which such distinctions can be made.

Universality

For the purpose of apportioning rewards, it would be ideal if all teachers, regardless of grade level of subject-area assignment, could be rated on a single performance scale, so that any teacher's performance could be compared with the performance of all others. Where such universality cannot be achieved, teachers must be evaluated within separate categories, which makes it more difficult to ensure that the best teachers are rewarded. The degree to which such categorization can be avoided is therefore a relevant consideration in judging an evaluation method.

Predictability by the Teacher

An incentive system is likely to be a better motivator if teachers are able to predict the performance ratings likely to result from their own teaching behaviors. Predictability reduces uncertainty, and uncertainty, for reasons already discussed, weakens incentives. A predictable relationship between behavior and performance ratings also helps teachers channel their self-improvement efforts. Therefore, the greater the predictability, the more likely that incentives will have their intended performance-enhancing effects.

Beneficial and Adverse Side Effects

Apart from providing the performance information needed to operate an incentive system, performance evaluation mechanisms may also have positive or negative side effects on the instructional process and the condition of the schools. Possible beneficial effects include reinforcement of state or district educational priorities and direct stimulation of improved teacher performance (i.e., the process of measuring and comparing

teacher performance might induce teachers to do better even in the absence of performance-contingent rewards). Possible adverse effects include distortion of the curriculum (e.g., undesirable "teaching to the test") rigidification of teaching methods, inhibition of innovative practices, and lower morale. These are only illustrative of the consequences that one should attempt to anticipate in assessing the various measurement approaches.

Cost and Burden

All the other characteristics of an evaluation method must be balanced against the costs and burdens it creates, since even an otherwise ideal evaluation system would be useless if it were too costly or difficult for a state or school system to operate. Cost, in this context, should be construed broadly. It encompasses not only the direct expenses of the evaluation process but also diversions of staff time and energy, instructional time lost by the students, and interference with the instructional process or the curriculum. For instance, systems requiring extensive and repeated classroom observation are likely to be very demanding of the time of evaluators, while systems based on student outcomes may require elaborate and specialized testing programs. How to measure performance adequately but at reasonable cost is one of the more difficult problems to be faced in designing a performance-based reward system.

THE PRODUCT APPROACH: CAN WE MEASURE THE TEACHER'S CONTRIBUTION TO LEARNING?

It is easy to see both why it would be desirable to rate teachers explicitly for their contributions to student outcomes and why it is

difficult to do so. The attraction is that a product-oriented evaluation focuses on the ends rather than the means of education, and thereby promises more valid performance ratings than a process-oriented approach. The principal obstacle to product-oriented evaluation also pertains to validity—namely, that it is not easy to separate the teacher's contribution from other influences on what students learn. There are other pros and cons as well. To deal with them systematically, I apply the foregoing criteria one by one to the product-oriented measurement approach.

Validity. To appreciate the validity problems arising from a product-oriented evaluation, consider the two steps needed to rate teachers according to their contributions to student outcomes. First, the relevant dimensions of student accomplishment must be measured, both at the beginning and the end of the evaluation period. Second, adjustments must be made for factors other than teacher performance that account for some of the differences in student progress among classrooms. Problems of validity arise at both stages, but those at the latter stage are far more severe.

The validity problems associated with measuring student progress are familiar to anyone ever peripherally involved with achievement testing. They are not problems peculiar to teacher evaluation, but they do have important implications for any system that rewards teachers for what students learn. One problem of content validity is that well-established, broadly applicable, and accepted outcome measures do not span all the relevant areas of learning but are concentrated mainly in such basic skills areas as reading, language, and mathematics. Even at the elementary level, one cannot judge teachers fairly by progress in basic skills alone, and at the secondary level, teaching basic skills is peripheral to most

teachers' assignments. Consequently, valid evaluation of the outcomes of teaching would require much broader-ranging achievement testing than is now the practice in most states and school systems. A second content validity problem is that standard achievement tests are unlikely to reflect the full range of instructional goals in their subject areas. In particular, they are likely to slight the learning of higher-order skills that presumably follows from superior teaching. Thus, even where the relevant subject areas appear to be "covered" by existing tests, it cannot be taken for granted that the products of teaching are being adequately or completely measured. In addition, other kinds of threats to validity can arise from student turnover and absenteeism, nonuniform conditions of testing, and even deliberate manipulation of the testing process by teachers or students. Thus, there are a number of impediments—not insurmountable but also not negligible—to the use of pupil achievement scores as the basis for rewarding teachers.

But the problems of measuring student achievement are minor and manageable compared with those of attributing achievement gains to teachers. There is no doubt whatsoever that much of the variance in pupil performance gains among classrooms is due to factors other than the quality of teaching, and hence that such factors must be taken into account ("controlled for") to get valid estimates of each teacher's contribution to educational results. Most important among these nonteacher factors are the characteristics of the students themselves: their abilities, prior educational experiences, economic circumstances, home environments, interests, and attitudes; the presence of "problem" or disruptive children; and, perhaps most important, what students have learned and what styles of learning

they have developed prior to entering a particular teacher's class. Also relevant are the resources available to each teacher (e.g., supporting staff) and a variety of school characteristics and external-to-the-classroom circumstances not under the teacher's control. It would be neither valid nor fair to compare pupil progress in different teachers' classes without somehow taking these factors into account.

In theory, it is possible, using multivariate statistical methods, to control for the factors other than teacher proficiency that cause student achievement gains to vary among classrooms. Millman (1981) explains how this can be accomplished by using analysis of covariance. This method (or the analogous multiple regression method) yields adjusted achievement gain scores, which, in essence, are statistically based predictions of the gains each teacher would have produced with a "typical" class in a "typical" teaching situation. The adjusted scores, rather than the original raw scores, would then be used to determine which teachers deserve performance-based rewards.[4]

I think it unlikely, however, that such statistical methods would be deemed acceptable as the basis for apportioning merit pay and promotions. The methods themselves would be incomprehensible to most of those affected

---

[4]There are actually two different ways to use the results of a regression analysis or analysis of covariance to estimate teachers' contributions to the observed achievement gains. One is to use the statistical model to estimate an "expected" average gain for each class (the gain expected, given student characteristics and other nonteacher factors, under an "average" teacher), and then to compare it with the actual gain. The difference between the actual and expected gains is then attributable to the teacher's performance. The alternative is to estimate a "teacher effect" on achievement gains for each teacher being compared by including a set of teacher dummy variables in the statistical model. In general, the two procedures will not yield identical results.

and difficult to justify or defend in public. Moreover, although such methods can correct for some of the differences in teaching conditions among classrooms, no method can take into account the full array of relevant factors. Any teacher with a special situation--several disruptive students, say, or an unusual mix of abilities--could complain, with justification, that the circumstances of his or her classroom had not adequately been considered. Moreover, it would soon become evident that any statistical adjustment procedure necessarily leaves much to the statistician's discretion (e.g., exactly which control variables to include and how to measure them), and that itself might be enough to cast doubt on the results.

In practice, therefore, outcome-based evaluation may depend on the validity of simpler, more comprehensible methods of assessing teachers' contributions to student learning. An example of a straightforward, nonstatistical approach is the following: rate teachers according to the gains made by their students during the period in question relative to gains of the same students in earlier years (or, even simpler, relative to the same students' initial levels of achievement). The rationale for this procedure is that initial achievement levels or prior rates of gain serve as proxies for expected gains by the same pupils. Consequently, comparing actual gain against expected gain measures the amount by which a teacher exceeds or falls short of expected performance.

Of course, adjusting only for prior achievement or achievement gains takes no account of the special circumstances that can render performance gains noncomparable across classes. To deal with such situations, it would be necessary either to adjust performance ratings on a case by case basis, which would introduce an undesirable element of subjectivity,

or to rely on replication of the measurement process. For instance, if teachers were assessed on the basis of their classes' performance during, say, four different semesters, there would be less reason for concern over special situations in any one period. (Note also that at the high school level, a rich body of data can be assembled by collecting data on achievement gains in all the classes a teacher teaches during each semester or school year.)

In sum, there are approaches that could yield reasonably valid, albeit far from perfect, estimates of teachers' contributions to students' learning. How these approaches would work out and which would prove to be acceptable in practice is unknown; but there is certainly no reason to assume at the outset that valid outcome-based evaluation is infeasible.

Reliability and Unbiasedness. These are the strong suits of the product-based evaluation approach. The great advantage, with respect to reliability and unbiasedness, of measuring student outcomes rather than teaching processes is that no subjective appraisals of teachers' classroom behavior are required. The method relies on objective pupil performance data (test scores) and on predetermined procedures for adjusting such data, as described above. At no point does an individual evaluator, such as a school principal, have to offer an opinion about a teacher's proficiency. Thus, the potential for bias and favoritism that so concerns teachers when performance-based rewards are proposed is eliminated by the impersonal nature of evaluations based on pupil gains.

The main threats to the reliability of such evaluations are those stemming from reliability of the achievement tests themselves and those stemming from the gain adjustment method. The former are likely to be

minor because the achievement gains in question are class averages rather than gains of individual pupils. The latter could be more serious because of the aforementioned difficulty of taking adequate account of variations in conditions among classes; but here too, replication and reliance on averages over multiple classes and time periods can mitigate the problem.

Discriminating Power. Evaluation methods based on student outcomes also rate highly in the ability to differentiate among multiple levels of teacher performance. Their discriminating power depends on the accuracy (reliability) of the underlying student outcome measures and on the statistical error introduced in the process of adjusting for factors other than teaching performance. Almost certainly, the adjusted student learning data will support more detailed distinctions than could be made on the basis of observations of classroom teaching.

Universality. There is some ambiguity about the applicability of product-based performance ratings across the range of teaching assignments. Obviously, students' gains in achievement in different subjects and/or at different grade levels are not directly commensurable. One can not, for example, measure gains in reading achievement in a third-grade class and gains in algebra achievement in a high school class and decide, by comparing those gains, whether the third-grade teacher or the high school mathematics teacher did a better job. What one can do, however, is to compare teachers of dirferent grades and subjects according to achievement gains in their classes relative to gains in other classes of the same kind. If one found, for example, that the third grade teacher has produced reading gains 110 percent as great as those produced by the average third-grade teacher, while the high school mathematics teacher has produced

gains only 90 percent as great as the average in that category, one could say that the third-grade teacher ranks higher in relative performance. In this relative sense, all teachers can be rated on a single performance scale.[5]

It should be noted, however, that relying on measures of relative performance within grades and subject areas constrains the design of the incentive system in a different respect; namely, it requires that teachers be compared within large enough units so that there are sufficient observations of performance at each grade level and in each subject area. The half dozen high school mathematics teachers in a small school district cannot simply be compared against one another, for example, because there is no reason to believe that their performance represents high school mathematics teaching in general. Instead, one would want an external standard, say, the average performance of high school mathematics teachers in the state, against which each teacher could be compared. But this need for a broad base of comparison implies an equally broad requirement for uniformity in evaluation instruments and methods. Exactly how broad depends on the type of teacher in question. There may be enough third-grade teachers in a medium-size or larger school district to make a wholly self-contained performance assessment feasible, but it is likely that more specialized categories, such as high school physics or music teachers, would have to be dealt with statewide.

---

[5]An important implicit assumption in the relative performance comparison method is that the average levels of performance in all subcategories of teaching are the same. If, for example, high school mathematics teachers were better performers, on average, than third-grade teachers (whatever that means), the relative performance method would unfairly favor the latter.

Predictability by the Teacher. The outcome-based approach rates relatively high in predictability because those being evaluated can be informed in advance of what results will produce what ratings. Also, each evaluee can monitor pupil achievement during the semester or school year, so his or her performance rating will not be a surprise. In particular, teachers can be advised of the specific norms (expected achievement gains) pertaining to their particular students and hence of the gains needed to earn a superior rating. However, the teacher cannot necessarily relate these achievement targets to his or her teaching behavior in the classroom, since there is no guarantee that any particular teaching approach or level of effort will generate a particular rate of student learning.[6]

Side Effects. An evaluation system based on pupil learning has a number of potential side effects, some beneficial and some adverse. On the positive side, the generation of extensive new data on pupil progress, coupled with the setting of implicit performance norms for each teacher's pupils, may itself stimulate more effective teaching, even in the absence of performance-based rewards. Also, teachers' knowledge that their rewards will depend on pupil progress in specified areas may help to enforce compliance with the curriculum and with the priorities officially assigned to different subjects of instruction. On the negative side, heavy reliance on achievement testing could distort the content of teaching. Teachers might be motivated to emphasize unduly those areas of the curriculum which count toward evaluations (assuming, as one must, that coverage

[6]In addition, a particular performance rating does not guarantee a particular reward, since there may be constraints on the number of performance-contingent pay raises or promotions that make the performance thresholds for rewards uncertain. The effects of quotas on incentives are discussed in Chapter IV.

will be incomplete). There is also likely to be extensive "teaching to the test"—a phenomenon that could be either desirable or undesirable depending on how well the tests reflect the full range of instructional goals. In addition, the requirement for uniform testing across large units, possibly including entire states, could have an inhibitory effect on curriculum diversity and local autonomy to shape programs.

Costs and Burdens. Finally, as to the costs and burdens of outcome-based evaluation, it is clear that several new costs would have to be incurred. Considerably more testing would be required than would otherwise be done (although there is a tendency toward more statewide testing even in the absence of outcome-based evaluation); record keeping and data processing capabilities would have to expand; and the analytical capacity would have to be created and maintained to produce adjusted gain scores on a regular basis. More time of teachers and pupils would be spent on testing, at the expense of instructional and other activities. It is worth noting, however, that the largest item of cost associated with more traditional evaluation methods would not have to be incurred--namely, the cost of large amounts of professional time spent in classroom observation of teacher performance. Compared with that item, the costs of outcome-based evaluation are likely to be relatively modest.

THE PROCESS APPROACH: CAN WE TELL HOW GOOD A TEACHER IS FROM WHAT THE TEACHER DOES?

Although there are nonobservational methods of obtaining data on what teachers do in their classrooms, process-based evaluation, in practice, is virtually synonymous with evaluation by means of classroom observation. Classroom observation, most often by the building principal, is by far

the most commonly used evaluation method in elementary-secondary education. It is also the main method proposed in nearly all the recently developed state incentive plans. Specifically, various combinations of evaluation by building principals, peers (often master teachers), or outside experts are featured in the plans recently enacted or proposed in Florida, Tennessee, Texas, and Delaware (U.S. Department of Education, 1984; Southern Regional Education Board, 1984). To assess process-based evaluation, therefore, is essentially to assess classroom observation as a means of judging teaching performance.

Validity. There are a number of well-known, serious threats to the validity of performance ratings based on classroom observation. I deal first with the most fundamental difficulty, the possibility that the basic assumptions underlying classroom observation are unsound, and then turn to some of the more specific problems that arise in practice.

The premise underlying observation-based evaluation is that certain specific, known, observable teacher behaviors are systematically related to teaching effectiveness. From this starting point, one arrives (after a certain leap of logic) at the proposition that one can infer a teacher's effectiveness by observing what the teacher does (or, more generally, what takes place) in the classroom, without having actually to measure what students learn. Opinions on this matter are sharply divided, sometimes along disciplinary lines. Some teacher evaluators, teacher trainers, and researchers, mainly from the educational psychology tradition, claim to have identified specific, behaviorally definable teaching "competencies," which, they say, are associated with student learning. Other researchers, including most social scientists who have studied educational effectiveness,

seem to believe, to the contrary, that (a) we have only tentative and fragmentary knowledge, at best, of how teaching behavior correlates with educational results, and (b) effective teaching behavior appears to be situational, varying by grade level, subject area, type of student, and instructional goal, and even according to the teacher's personality. The former view implies that valid assessments of teaching process are possible, the latter that they probably are not.

An important distinction, in this regard, is between evaluations of minimum teacher competency and evaluations aimed at differentiating superior or outstanding from average performance. According to Wise et al. (1984), the argument that specific performance-related behaviors can be observed in the classroom is more plausible with respect to the former than the latter because relatively gross phenomena (e.g., inability to maintain control or to present subject matter coherently) distinguish incompetent from competent teaching. In comparison, the behaviors that distinguish superior or excellent teaching from ordinary teaching are subtler, more particular to the subject and the setting, and less well understood. These more-difficult-to-make distinctions are, of course, the relevant ones in the context of performance-contingent rewards.

Fortunately, it is not necessary to resolve the underlying scholarly disputes about what we do and do not know about teaching to decide whether classroom observation, as actually practiced, is likely to yield valid performance ratings. The aspects of teaching behavior that evaluators are asked to rate in practice bear only the faintest resemblance to the teaching "competencies" discussed in effectiveness research. While the latter are specific and operational, the former are broad, vague, and

subjective. In a true assessment of "competencies," the observer might be asked, for example, to record how often a teacher uses particular types of questions, cues, and directions, and the frequency of using effective types would become part of the performance rating. In contrast, under the rating methods now being proposed for incorporation in certain incentive systems, observers would be asked to judge teachers' "preparation for instruction," "use of appropriate teaching techniques," and "classroom management."[7] Such broad-brush, high-inference items invite, in fact compel, impressionistic rater responses. There is no evidence whatever that ratings on such gross criteria correlate with teaching effectiveness or that such questionnaires can yield valid performance ratings.

I mention more briefly three other factors likely to detract from the validity of observation-based performance ratings:

First, apart from the more general lack of predictive validity cited above, the validity of teacher rating procedures is further degraded by the inclusion of rating criteria with only peripheral relevance to teaching effectiveness. For example, along with the relevant, albeit hazily defined, criteria cited above, the proposed Tennessee/Delaware rating system calls for assessing such things as the teacher's preparation of lesson plans, pursuit of graduate courses and advanced degrees, and leadership and community relations activities. There is not even logical,

---

[7]The items cited are derived from the Tennessee Career Ladder Evaluation System, as described in Cresap, McCormick and Paget (1985) in setting forth a proposed career ladder plan for Delaware. More detailed explanations are provided, in conjunction with the rating forms, of the categories of behavior subsumed under each broad heading, but the raters are asked only to respond holistically to the broad items themselves, and not to the more detailed, behavioral elements underlying them.

much less empirical, reason to believe that these factors predict student learning. Hence, even if the rating method were otherwise valid, that validity would be undercut by assigning weight to essentially irrelevant criteria. In this case, less is more. Including items with demonstrated ability to predict effectiveness is helpful; anything else detracts.

Second, the validity of performance evaluations is threatened by the inadequately small, and hence unrepresentative, samples normally allowed for under observation-based rating schemes. The two, three, or four annual visits typical of such schemes do not suffice even to sample teaching in all the major subject areas (in the case of elementary teachers) or to observe teaching in different situations and with different classes (in the case of secondary teachers), much less to allow for sampling variation.

Third, and most important, validity is gravely impaired by the obtrusiveness of classroom observation. The presence of any observer, but especially an evaluator, changes classroom activity drastically, eliciting unnatural behavior from both teachers and students. It does not matter whether observation is expected or unexpected. The former leads to intentional, even rehearsed, artifical behavior, but the latter also disrupts the classroom environment, so there is reason to doubt that anything typical can be observed. Considering also the interaction between obtrusiveness and small sample size (i.e., the rarity of visits makes them even more special), it seems unlikely that real classroom behavior would ever be assessed.

Taking all the foregoing factors (and others) into account, Scriven (1981) concludes, in part, as follows:

> Using classroom visits by colleagues (or administrators or "experts") to evaluate teaching is not just incorrect, it is a disgrace....[N]othing that could be observed in the classroom (apart from the most bizarre special cases) can be used as a basis for ... any conclusion about the method of the teaching....There are no valid indicators to be seen, no matter who looks.

While I would not go quite so far, I do conclude that the validity problems in observation-based, process-oriented assessment are at least as severe as those encountered in evaluations based on student outcomes.

Reliability. Doubts about the reliability of ratings of classroom teaching reinforce the doubts about their validity. The vaguely defined performance criteria and the inherently subjective, highly personalized nature of the judgments to be made work to ensure that inter-rater reliability will not be high. These problems are aggravated by the small number of observations allowed for under typical evaluation schemes. Given the great variability of classroom activity from day to day and hour to hour, even the same raters' assessments of performance on different occasions are likely to conflict.

Unbiasedness. Concerns about bias in observational rating methods are at the heart of teachers' opposition to their use in allocating performance-based rewards. The likelihood of bias (deliberate or not) is especially great when the evaluator is the building principal or another teacher from the same school. There are unavoidable conflicts of interest between the principal's role as supervisor, or the peer's role as colleague, and the role of objective evaluator. Both principals and peers may have interests, unrelated to teaching performance, in whether the evaluee succeeds or fails, advances or falls behind, or remains in or leaves

the school. Yet reliance on principal evaluation is pervasive and, as noted above, would be institutionalized in some of the proposed and pending state incentive schemes.

In addition to personal biases based on relationships within the school, evaluators may also have more generic biases related to teacher characteristics (age, race, sex, etc.) or to particular teaching styles. The latter can be particularly insidious because the evaluator typically believes that his or her personal preferences among teaching approaches reflect valid distinctions among more and less effective modes of teaching (Scriven, 1981). In some respects, these generic biases are more troubling than the personal ones, because while the latter can be avoided by selecting evaluators appropriately, the former are much more difficult to weed out. Nothing less than a corps of trained, professional, outside evaluators is likely to work, and even that may not suffice.

Discriminating Power. The power of an observational rating system to discriminate among degrees of superior performance is unclear. Ultimately, it derives from the reliability of the rating method, so if the reliability is low, the ability to discriminate is low too. But in addition, discriminiating power in some evaluation system is limited by design. as, e.g., where the system allows only "satisfactory" or "unsatisfactory" ratings to be assigned to teachers. Whether this can be rectified depends on the performance criteria: can they be defined in such a way that distinctions among multiple levels of performance can be made operational? This hinges on how specialized the criteria must be to fit different grade levels, subjects, and teaching assignments (an issue discussed under "universality," below).

Universality. Some observation-based evaluation methods implicitly claim universal applicability, but such claims need to be viewed skeptically. The aforementioned Tennessee/Delaware evaluation plan, for example, purports to rank teachers on generic criteria that apply equally to all grade levels, subject areas, teaching assignements, and types of children. For example, it asks, regardless of setting, "does the teacher use appropriate teaching techniques?" and "how well does the teacher manage the classroom?" But to ask the question does not mean that it is answerable meaningfully, or in terms that mean the same thing regardless of the teaching situation. According to Wise et al. (1984), asking about generic teaching skills may suffice when only minimum competency is to be assessed, but more specialized criteria apply and more specialized evaluators are needed once the focus shifts to superior performance. Whether "appropriate teaching techniques" are used may be as reasonable a question to ask about high school physics instruction as primary reading instruction, but the meaning of "appropriate" varies, as does the background needed by an evaluator to make an informed judgment. Consequently, although the form of evaluation findings may be universal, the content is likely to be particularized, and the ability to make valid comparisons may well be limited to relatively homogeneous subgroups of teachers.

Predictability by Teachers. It is impossible to generalize about this attribute of observation-based methods because predictability depends entirely on how the performance criteria are defined. The more detailed and specific the criteria, the better will be the teachers understanding of the behaviors needed to earn a given performance score. But the rating scales more likely to be used, judging by incentive proposals to date,

contain high-inference, vague, and subjective items, which give teachers much less basis to predict observers' impressions and ratings. Stylistic differences between evaluator and evaluee, and among the evaluators themselves, can create great uncertainty about how particular classroom approaches will be received.

Side Effects. Unlike an outcome-based approach, a process-oriented evaluation method is unlikely to distort the content of the curriculum (there is no test to teach to), but it is more likely to distort teaching styles and methods. Intentionally or not, any performance rating scale that is not completely vague conveys messages about preferred teaching styles. Thus, there is the danger that linking rewards to observational rating scales would tend to skew teaching styles toward officially preferred approaches and discourage stylistic innovation. If it is correct that appropriate styles vary by subject, grade level, etc. (and even in relation to teacher personality), such rigidification would be an undesirable effect.

Costs and Burdens. The major costs of process-based evaluation are the costs of evaluators' time. The magnitude of these costs depends on (a) how extensively each teacher is observed to produce a performance rating, and (b) how often teachers are rated. There is a direct trade-off between the cost of evaluation and such attributes as reliability and discriminating power, which also depend on the number of observations per teacher. There is little doubt that reliable process-based evaluations, which would require multiple evaluators and multiple visits to a classroom by each evaluator, would cost substantially more than assessments of educational outcomes. Unlike the latter, however, with their requirements

for extensive testing, the former would not entail significant diversions of class time away from instructional activity (unless one counts as diversions the time spent in the artifical classroom visit situations).

COMPARATIVE ASSESSMENT

The overriding concern about either method of evaluation is that the performance ratings may be invalid—that is, not good predictors of teachers' contributions to learning. In the case of an evaluation based on classroom observation, the danger is that teachers will be judged for behavior unrelated or only weakly related to results; in the case of an evaluation based on student outcomes, it is that outcomes will not be measured well and/or that factors outside the teachers' control will not be taken adequately into account. The consequences of error are quite different in the two cases, however. Faulty measurement or faulty adjustment of outcomes under the product-based approach would result in unfair and inconsistent treatment of some teachers, but unless the procedures are egregiously bad, there will still be a positive association between performance ratings and effective teaching. Thus, effective teaching will, on average, be rewarded. On the other hand, a failure to identify outcome-related classroom behaviors under the process approach could lead to rewarding, and hence encouraging, mediocre or even counter-productive modes of teaching. In both cases, the wrong teachers might receive high ratings and rewards, but under the process approach there is more of a risk to educational quality.

The criteria of reliability, unbiasedness, and discriminating power generally favor the outcome-based approach. The problems of measuring

teachers' contributions to outcomes reliably are considerable, but those of obtaining reliable ratings from classroom observation are more severe. Impersonal ratings based on student test scores avoid the dangers of bias and subjectivity inherent in any system that relies on evaluators' judgments. In the case of some of the classroom observation systems now being proposed, which rely on evaluations by school principals or other interested parties, those dangers are so severe as to render the evaluations useless for purposes of apportioning rewards.

It is not clear how the side effects of the two evaluation approaches balance out because what is a negative side effect for some people is sometimes a positive effect for others. Evaluation systems based on classroom observation are likely to bring about a narrowing of teaching methods around those favored, explicitly or implicitly, by the evaluation protocols. Those who feel that more standardization and control of teaching is needed (and believe that we know what "good" methods are) might find this desirable; those who emphasize the need for diversity in teaching to cope with varied situations will find it a considerable loss. Evaluations based on student outcomes are likely to influence the content of the curriculum and promote "teaching to the test," which may or may not be beneficial, depending on how it is done. In addition, such evaluations are likely, because of the requirement for large-scale, uniform measurement of outcomes, to promote statewide standardization and central control of curricula. Until recently, the latter might have been enough, by itself, to rule out the outcome-based approach. Now however, with states engaged in specifying and tightening curricula, mandating additional statewide testing, and, of course, installing statewide incentive plans,

that same centralizing effect might be viewed as complementary to the other reforms (Goodwin and Muraskin, 1985).

Mainly on the basis of the validity and bias arguments, I consider outcome-based evaluation to be, in principle, the favored approach, and in my view it is the approach we should work to perfect for the longer run. But in the shorter run, there is an important practical point to consider. Tests suitable for evaluating the outcomes of instruction by particular teachers in particular classes are now available for only a very limited range of subjects and grade levels--primarily basic skills subjects in the elementary grades. Moreover, the procedures for adjusting test scores and establishing performance norms have not been worked out in detail. Consequently, the opportunities for near-term implementation of outcome-based evaluation are limited, and substantial development work will be required to enlarge them in the future. In comparison, process-based evaluation, mainly by means of classroom observation, is commonplace in the schools and is applied, for better or worse, to teachers of all grades and all subjects. While much of this types of evaluation is highly flawed and/or inappropriate for rating performance, it is likely that with some effort it could be reoriented to the task and its worst flaws alleviated. Realistically, if there is to be any substantial implementation of merit pay and career ladder plans in the next few years, much of the evaluation will have to be done by traditional classroom observation methods. Later, it may be feasible to rely on outcomes more heavily. For the time being, process-oriented methods or combinations of process and outcome approaches are likely to dominate the field.

IV. LINKING REWARDS TO PERFORMANCE

Compared with all the attention given to teacher evaluation, the other half of the incentive design problem, relating rewards to performance, has been largely neglected. Although every incentive proposal necessarily relates rewards to performance ratings in some manner, there has been little analysis of how rewards should be structured or apportioned, and there are no established guiding principles or rationales.[1] Proposed designs vary widely. Among the plans recently adopted or now being considered by states, some offer large rewards, others only tokens; some reward many teachers, some few; some would make rewards permanent, others temporary; some involve promotion and new responsibilities, some only increments in pay; some recognize multiple degrees of merit, others only one; some cover all teachers, others only volunteers; and so forth through many other design features. It seems mainly a matter of accident that a particular state favors a particular incentive approach. In this chapter, I consider whether there is a more systematic way to choose an appropriate design and to match the features of an incentive plan to the purposes and circumstances of a locality or state.

For the purpose of this discussion, I temporarily set aside the problems of performance measurement discussed in Chapter III and proceed as if it were feasible to generate valid, reliable, fair, and otherwise acceptable ratings of each teacher's performance. This makes it possible to focus on an issue logically separable from evaluation: how, given

---

[1]One of the few papers that does deal systematically with a series of specific design issues is Hatry and Greiner (1984).

performance ratings for every teacher, should performance be rewarded? At some points the fiction that there are no obstacles to performance evaluation cannot be maintained because issues of system design interact with issues of measurement, but by and large, separating the other design issues from issues of teacher evaluation facilitates the assessment of alternative incentive systems.

Much of what I say in this chapter pertains equally to merit pay and career ladder plans, which is natural since career ladder plans virtually always subsume performance-based pay. Career ladder proposals do raise certain special issues, however, that do not arise when the rewards are limited to salary increments. I note below the topics that require special treatment when promotion is one of the rewards, and I also deal separately with certain issues that arise only in connection with career ladder plans, such as the appropriate roles and responsibilities of teachers promoted to higher ranks.

For expository purposes, I have organized the following discussion around a series of major design issues, as follows:

- o The form of rewards
- o The duration of rewards
- o Incorporating performance-based rewards into existing reward structures
- o The number and size of rewards
- o The hierarchical structure of rewards
- o Performance thresholds and rationing
- o Eligibility and participation
- o Evaluation units and comparison groups

These divisions are sometimes artificial because the categories overlap, but they prove useful for imposing some order on a complex web of design questions. Each cluster of issues is considered in a separate section of this chapter. In addition, the last two sections deal, respectively, with interactions among design features and with the implications of this analysis for some of the recently developed state incentive plans.

THE FORM OF REWARDS

The main forms of performance-contingent rewards now being developed in the states are (a) performance-based pay alone and (b) performance-based pay coupled with promotion along career ladders. Also featured in some proposals, either alone or in conjunction with one of the above, are such other reward forms as nonmonetary recognition, special assignments, improved working conditions, and extra perquisites (Cresap, McCormick and Paget, 1984).[2] I have already mentioned, at the end of Chapter II, some of the considerations involved in choosing among these forms, but I now discuss certain points in greater detail.

The formal distinction between merit pay and career ladder plans is that the latter involve multiple ranks and promotions in addition to performance-contingent pay; but as explained earlier, a more meaningful distinction is between plans that do and do not assign significantly different roles to teachers of different ranks. Promotion per se (without

---

[2]In addition, there are certain specially targeted forms of rewards that I do not discuss in this paper, including pay differentials related to such particular skills as expertise in mathematics or science and various forms of subsidies (scholarships, loans, work opportunities) for persons in teacher training programs.

role differentiation) is a form of honor or recognition and, as such, only a minor departure from merit pay, while promotion with role differentiation is a qualitatively different approach. If there are important role differences, the career ladder plan must be assessed not only as a system of rewards but also as a method of reorganizing the instructional staff and the delivery of services within schools. At this point, I am concerned mainly with the incentive aspects of reward plans, but I comment, where appropriate, on the differentiated staffing aspects as well.

Considering the incentive effects only, are there good reasons to set up a career ladder plan rather than just a system of performance-based pay? Or, putting the same question differently, are the nonsalary aspects of career ladder plans likely to contribute significantly to teacher incentives in their own right? To focus the issue, consider two plans that relate pay to performance in exactly the same way. In plan M, a merit pay plan, teachers who attain each of three successively higher performance thresholds earn progressively higher supplements to their regular scheduled salaries. In plan C, a career ladder plan, the specified performance levels and pay differentials are the same, but each performance threshold is associated with a rank, title, and status, and some ranks, at least, carry special nonteaching responsibilities. Those who attain the first performance threshold are promoted from "apprentice teacher" to "teacher" and awarded tenure; those who reach the next level are designated "senior teachers" and assigned to help other teachers improve their skills; and those who attain the highest level ("master teacher") are made responsible for supervising and evaluating other teachers and developing the instructional program. How do the incentives provided

by plan C compare with those provided by plan M? I see four principal differences, as follows.

First, plan C offers professional recognition not offered by plan M. Promotion to each successively higher rank is considered an honor and presumably advertised as such. There is no way to estimate what it is worth to a teacher to receive such an honor (abstracting from the monetary reward that goes with it), but it is reasonable to believe that the value would be positive to some if not all teachers, and hence that adding recognition to performance-based pay would increase the incentive effect. (If people did not value such recognition, it would be hard to explain all the medals, certificates of appreciation, and testimonial dinners—not to mention teacher-of-the-year awards—that figure in our public life.)

Second, the special nonteaching responsibilities associated with the higher ranks are likely to affect teachers' incentives to attain those ranks; but—an important point—it cannot be assumed that the effect would be positive for all teachers. Some teachers would welcome the teacher training and evaluation assignments, not only for the status associated with such roles but also as intellectual challenges, refreshing changes from classroom teaching, and potential stepping stones to administrative positions in the future. Other teachers might find the same assignments burdensome and distasteful, however, or simply might view them as unwelcome diversions from what they like to do and feel comfortable doing—teaching children. Thus, although the differentiated staffing elements of a career ladder plan are likely to strengthen the incentive for some teachers to perform well, they are also likely to weaken the incentive for others. In the case of a voluntary career ladder plan

(see the discussion of voluntariness later in this chapter), teachers in the latter group might be deterred from participating if promotion along the career ladder carries the obligation to assume significant nonteaching responsibilities.[3]

Third, the promotion and differentiated staffing elements of a career ladder plan may add to the acceptability, and hence the effectiveness, of performance incentives. That is, people both inside and outside the school system who find unadorned pay-for-performance plans unappealing might consider more palatable monetary rewards linked to career progression, professionalism, broadened responsibility, and other virtues evoked by the career ladder concept. The importance of this employee relations, or public relations, criterion is unknown, but the predominance of career ladder plans over pure merit pay plans in recent state proposals suggests that there may be something to the idea.

Fourth, based on the state plans proposed to date, it appears that significant differences in the timing and duration of rewards may be associated with the choice between career ladders and merit pay. Rewards under career ladder plans are likely to be permanent, but long waits may be required to become eligible for each successive promotion. The rewards under merit pay plans may be either permanent or temporary but, in either case, are likely to be accessible with less delay. These timing differences may affect the strength of the incentives considerably.

---

[3]It should be possible to avoid this problem by setting up alternative types of assignments for high-performing teachers not attracted to supervisory, evaluative, or other noninstructional roles or simply by making nonteaching activity a voluntary activity for teachers who attain higher ranks.

I defer comment on the effects to the following section, which deals specifically with the timing and duration of rewards.

Looking beyond the direct incentive effects, there is another major difference between merit pay and career ladder plans to consider: the manner in which high-performing teachers are used. Under pure merit pay plans (or career ladders with largely honorific ranks), such teachers remain in the classroom. Consequently, any gains in performance stimulated by merit pay are realized immediately in the form of improved teaching. In contrast, under "true" career ladders (those with significantly differentiated responsibilties), the best performers—"master" or "mentor" teachers—spend significant time in nonteaching roles. Thus, other things being equal, there is likely to be less of a short-term gain in classroom performance under a career ladder than under pure merit pay. On the other hand, the mentor/master role constitutes investment in the future—time spent in evaluating other teachers and helping them to improve. If the plan succeeds, long-term performance may be enhanced. What is at issue, therefore, is a potential trade-off in time: maximum present benefits by leaving the best teachers in their classrooms versus future benefits from diverting them to disseminate their skills.[4]

---

[4]At least two other factors need to be considered in assessing this trade-off. One is the efficacy of high-performing teachers as teacher evaluators and trainers, as compared with that of specialists in those roles. Another is the trade-off between proficiency in those roles and the criteria for selecting teachers for merit awards. The possibilities regarding the latter are, on one hand, to promote teachers solely on the basis of their performance with children and to hope that they will also be effective performers with adults or, on the other, to promote teachers partly on the basis of their prospective performance as teacher trainers and evaluators, with the risk that the incentive for good classroom performance will be undercut.

THE DURATION OF REWARDS

How long rewards last (i.e., how long a teacher continues to receive them once they are earned) is likely to be an important determinant of the effectiveness, cost-effectiveness, and equity of an incentive system. Among the possibilities discussed in the literature (Hatry and Greiner, 1984; Cresap, McCormick and Paget, 1984) and represented in proposed incentive plans are

1. Permanent merit pay increments and/or promotions, in which a performance-based increment in pay, once earned, becomes part of the teacher's regular salary, and a performance-based promotion, once conferred, establishes the teacher's new permanent rank;
2. Term pay increases or promotions, in which good performance earns a pay increment or promotion valid for a specified term of years, with subsequent renewal contingent on a new demonstration of superior performance;
3. Nonrecurring rewards, usually in the form of one-time performance bonuses, in recognition of performance during a specified interval (usually one year), and for which teachers compete anew each period.

A strong case can be made that both the strength of the incentive to teach well and the fairness of the incentive system can be maximized by emphasizing short-term rewards--that is, performance bonuses rather than permanent raises and promotions. Limiting the rewards to one-time

payments for performance during a particular period maximizes the performance-contingent fraction of each year's salary pool and hence the number and/or size of rewards that can be offered for performance during each period. Under the reasonable assumption that the potency of a reward is an increasing function of its expected value, this also maximizes the strength of the incentive to perform well.[5] Moreover, limiting the duration of rewards minimizes the potential commitment of merit pay funds to teachers who performed well at some point in the past but who no longer exhibit superior performance. Holding down such payments is desirable for both efficiency and equity reasons. In terms of efficiency, such payments are wasted because they buy no improvement in performance. From the standpoint of equity (and morale), eliminating such payments avoids the appearance of unfairness attendant on paying more to teachers who may have performed well in the past but whose current performance is below that of teachers not receiving merit pay.

Consider, for example, the difference between a plan under which a teacher who meets a specified performance standard for each of, say, three consecutive years receives a permanent $2,000 per year pay increase and an alternative plan that pays a $2,000 bonus to each teacher who satisfies the same standard in any particular year. Under the former plan, the $2,000 per year pay increment, once awarded, is committed whether performance is maintained or not; under the latter, it is contingent on continued high performance every year. In terms of incentive effects,

---

[5]The expected value of a reward is the mathematical product of the size of the reward and the probability of receiving it. Thus, expected value increases with both the size and the number of rewards.

two things seem clear. First, the permanent pay increase provides a stronger incentive for superior performance during the initial qualifying period. Specifically, the permanent-reward plan offers a reward with a present value of about $14,000 for three years of superior performance, while under the short-term plan, a teacher would have to do well each year to earn a $2,000 bonus.[6] But second, the incentive provided by the permanent pay increase falls to zero after year three, whereas the bonus plan offers a continuing year-by-year incentive to perform. Thus, the choice is between a strong incentive to perform well at the outset and a smaller incentive to perform well for an entire career.

There are some legitimate arguments to be made, however, in favor of permanent, or at least multiyear, pay increases and promotions for those who exhibit superior teaching performance. In the case of a career ladder system with significant differentiated staffing, continuity is likely to be an important consideration. For instance, those who become master teachers, responsible for training and evaluating other teachers, need to accumulate experience in those roles to carry out their responsibilities satisfactorily. Elevating individuals to the master teacher level for only one year at a time, with retention of the title always in doubt and contingent on new performance evaluations, would probably not result in an effective or commited master teacher corps. Conferring permanent promotions may be undesirable for the reasons given above, but reselecting master teachers every year seems equally undesirable

---

[6]The $14,000 figure is the present value, calculated at a 10 percent discount rate, of a stream of payments of $2,000 per year that begins three years from now and extends into the future for an assumed career length of 30 years.

for other reasons. Something in between--say, promotion for a term of several years, with renewal for subsequent terms contingent on sustained performance--may be the most sensible solution.

PERFORMANCE-BASED REWARDS AND EXISTING REWARD STRUCTURES

A consideration that influences and constrains the form of performance-contingent rewards is the need to relate them to existing reward structures. Two aspects of this relationship are (1) the connection between performance-based pay and existing salary schedules, and (2) the connection between performance-based promotion and the existing tenure system.

Merit Pay and Existing Salary Schedules

Under the remarkably uniform philosophy of teachers' compensation prevailing throughout the United States, a teacher's salary is determined mainly by two factors: education, as measured by degrees held and/or post-B.A. credit hours completed, and years of teaching experience or seniority. Reasonably typical 1984-85 salary schedules offer starting salaries of $14,000 to $16,000 to teachers with B.A. degrees and no experience; increments of $1,500 to $3,000 for M.A. degrees, plus additional increments for more credit hours; and increments of $400 to $600 per year of seniority up to a maximum of 10 to 15 years. In addition, extra pay is often given to teachers who perform certain special functions, such as serving as department heads and coaching athletic teams. The ratio of the highest salary paid by a district to the lowest is likely to be on the order of two to one. There is, of course, no provision for salary to vary as a function of performance.

Performance-contingent pay can be incorporated into the salary structure in any of the following five ways:

1. Performance increments can be superimposed upon the existing structure, leaving the original salary schedule unchanged. That is, each teacher continues to receive his or her regularly scheduled salary, but teachers who attain specified performance thresholds receive, in addition, specified merit increments in pay.
2. Performance increments can take the form of specified percentage increases in salary for teachers who satisfy stipulated performance criteria.[7]
3. Separate salary schedules of conventional form can be applied to teachers who reach specified performance thresholds or steps on a career ladder.
4. The annual scheduled increase in pay can be adjusted up or down (perhaps even to zero) on the basis of the teacher's performance.
5. Finally, a completely new form of salary schedule can be developed, in which salary is specified as a function of performance and other factors, possibly but not necessarily including the traditional experience and/or education factors.

---

[7]One can also conceive of more complicated relationships than simple additive performance increments or simple percentage increases. For example, a hybrid of the two might provide a pay increase of, say, $1,000 plus 10 percent of regular salary to a teacher who attains a specified level of performance.

To assess these options thoroughly would require an analysis of the rationale for and implications of the existing education and experience-based teacher salary schedules, but as that would take me too far afield from the topic of the paper, I confine myself instead to a few observations regarding the rationales for and uses of the different methods listed above.

The first two options are the simplest, and one of them, the additive performance increment, is the approach found most frequently in recently developed state plans. These options reflect the idea that each teacher has a "regular" scheduled salary, not itself linked to performance, but to which performance increments may from time to time be appended. They are the natural methods to choose, for two reasons, for a system of short-term performance bonuses: first, they make awards of performance-based pay readily reversible; second, they require no modification of the underlying regular salary structure. In addition, they are suitable for use in longer-term merit pay or career ladder plans where a "mild" form of performance-contingent pay is desired--that is, where superior performance is rewarded, but teachers' regular salaries and salary increases are not at risk.[8]

The third alternative, separate salary schedules for teachers at each performance level or each rung of the career ladder, differs from the preceding options in two respects. First, it allows for more complex forms of performance-based salary differentiation. Instead of merely

---

[8]Later in this chapter, I argue that it is unlikely, in the long run, that there can be winners only and no losers under a performance-based reward system. That is, even though a system nominally gives "regular" pay raises to teachers who fail to qualify for merit pay, those raises are likely, over time, to be smaller than those that would have been forthcoming if none of the salary budget had been diverted to merit pay.

adding a fixed amount or fixed percentage to regular salary, one could vary the "shape" of the salary schedule from one category of teacher to another. For example, the salary schedule for a master teacher might offer larger seniority increases than the other schedules to encourage master teachers to remain in the system and might continue to pay experience increments for up to 20 or 25 years of experience, as opposed to only 10 or 15 for lower-ranked teachers. Second, shifting a teacher from one salary schedule to another has more of a connotation of permanence than merely adding a performance increment to regular salary. This makes the option useful primarily in connection with career ladder systems and nonreversible awards of merit pay.

The fourth alternative, modifying teachers' scheduled pay increases on the basis of performance, is a more drastic approach. To adopt it is to abandon the idea that teachers are entitled to regular salary increases whether or not they teach well. Instead, this alternative would link the size of the annual increase to how well a teacher performs. There are a number of ways in which performance could be taken into account. For example, taking the regular scheduled rate of increase as the base, one could define a minimum performance threshold below which no raise, or only a reduced raise, would be paid and one or more thresholds of superior performance above which multipliers would be applied to the regular percentage rate. Under a strong version of this formula, a sub-standard teacher might receive no raise at all (not even an inflation adjustment) for several consecutive years, whereas a superior teacher might enjoy raises of 150 or 200 percent the average percentage. Such a reward structure would magnify the incentive for teachers to do well

and create strong incentives for good teachers to remain and poor teachers to leave the profession.

Finally, the fifth option, a performance-based salary schedule, is potentially the most radical method because it makes a teacher's entire salary, not only the annual increase, dependent on performance. By adopting it, a state or district would be declaring that performance is not a secondary or supplementary consideration in the determination of pay but at least coequal with experience and training. What would such a salary schedule look like? That depends, in part, on whether the existing salary schedule factors were retained. One possibility is that the performance factor might be substituted for the traditional education factor in teacher salary schedules. Another is that both experience and education would be retained, but the amounts paid for each year of seniority, or for course credits or advanced degrees, would all depend on levels of performance. Thus, teachers with the same education and experience could receive significantly different annual salaries.

Once one recognizes performance as a legitimate salary factor, however, it becomes hard to escape the question of whether experience and, especially, teacher education should continue to be rewarded as at present. The cumulative findings of research on teacher effectiveness leave little doubt that the education factor is unrelated to teaching effectiveness, and they also suggest, only slightly less strongly, that the benefits of experience are fully realized after the first few years. A salary schedule that rewards both teaching performance and post-college credits and degrees can almost be said to contradict itself. Moreover, such a schedule provides a mixed message to teachers: "improve your performance

and raise your pay, or take graduate courses, which add nothing to your performance, and raise your pay by a comparable amount." Substitution of performance for education in salary schedules would go far toward rationalizing the teacher reward structure.

Performance-Based Promotion and Tenure

A rudimentary form of performance-based promotion already exists in most school systems: the elevation of new teachers to permanent status and tenure after an initial probationary period. How should this transition be accommodated, if at all, into a performance-based promotion, or career ladder, system?

One possible answer has already been provided by some of the newly developed state career ladder plans: make tenure one of the performance-based promotions and rewards. The first step up the ladder on a proposed plan recently designed for the state of Delaware (Cresap, McCormick and Paget, 1984b) and the second step up the career ladder now being implemented in Tennessee is that from Apprentice Teacher, a nontenured position from which teachers are subject to dismissal, to an initial tenured rank labeled Career Level I.[9] These arrangements modify the traditional tenure-granting arrangement in three noteworthy respects. First, they appear, at least in theory, to impose more substantive performance requirements for promotion to tenured rank than one finds in conventional systems. Second, there is a longer period during which a district can decide whether to retain a teacher than the standard single probationary year. Third, a performance-

[9]The Tennessee plan allows for both a one-year probationary teacher period and a subsequent three-year apprentice teacher period, neither of which is subject to renewal (Southern Regional Education Board, 1984).

based increment in pay, larger than the regular one-year seniority increase, is associated with the promotion to permanent status. All three contribute to the incentive to perform well at the beginning of the teaching career.

But what about the period after tenure is granted? Does not tenure itself conflict with and undercut the principle of performance-based promotions and ranks? In my view, whether or not such a conflict exists in principle, its importance in practice depends on how the merit-pay part of the system is structured. If performance-based pay is treated as an appendage to regular pay, so that a poor teacher continues to receive regular pay increases without regard to performance (as in salary structure options 1, 2, and 3, described above), then tenure could be a serious problem. Under those options, a tenured teacher whose work becomes unacceptable faces minimal financial penalties and no compelling reason to leave. On the other hand, under the more thoroughgoing forms of performance-based pay described above, such a teacher could receive zero increases each year (option 4) or even actual reductions in pay (option 5). Thus, a strong incentive to leave teaching would be created, and the likelihood of having to institute dismissal proceedings would be reduced. A strong system of performance-based pay reduces the need for performance-based dismissal and avoids the bitterness such an attack on tenure would entail.

THE NUMBER AND SIZE OF REWARDS

Two issues that must be addressed by anyone designing an incentive system are "how many teachers should be rewarded?" and "how large should the rewards be?" These questions, separable in principle, are linked in practice by resource constraints. Assuming, not unrealistically, that

the budget for teacher incentives is fixed, the trade-off is direct: doubling the rewards means halving the rewardees. Thus, one must consider not only the size and number of rewards but also the optimal combination of the two. In addition, these questions are greatly complicated by the fact that real-world reward systems are likely to provide rewards in multiple sizes to fit different levels of superior performance. To keep the discussion manageable, I deal first with the issues of size and number when there is only one type of reward. Then, in the following section, I consider the more complex hierarchies of rewards.

Size of Reward and Incentive Effect

One key consideration is how the size of a reward affects the strength of the incentive to perform. Several writers have pointed out that there must be more than token rewards to influence behavior (e.g., Hatry and Greiner, 1984). The amounts at stake must be substantial enough to "make a difference" to the teacher's level of economic well-being and more than compensate the teacher for the costs of generating a satisfactory response. Such costs can be substantial. Teachers, after all, would be asked to make significant changes in behavior, possibly including putting in additional hours, working more intensely, abandoning customary and "easy" teaching methods for harder ones, and investing time and energy in self-improvement activities. It is unreasonable to assume that such responses can be purchased for $100 prizes.

Unfortunately, it is easier to enunciate the principle that rewards must be "substantial" than to name specific dollar amounts. There is no formula for deciding how much is enough. Pending the analysis of

response data from large-scale statewide incentive plans, one can only attempt to judge subjectively what may be required to elicit the desired behavior. The prevailing guess seems to be that rewards must be at least on the order of 10 to 20 percent of prevailing salary levels to motivate teachers significantly, which is to say, in the range of $2,000 to $5,000 per year. Salary increments in this range would be provided under many of the plans now being implemented or considered in the states.

From the point of view of the employer (school district or state), there is a different question about the size of rewards to resolve—not how much it costs to obtain higher performance but how much higher performance is worth. This is a difficult question even to formulate precisely, since we are not used to thinking about hiring different quality grades of teachers at different prices. Nevertheless, in one way or another, explicitly or not, the problem must be confronted. If the price of quality proves to be too high—i.e., if larger rewards are required to stimulate performance or attract high quality teachers than school systems (or the public) feel they can afford—the enthusiasm for incentives is likely to be short-lived.

One potentially useful way to think about the issue of worth is to compare subjectively the value of higher-quality teaching against the values of teacher attributes for which specific sums are now being expended. Typical salary schedules, as noted earlier, offer pay increments of $1,500 to $3,000 for a master's degree and $2,000 to $3,000 for five years of experience. Given the demonstrated willingness of districts to pay those amounts for credentials that appear to have no relationship (in the case of advanced degrees) or little relationship (in the case

of experience) to the quality of teaching, it seems plausible that they would be willing to pay similar amounts—premia in at least the $2,000 to $3,000 range—to obtain noticeably higher quality teaching.

The question of what size rewards will bring forth what size responses is ultimately answerable only from empirical experience. This makes it a matter of some urgency that states implementing merit pay plans or career ladder programs offer substantial enough rewards to elicit responses or, better, rewards of varied sizes to permit estimation of differential response rates. There is some danger that financial constraints and/or pressures to spread rewards thinly may lead to the conclusion that performance-based rewards "don't work" when the real problem is that the prices offered were too low.

The Number of Rewards

The effectiveness of an incentive plan also depends on how many rewards it can offer or, to be more precise, on the percentage of teachers to be rewarded. Assuming, for the moment, that there is only one type of reward and taking its size as given, raising the number of rewards means raising the probability of becoming a rewardee and hence the expected value of a reward to the average teacher. If rewards were restricted to a small stratum of outstanding teachers, say only the top 5-10 percent, large numbers of teachers would conclude, correctly, that they had little chance to qualify and hence would feel no incentive to perform. Raising that percentage, say to the 20-30 percent range, would convince many more teachers they had a chance and hence stimulate them to compete. Thus, up to a point, offering more rewards increases the incentive effect.

Eventually, however, further increases in the number of rewards have two negativ consequences, which first attenuate and then cancel out the incentive effects. First, raising the number of rewardees requires lowering the quality threshold at which a teacher qualifies for performance-based pay. Hence, as the percentage of rewardees rises, the rewards buy progressively smaller performance gains. Second, as the performance threshold falls, increasing numbers of the better teachers will earn rewards without doing anything to improve their teaching. In the extreme case, rewarding a large majority of teachers, say 70 or 80 percent, would allow most above-average teachers to qualify without teaching any better than they would have taught with no incentives at all. For them, the reward system would cease to offer any inducement to improve. From the point of view of the school district or the state, there would be little performance improvement to show for a large investment in performance-based pay.

In the case of a career ladder plan of the master/mentor teacher type, there is also a nonincentive factor to take into account: the number of high-ranked teachers needed to perform nonteaching assignments. Only so many people are needed to play the teacher trainer and teacher evaluator roles. For example, the California mentor teacher program limits the number of mentors to five percent of each district's teaching force. The significance of this depends on how the rest of the promotion system is designed. If other rewards are open to outstanding teachers without constraint, the significance of limiting the numbers of masters or mentors may be minimal; if not, rationing may undermine the incentive effects (see the comments on "rationing of rewards," below).

The Trade-off Between Number and Size

It follows from the above that the right answer to "how many teachers should be rewarded?" is not "the more, the better." There is a point after which increasing the number of rewards becomes counterproductive, even if the expense of providing rewards is not an issue. In real life, of course, cost is also a major concern. Thus, there are two reasons, one related to the effectiveness of incentives and one to constraints on cost, to seek the best combination of the number and the size of rewards.

Without quantitative data on how teachers respond, there is no way to be precise about where that optimal balance lies, but several relevant points can be noted. First, requiring that rewards be "substantial" sets a lower bound on size and hence an upper bound on the number of rewardees. Second, limiting the number of high-ranked teachers under a career ladder plan for nonincentive reasons may further constrain the trade-off between number and size. Third, a key consideration is the relative value assigned by the school system to performance improvements in different parts of the performance range. A school system interested in bringing up the lower end of the quality distribution (i.e., stimulating mediocre to average teachers to teach better) should emphasize numbers of rewards, recognizing that this will dilute incentives for teachers who already perform well. Conversely, one interested in promoting excellence should emphasize high performance thresholds and hence smaller numbers of higher-value rewards. But fourth and finally, the need to make such trade-offs can be circumvented, in part, by establishing multiple levels of rewards, as I discuss in the following section.

THE HIERARCHICAL STRUCTURE OF REWARDS

I use the term "hierarchical structure" to refer to the multiple levels of performance recognized under an incentive system and the corresponding multiple levels of rewards. To describe such a hierarchy, one must specify (a) the number of levels, (b) the performance criteria for attaining each level, (c) the rewards offered at each level, and (d), in the case of a career ladder plan, the differentiated roles and responsibilities assigned to teachers at each level. The reward hierarchies in recently discussed state plans range from California's single-step mentor teacher plan to Tennessee's five-step career ladder (Southern Regional Education Board, 1984). I consider here, first, why one would want a multilevel plan and, second, how the multiple levels should be structured.

The main incentive-related reason to establish multiple levels of rewards is to offer effective incentives to a broader range of teachers than can be reached by any single-level plan. To appreciate what "effective incentives" means in this context, consider how teachers at various levels of performance would be affected if there were only a single performance threshold and a single level of reward. As explained in the foregoing discussion of "number of rewards," if the threshold were set high, say at the 90-percentile level, many average and below-average teachers, appraising their own capabilities, would realize that their chances of success percent were small and would have little incentive to compete against such high odds. Only teachers confident of being well above average would have any real incentive to change their behavior in response

to such a plan. If, on the other hand, the performance threshold were set low, say at the 20-percentile level, all or nearly all teachers would be within range of qualifying for a reward, but a different threat to the effectiveness of the incentives would emerge. Because the performance standard would have to be set low, any above-average teacher would qualify at once and hence would face no incentive to improve. By the same reasoning, if the performance threshold were set at some intermediate value, teachers at either end of the performance spectrum would have little incentive to improve. For those at the top, the standard would be too easy; for those at the bottom, it would be too hard. No one-level plan, whatever its performance threshold, would motivate more than a fraction of the teaching force.

A multilevel reward structure can circumvent this problem. With such a structure one can establish a series of performance thresholds of successively greater difficulty, each corresponding to a successively greater reward. In a four-step merit pay system, for example, the highest performance threshold might be set at a level that only 10 percent of teachers can reach, the next threshold at a level 30 percent can reach, and the third at a level 80 percent can reach. (Under a career ladder system, these same thresholds might constitute the criteria for promotion to, say, "master teacher," "senior teacher," and "teacher," respectively.) Not all teachers will be in contention for the highest level, but most will be able to aspire to one of the other performance rewards. No one will be out of range, and nearly everyone will have something to gain from further performance improvement.

How many levels of rewards should there be? Under a pure merit pay system (no ranks or promotions), there would be no need, in principle, to limit the number of steps. One can even envision a continuously variable (infinite level) relationship between performance ratings and pay.[10] In practice, however, the number of gradations is limited by the discriminating power of the performance evaluation instruments. As explained in Chapter III, with the performance rating methods now available, even distinguishing among satisfactory, superior, and outstanding teachers is difficult and entails a considerable probability of error. Attempts to make finer performance distinctions are likely to be frustrated by the imperfections of the measurement art. For this reason alone, limiting the reward structures to three, four, or five levels, as in most of the recently developed state plans, appears the prudent thing to do.

PERFORMANCE THRESHOLDS AND RATIONING

There are two ways in a system of performance-based rewards to establish the various performance thresholds and the corresponding numbers of rewardees. One is to decide first on the levels of performance for which teachers will be rewarded and then to reward all teachers who qualify; the other is to predetermine the numbers of rewards and allow the performance cut-offs to adjust to match them. The former leaves the number of rewards uncertain; the latter involves rationing of rewards. What can be said about the choice between the two?

---

[10]For example, one might calculate salaries according to a formula of the type, SALARY = BASE SALARY + K(PERFORMANCE - BASE PERFORMANCE), according to which a teacher earns an increment of K dollars over base salary for each "point" by which the teacher's performance rating exceeds a specified base level of performance.

An understandable motive for wanting to ration rewards (i.e., to fix their numbers without regard to how many teachers perform well) is the desire to control cost by avoiding having to pay more merit increments than planned. Other motives include the desire to maintain the prestige of rewards by making them relatively exclusive and, in the case of master teacher or mentor teacher career ladder plans, to avoid having more master or mentor teachers than there are assignments for them. Notwithstanding the reasonableness of these motives, rationing rewards by setting quotas in advance alters the incentive effects of performance-based rewards in several undesirable ways.

Rationing creates uncertainties about the performance required to earn a reward. In contrast to a system that offers rewards to all teachers who meet a stated performance standard, one that limits rewards to a specified number or fraction of teachers leaves the performance threshold unknown. How well each teacher must perform to win a reward depends on how well other teachers perform during the same period, not on the individual teacher's success in attaining a performance goal. As explained in Chapter II, more uncertain rewards are less attractive, and the incentives they provide are weaker. Therefore, by adding a new element of uncertainty (in addition to the uncertainty already inherent in any incentive scheme), rationing is virtually certain to diminish the effectiveness of performance-contingent rewards.

Rationing also introduces a form of head-to-head competition among teachers that is absent when the performance threshold for each type of reward is established in advance. In the latter case, no teacher's success threatens another teacher's opportunity (or at least the threat

is not direct and visible).[11] Under rationing, however, allocation of rewards becomes a zero-sum game: each contender's chances decrease if other teachers do well. Thus, it would be in a teacher's interest not to have his or her fellow teachers do well. As a number of writers have pointed out (e.g., Rosenholtz, 1985; Johnson, 1984), incentives could have deleterious effects on schooling if they diminish collegiality and mutual aid among teachers within schools. Such effects are likely to be minor when performance standards are predetermined. In fact, by including school-level as well as individual rewards in the plan (see the comments on "collective rewards," below), one could design incentives to augment rather than diminish collegiality. But rationing could make interteacher rivalry a major negative factor, with adverse consequences for school climate and, ultimately, educational results.

An alternative form of rationing that mitigates some of the aforementioned problems is to restrict the total pool of performance-based pay rather than the number of recipients. With this approach, there would be a predetermined performance standard to qualify for a reward, but the dollar amount of the reward would depend on how many teachers exceeded the specified threshold. Thus, there would be no uncertainty about the level of performance required to win but only about the size of the prize. The adverse effect of uncertainty on motivation under this arrangement is likely to be much less than under the quota system

---

[11] One might argue that even in the absence of quotas there is some reason for teachers not to facilitate other teachers' success, since if too many teachers qualify for rewards it is likely that standards will be raised in the future or the value of rewards reduced. It seems unlikely, however, that such calculations would stimulate anything like the degree of competitiveness likely to develop under a fixed quota of rewards.

discussed above. In addition, potentially destructive head-to-head competition for a limited number of rewards would be avoided. It remains true that one teacher's gain could detract from another's, but only in the limited sense that more teachers' winning would reduce each winner's reward. The loss to any one teacher from his or her immediate colleagues doing well would be so minute, however, that it is unlikely collegiality and mutual aid within schools would be impaired.

In the case of a master or mentor teacher plan, it makes sense to limit promotions to such ranks to the numbers needed to perform special master/mentor functions, but several steps can be taken to minimize the adverse effects. First, although the number of master or mentor teachers may have to be specified in advance, the same need not be true of promotions to intermediate ranks. Rationing can be limited to only the top. Second, even at the master/mentor level, one could consider a dual selection method, in which promotion itself is not rationed but special assignments (and, perhaps, corresponding extra pay) are given only to some of those accorded master rank. For example, 7 percent of a district's teachers might satisfy the criteria for elevation to master status, but only 5 percent might be assigned teacher training and evaluation functions.[12]

An open-ended reward system may seem disconcerting from a management perspective because of the attendant fiscal uncertainty, but the fiscal

---

[12]This arrangement is also attractive for two reasons not directly related to incentive effects, namely that (1) outstanding performance as a classroom teacher is a sufficient criterion for allocating rewards but not necessarily for selecting trainers and evaluators of other teachers, and (2) some teachers who qualify fully for the master or mentor ranks on the basis of performance may not be interested in the nonteaching roles. Differentiating between the rank itself and the associated special assignments is a way of accommodating these concerns.

problems are unlikely to be large or lasting. A system that provides, say, rewards averaging 20 percent of base salary to 25 percent of all teachers involves only a 5 percent increase in the total salary budget, and consequently a fractional error in forecasting the number of rewardees would be likely to translate into no more than a 1 or 2 percent change in the salary budget. This percentage can be reduced by phasing in the program gradually and using the early experience to make better forecasts of how many teachers will qualify for rewards. Moreover, if either more teachers or fewer teachers qualify than expected, budgetary equilibrium can be restored in future periods by raising or lowering the performance thresholds accordingly. Thus, any budget stresses caused by forecasting performance incorrectly are likely to be transient phenomena.

ELIGIBILITY AND PARTICIPATION

The rules governing teacher eligibility and participation in performance-contingent reward systems are another feature that can influence the incentive effects. In this section, I address two issues concerning those rules: (1) whether eligibility for rewards should be linked to seniority, and (2) whether participation in an incentive plan should be mandatory or voluntary.

Seniority and Eligibility for R wards

Under some of the recent state incentive proposals, eligibility for rewards, especially promotions, is tightly tied to seniority. Both the Tennessee plan and the proposed Delaware plan, for example, require three years of teaching at the apprentice level before becoming eligible for the next step up the ladder ("Career Level I"), five years of additional

experience to become eligible for Career Level II, and yet another five years to become eligible for the highest step, Career Ladder III—that is, a minimum of 13 years of teaching to be considered for the highest rank. Until these seniority requirements are satisfied, teachers may not earn promotions or the accompanying pay increases regardless of the excellence of their teaching.

Viewing the career ladder plans as leadership systems, one can see some rationale for these seniority requirements, but viewing the plans as incentive mechanisms, one has to question their effects. The applicable general principle is this: a reward delayed is a reward diminished. Specifically, it seems clear that any attractiveness that a performance-contingent reward system might otherwise have for new or prospective teachers would be attenuated severely by the long delay before superior performance could earn a substantial reward. For example, assuming a moderate time discount rate (by current standards) of 10 percent per year, the present value of a permanent (lifetime) increment in pay beginning five years from now is only 67 percent as great, and that of an increment beginning ten years from now only 39 percent as great, as the present value of the same annual increment beginning today. Or, putting it differently, an offer of a 25-percent permanent pay increase to a teacher who attains master rank is worth the full 25 percent to someone immediately eligible but is equivalent to only about a 9-percent increase to someone who will not be eligible for ten years.[13] Consequently, while such a

[13]The 9-percent figure is obtained by comparing the present value, at a 10 percent discount rate, of an immediate increase payable over an assumed career of 40 years with that of an increase of the same amount 10 years from now, payable for the remaining 30 years of the same 40-year career.

reward structure may stimulate the performance of teachers who have been in the system for many years, it is likely to do little for those debating, after, say, two or three years, whether to remain in teaching.

Five years is a long time to wait, either to reward a teacher for doing well or penalize a teacher for doing poorly. During those intervals, waiting for seniority to accumulate, teachers are effectively off the incentive system. Neither their status nor their pay depends on performance until the appointed time has elapsed. Thus, in any given year, only a fraction of the teaching force feels any direct incentive to perform.

What this suggests to me is that if there are to be career ladders, with eligibility for high rank limited to seasoned, veteran teachers, there should also be shorter-term performance-based rewards along the way. These could take the form of performance bonuses, over and above the pay scale associated with a particular rank, or movements along a performance-based salary schedule. The specific form is immaterial. What matters is the principle: rewards should be performance-contingent for as many teachers as possible, as much of the time as possible, to maximize the incentive to teach well.

The treatment of seniority may turn out to be the most important practical distinction between career ladders and merit pay. The main arguments for linking eligibility to seniority under a career ladder system do not apply when rewards consist of merit pay alone. When a teacher's performance is rewarded with a special increase in pay, there need be no implication that he or she is senior, higher-ranked, or in a leadership position relative to teachers who do not earn similar performance awards. Consequently, there is less reason to insist that teachers who

receive the larger rewards must generally be older or more experienced than those who do not. Moreover, merit pay is reversible in a way that promotion is not, and without the trauma that accompanies demotion to a lower rank. This greater flexibility with respect to, among other things, eligibility for rewards is one of the main advantages of merit pay over the career ladder approach.

Voluntary or Mandatory Participation?

Some proposed incentive plans would make participation voluntary, letting teachers opt out without explicit adverse consequences, while others would automatically cover the entire teaching force. The suggestions for voluntariness seem to be motivated by such considerations as the desire to make performance-based rewards less threatening and hence more acceptable to the affected parties (Hatry and Greiner, 1984). Also, in the case of certain career ladder plans in which promotion carries with it leadership roles and nonteaching responsibilities, voluntariness is essential to avoid thrusting such roles on unwilling teachers.[14]

Whatever the rationale for voluntariness, one of its consequences would be to dilute incentives for better teaching performance. Among the most likely nonvolunteers in a voluntary system are teachers who perform below average, know it, and do not intend or expect to improve.[15]

---

[14]As noted earlier, however, it is possible to make the assumption of non-teaching roles voluntary without making participation in the incentive program voluntary as well.

[15]Other likely nonvolunteers include teachers of varying levels of performance who find competition distasteful, possibly including some who deem competition among teachers wrong in principle, and teachers who object to summative evaluation in general or to the particular evaluation methods or criteria adopted by the state or district in question.

By remaining outside the incentive system, such teachers would be shielded (albeit only partially and temporarily, for reasons I discuss below) from the consequences of working under a system of performance-based rewards.

The extent to which voluntariness can shield nonvolunteers from the risks and uncertainties of incentives is unclear because one aspect of the design of a voluntary system has always been left vague: how would nonvolunteers be rewarded compared with teachers who do choose to participate but fail to win merit pay or promotion? Under a "mild" incentive plan, the answer is simple: nonvolunteers receive their regularly scheduled salary increases but do not qualify for merit increments or promotion to higher ranks. But consider a more potent merit pay plan under which teachers who fail to meet minimum performance standards receive no pay raises at all. Under such a plan, a teacher who chooses to participate but falls short of the standard would be denied a raise; but what of the teacher who opts out? To give such a teacher a "regular" raise seems unfair, on the grounds that those who try and fail should not be treated worse than those who do not try at all. On the other hand, to deny all raises to nonvolunteers would make a mockery of the nonparticipation option. I see no acceptable way out of this dilemma. If teachers are to be penalized for teaching poorly as well as rewarded for teaching well, it is neither feasible nor fair to let some teachers opt out.

The idea of making participation voluntary seems to arise from the understandable desire to have a system with rewards but no punishments—one in which some teachers win but no teacher loses. But such a system is probably not feasible in the long run; nor, if feasible, would it be desirable. If the welfare of children is the ultimate objective,

as everyone seems to agree, then encouraging poor teachers to leave teaching is as important as stimulating teachers to improve or inducing superior teachers to enter and stay. Allowing incompetent teachers to escape the consequences of their performance runs counter to the whole point of shifting to an incentive approach.

In any event, the notion of a "no-loser" system--one in which some teachers earn extra pay for good performance but no one earns less--ultimately must break down unless there is a permanent infusion of extra funds from the outside, effectively earmarked for extra, performance-based pay.[16] Some states, to be sure, have undertaken to provide special funds for performance incentives or increase state aid as incentives are introduced, but the former seem to be viewed as temporary, start-up contributions, and the latter are likely to be one-time events. It is likely, in the longer run, that performance-contingent increments in pay will become integral parts of district salary structures, funded out of general education revenues. If so, higher pay for good teachers will have to be balanced out by lower pay for poor teachers. Explicit reductions need never occur. Instead, salaries for those who do not earn performance-based increments could be allowed gradually to decline relative to pre-incentive expectations, and those who opt out would gradually become losers even though there would be no explicit reductions in their pay.

---

[16]Note that the availability of extra outside aid, nominally for the support of an incentive system, is not sufficient. There must also be effective provisions to ensure that such aid is used to maintain the basic salary structure at "what it would have been" in the absence of merit pay, rather than, for example, to fund additional rewards or hire additional teachers. Such earmarking provisions are notoriously difficult to design, and it is not clear that a state could make them effective even if it deliberately set out to do so.

If the goal is improved educational quality for children, the idea that poor-performing teachers will be made worse off by the incentive system should be seen as a positive feature, not as something to be avoided by making participation voluntary. Making teaching less attractive and economically viable to those who teach poorly and do not improve may be as effective a method of raising quality as making teaching more rewarding to those who teach well. It is not surprising that those who have to develop incentive plans in the real world--to bargain with teachers' unions and generate political support--should emphasize the "carrot" and not the "stick." The danger, however, is that states and school districts may actually attempt, at least in the short run, to create "hold harmless" systems in which no one's pay falls below what it would have been under the old regime; but this would serve mainly to dilute and delay the contribution of incentives to a higher-quality teaching force.

CONCLUSIONS AND IMPLICATIONS

The degree to which the potential benefits of performance-based rewards would be realized under a state or local incentive system depends on a series of specifics of system design. It is reasonably clear which design features, or combinations of features, offer the strongest incentive effects. In some cases, however, there are trade-offs to be made between the strength of the incentive offered to one category of teachers and to another (e.g., between incentives for newly hired versus senior teachers) and between short-run and long-run effects on performance. In other cases, states or school districts may reasonably choose to compromise

the effectiveness of incentives to achieve other educational goals or to win acceptance of the incentive approach.

The strength of incentives depends on certain clusters of design features in the following ways:

First, the potential effectiveness of an incentive plan is positively associated with the sensitivity of pay and status to performance and with the proportion of salary that is contingent on performance. Specifically, an important line of demarcation is between plans that leave "regular" salary structures untouched, merely appending to them performance-based rewards, and those that make all pay increases contingent on performance. The latter, obviously, are likely to be more potent, especially in affecting the retention decisions of low performers.

Second, several aspects of timing are important. In general, the inducement to raise one's performance and to sustain the increase over time is enhanced by linking rewards to performance in each time period, making the rewards short-term or reversible, and avoiding long waits to establish eligibility for successive rewards. However, the advantages of sensitive and variable rewards must be traded off against the need for reliable performance measurement, which implies multiple assessments of teaching performance over more than a single year.

Third, the incentive system is likely to be more effective if it is structured so that large numbers of teachers are within "striking distance" of earning a reward, which implies that there should be multiple levels of rewards corresponding to multiple performance thresholds. This criterion is <u>not</u> satisfied by offering a sequence of rewards tied to different levels of seniority, since such an arrangement still offers

only a single level of reward to a teacher at any given stage of his or her career.

Fourth, effectiveness would be enhanced by making the plan applicable to all teachers, which means (a) ensuring that teachers can earn rewards at each stage in their careers, (b) avoiding delays to establish eligibility (even merit-based differentials in starting salaries should be considered), and (c) making participation for everyone automatic, with no option to avoid the consequences of performance-based rewards.

Fifth, the inducement to perform well is reinforced by clarity about what is required to earn rewards, which implies, in addition to specificity about the performance criteria themselves, the avoidance of arbitrary quotas on numbers of rewardees.

In light of the above, it can be seen that some of the state plans now being implemented and proposed have features that are not especially conducive to the strength of incentives and that, consequently, must be justified, if at all, on other grounds. The most popular approach, at least among the plans that have arrived earliest at the stage of implementation, is a career ladder of three, four, or five steps. Promotions along such ladders are permanent, eligibility for promotion is tightly linked to seniority, participation is often voluntary, and the performance pay increments take the form of lump-sum additions to regular scheduled salaries. Because the proferred rewards are substantial, and buttressed by the recognition associated with promotion, it is likely that they will produce significant inducements to perform; however, their potential effectiveness is less than what it could be in a number of respects. Perhaps most important are the timing aspects of the recently proposed

plans. Once elevated on the career ladder, a teacher does not become eligible for further performance-based rewards for a number of years, and, in the interim, retention of rewards is not contingent on sustained performance. Thus, the inducement to perform is intermittent rather than continuous, and the danger of backsliding is considerable. Also, the deferral of eligibility until a certain level of seniority is attained greatly diminishes the attractiveness of the rewards to prospective and recently hired teachers. Also notable is the lack of an explicit "stick" to go with the "carrot." Because performance-based pay is merely superimposed on regular pay, teachers continue to receive pay increases even if they fail to reach performance thresholds; and because participation is voluntary, they continue to receive pay increases even if they opt out. This reduces sharply the likelihood of a quality-enhancing effect on teacher turnover and also makes it easier for teachers not interested in competing to continue with "business as usual." In addition, incentives are weaker than they might be under such career ladders because only one performance threshold faces a teacher at any point in his or her career. For some teachers, that threshold will be too easy to attain; for others, too hard. Either way, the incentive to teach better will be undercut.

What alternatives might be worth considering? The appropriate objective in my view, is to try to ensure that in each period as many teachers as possible have something to gain from good performance and something to lose from poor performance. Without attempting to draw out the full design implications, I suggest that the following features would contribute to attaining this goal:

- o Performance-contingent pay increases. A system in which annual percentage increases range from zero (for teachers whose performance is below minimum standards) to a multiple of the average rate. Note that this is not incompatible with having a career ladder or a corps of master teachers, since a different performance-contingent pay icnrease schedule could be established for each rank.
- o Universal coverage. Every teachers' pay would be determined according to the foregoing schedule. (Note that this does not preclude a voluntary system of competition for higher ranks and/or for special non-teaching roles).
- o Differentiation among multiple levels of performance. There should be gradations of rewards (pay increases) corresponding to gradations of performance, not a "yes or no" decision as to whether a teacher qualifies for a reward.
- o Explicit, predetermined criteria for reward. The performance levels required to earn various rewards should be known in advance, and rewards should not be limited to fixed numbers or percentages of teachers.

REFERENCES

Boyer, Ernest, High School, Carnegie Foundation for the Advancement of Teaching, New York: Harper & Row, 1983.

Cohen, David K., and Richard J. Murnane, "The Merits of Merit Pay," Winter 1985.

Cresap, McCormick and Paget, Teacher Incentives: A Tool for Effective Management, prepared for National Association of Elementary School Principals, American Association of School Administrators, National Association of Secondary School Principals, Reston, VA, 1984.

Cresap, McCormick and Paget, State of Delaware: Career Development and Accountability Program, Draft PLan, Washington, D.C., 1985.

Darling-Hammond, Linda, Arthur E. Wise, and Sara R. Pease, Teacher Evaluation in the Organizational Context: A Review of the Literature, In Review of Educational Research, 53-3, Fall 1983, 285-328.

Education Week, "Changing Course: A 50 State Survey of Reform Measures," February 6, 1985.

Goodlad, John I., A Place Called School, New York: McGraw-Hill, 1983.

Goodwin, David, and Lana D. Muraskin, Regulating Excellence: Examining Strategies for Improving Student and Teacher Performance, National Association of State Boards of Education, Alexandria, VA, March 1985.

Hatry, Harry P., and John M. Greiner, Issues in Teacher Incentive Plans, The Urban Institute, Washington, D.C., January 1984.

Johnson, Susan Moore, "Merit Pay for Teachers: A Poor Prescription for Reform," Harvard Educational Review, 54-2, May 1984.

Medley, D. M., Teacher Competency Testing and the Teacher Educator, Association of Teacher Educators and the Bureau of Educational Research, University of Virginia, Charlottesville, 1982.

Millman, Jason (ed.), Handbook of Teacher Evaluation, National Council on Measurement in Education, Sage Publications, 1981.

National Commission on Excellence in Education, A Nation at Risk: The Imperative for Educational Reform, Washington, D.C., April 1983.

Peterson, Paul E., "Did the Education Commissions Say Anything?" The Brookings Review, Winter 1983.

Rosenholtz, Susan J., "Political Myths About Education Reform: Lessons from Research on Teaching," Phi Delta Kappan, January 1985, 349-355.

BEST COPY AVAILABLE