# Texas Federation of Teachers
## aft, afl-cio

3000 South IH-35, Suite 175 ◆ Austin, Texas 78704-6536
512/448-0130 ◆ 1-800-222-3827


Testimony of Eric Hartman
Legislative Director, Texas Federation of Teachers
Before the Senate Select Committee on Education Reform and Public School Finance
Monday, February 27, 2006

The Texas Federation of Teachers represents more than 51,000 teachers and other education employees in public schools and higher education. We are deeply committed to improving education.

Our local affiliates in major urban school districts, such as Dallas, San Antonio, and Corpus Christi, currently are working with administrators in those districts on an array of programs to improve the quality of instruction. A major current emphasis of these efforts is on meaningful professional development to equip teachers with the knowledge and tools they need to be more effective in the classroom. Many of our affiliates also are teaming up with their district administrators to redesign schools in need of improvement.

These current efforts extend a commitment to education reform that TFT has upheld consistently in support of a series of legislative initiatives that have had positive results for our students. We supported the standards-based reforms of the 1980s such as HB 72. In the 1990s we supported the standards-based revision of the Education Code, the governor's reading initiative, and the legislation also initiated by then-Gov. Bush to stop social promotion. TFT supported the toughening of high-school graduation standards and the adoption of the Recommended Curriculum, a more rigorous, college-preparatory curriculum, as the standard for all students. TFT also has supported Gov. Perry's initiatives to establish master-teacher programs. And we have supported legislation by Sen. Shapiro to improve the financial accountability of school districts.

We are always open to changes in policy if we are convinced that they are changes for the better. However, not every idea labeled "reform" is worthy of support. And before you embark on a new round of policy changes in the name of reform, you have unfinished business that must be addressed.

First Things First

In 2003, just as our school districts were reaching the limits of their local capacity to raise revenue to meet state requirements, and just as a host of new requirements took effect, we saw the state cut back on its commitments to public education. The Student Success Initiative, master-teacher programs, kindergarten and prekindergarten grants, school

textbooks, school employees' health insurance, and school employees' pension benefits all suffered cuts. The 2005 session did not reverse many of those cuts, and in fact further cuts in employee benefits were enacted.

Again, this negative trend in compensation occurred as our schools face tremendous challenges to meet new state requirements, such as the mandate that all students take the far more demanding Recommended High School Curriculum and demonstrate proficiency in a full array of subjects including higher-level English, social studies, math, and science courses, in order to graduate. Just to give you a sense of the scale of this challenge, I would note that the Texas Public Policy Foundation 2003 Briefing Book cited a need for a 60-percent increase in resources at the high-school level just to meet the requirements of the Recommended High School Curriculum.

Another measure of the challenge we face is the percentage of teachers assigned to teach outside their field of preparation and expertise–according to a 2004 study by Dr. Ed Fuller, research associate at UT-Austin, 62.7 percent of teachers at low-performing Texas schools were teaching out of field, compared to 20 percent at high schools rated acceptable or better.

How to Improve Compensation

The legislature's priority in the area of teacher compensation should be to reverse the erosion of benefits and to start significantly increasing pay and benefits again, as we saw the state legislature do with the $3,000 teacher pay raise of 1999, the increases in the retirement multiplier in 1999 and 2001, and the allocation of state funding to school employees' health care in 2001.

The first step should be to restore the full $1,000 health-care stipend for all school employees. This step would go a long way toward restoring the credibility of the state's commitment to support Texas teachers and school employees.

In the longer run, filling our classrooms with appropriately trained and qualified teachers will take a much larger increase in compensation. The Texas Higher Education Coordinating Board in October 2002 found that more than 50,000 classroom slots were filled by teachers assigned to teach out of field. The Coordinating Board said it would take an increase of 30 percent in teacher pay, adjusted for inflation, in order to provide the estimated 360,000 teachers, teaching subjects they know how to teach, who will be needed by 2015. Roughly speaking, that would require a $12,000 increase in average teacher pay, in constant dollars.

In addition to a substantial, state-funded increase in compensation for Texas teachers across the board, another large part of the solution will be to restore the capacity of school districts to enhance compensation on their own initiative. We need increased per-pupil funding for Texas school districts. Improving district capacity to increase compensation for all employees and provide other resources to support student achievement is a key benchmark of success for any proposed solution to our state's school-finance system.

Differential Pay

TFT also supports several forms of differential pay, on top of a foundation of competitive base pay and benefits. We support:

–stipends for mentors;
–extra pay for teachers in shortage subjects;
–extra pay for demonstrating higher levels of teaching mastery such as National Board Certification that clearly correlate with higher levels of student achievement;
–and extra pay to ensure that hard-to-staff schools can recruit teachers appropriately qualified to teach their subject.

On the question of so-called "performance pay" tied to a "value-added" measure that links students' achievement test scores to individual teachers, we agree with analysts from the RAND Corporation who reviewed the "value-added" research and concluded that: "The research base is currently insufficient for us to recommend the use of VAM [value-added methodology] for high-stakes decisions" regarding employee evaluation and compensation.

The Rand researchers concluded that no current value-added models account credibly for variables, beyond the current-year teacher's performance, that influence student achievement: e.g., controlling for student backgrounds, disentangling school and district effects from teacher effects, disentangling the effects of earlier teachers and schools from estimated teacher effects, effects of incomplete records, effects of missing records, and criteria for linking particular students to particular teachers.

As Dr. Hanushek back in 2004 testified before a Texas House committee that existing performance-pay models around the country "have failed." Positive results of such experiments should not be assumed just because they begin with good aims and good will.

The Texas experience with the career ladder illustrates further pitfalls of a program that begins with fanfare and then is abandoned when too many teachers qualify for the extra pay. The mistrust of state incentive pay caused by that experience is still strong in our schoolhouses.

Before you are tempted to embrace experimental models from other states, please pay heed to the context of those models–such as the collective-bargaining agreements, negotiated between school districts and their local teacher union, that assure teachers a compensation system will be a legally enforceable deal in states like Colorado. The foundation for a plan like the one teachers have supported in Denver is lacking in Texas.

It is worth taking a moment to consider the contrast between the Houston plan announced last month and the Denver plan. The Houston plan announced last month, for example, does nothing to address Houston ISD teachers' low base salaries, which are not competitive with pay in nearby districts. The plan is confusing and not understood by teachers. The biggest bonuses are reserved for administrators.

Teachers who spoke out against the plan last month noted that teachers of subjects not tested on the exams will not be eligible for most of the bonus money, and they will not have much chance to affect the student test performance that may trigger a school-wide reward. This Houston plan reflects what happens when a mechanism is lacking for meaningful partnership and collaboration with teachers in designing and implementing an incentive-pay plan.

In contrast, the Denver plan was carefully negotiated over years of formal, contractual bargaining between the Denver teachers' union and the school district. The teachers of Denver had to approve it by majority vote in a referendum before it could take effect, and no teacher already on the faculty could be compelled to participate in the plan. And whether they participate in the plan or not, all teachers in Denver are entitled to ongoing cost-of-living increases. Under the Denver plan, a teacher who earns a bonus has a legally enforceable right to that bonus.

Before the Denver plan could take effect, the citizens of Denver also had to vote for increased school funding that will raise teacher salaries by 12.5 percent overall. The plan has a dedicated funding source. And the use of test scores to gauge performance is just one small facet of the overall Denver plan.

The Houston and Denver plans do have one thing in common, though. They are both experiments, not programs of proven effectiveness. The same can be said of the Teacher Advancement Program's project in Richardson ISD and other local initiatives around the state. Before committing the state to a full-scale, statewide implementation of any incentive-pay program, wouldn't it be wise to see how these pilot projects perform?

RAND Education Research Brief RB-9050-EDU (2004)

# The Promise and Peril of Using Value-Added Modeling to Measure Teacher Effectiveness


## Abstract

**Value-added modeling offers the possibility of estimating the effects of teachers and schools on student performance, a potentially important contribution in the current environment of concern for accountability in education. These techniques, however, are susceptible to a number of sources of bias, depending on decisions about how the modeling is executed and on the quality of the data on which models are based. If teachers are to be held accountable for the performance of their students, strategies for measuring the impact of their work must be refined or, at least, the uncertainties of these measurements must be taken into account in assessing the impact of teachers and schools on student performance.**


Value-added modeling (VAM), a collection of statistical techniques that uses multiple years of student test score data to estimate the effects of individual schools or teachers, has recently garnered a great deal of attention among both policymakers and researchers. For example, several states — including Tennessee, Pennsylvania and Ohio — are providing at least some of their schools and school districts with feedback about their performance based on VAM, and, in some statehouses, the idea of using VAM results to evaluate and reward administrators and teachers has been discussed.

This interest on the part of policymakers reflects the promise of VAM, but many technical issues must be considered in the execution and application of VAM to ensure that policy decisions are based on sound information. Although there have been reviews of particular approaches, no previous reviews carefully compared recent VAM efforts or systematically discussed the wide variety of issues they raise. To address this problem, RAND researchers, funded by the Carnegie Corporation of New York, undertook a systematic review and evaluation of leading approaches to VAM. The goals of this investigation were to

- delineate the technical issues raised by the use of VAM for measuring teacher performance
- evaluate the practical impact of decisions regarding modeling techniques, variations in the quality of the data used in modeling processes, choices of outcome measures, and techniques for sampling student performance
- identify gaps in the literature that could benefit from further research
- inform the debate among both researchers and policymakers about the potential of VAM.

In addition, the research team estimated the effects of math teachers for students in Grades 3-5, using math scores from a sample of schools in a large suburban district. This independent analysis permitted examination of the effects of certain variations in modeling strategies.

## Value-Added Modeling Has the Potential to Identify Effects of Teachers on Student Performance

VAM attempts to determine the incremental effects of inputs into education, controlling for the prior achievement level of students. In practice, VAM is used to estimate the unique contributions of the school or teacher on students' progress over the course of a year rather than the cumulative effects of education or student background factors.

Two factors have contributed to recent interest in VAM. First, in theory, VAM has the potential to separate the effects of teachers and schools on student performance from the powerful effects of noneducational factors such as family background. This isolation of the effects of educational and noneducational factors is critical for accurate evaluation of schools and teachers. Second, some recent VAM studies purport to show very large differences in effectiveness among teachers. If these differences can be substantiated and can be causally linked to specific characteristics of teachers, significant improvements in education could be made through the selection of effective teachers or through training to improve teacher effectiveness.

## Variations in Teachers Affect Student Performance, but Size of Effect Is Uncertain

The recent literature on VAM suggests that teacher effects on student learning are large, accounting for a significant portion of the variability in growth, and that they persist for at least three to four years into the future. RAND researchers critically evaluated the methods used in these studies and the validity of the resulting claims. They concluded that teachers do, indeed, have discernible effects on student achievement and that these teacher effects appear to persist across years.

The shortcomings of existing studies, however, make it difficult to determine the size of teacher effects. Nonetheless, it appears that the magnitude of some of the effects reported in these studies is overstated. To determine the true size of teacher effects, several important statistical and psychometric issues must be addressed.

We group these issues into four categories: basic issues of statistical modeling; issues involving omitted variables, confounders, and missing data; issues arising from the use of achievement test scores as dependent measures; and uncertainty about estimated effects.

### Impact of Alternative Statistical Modeling Strategies on Estimates of Teacher Effects

Modeling choices could have a significant impact on estimates of teacher performance. The problem of small classes is a case in point. When the number of students taught by a particular teacher is small, estimates of teacher effects can be heavily influenced by the performance of only a few students. One modeling approach to addressing this problem involves using data from small classes without adjusting for class size. This approach, however, tends to classify too many teachers of small classes as either highly effective or highly ineffective. An alternative approach, used in many of the most prominent recent VAM studies, "shrinks" estimates for individual teachers back toward the overall mean. That is, estimates of the effects of teachers who teach small numbers of students are statistically adjusted so that they are similar to the average effect of all teachers. This approach offsets the problem of distortions in the overall effects of teachers, but it makes identifying particularly effective or ineffective teachers who teach small classes

considerably more difficult.

**Impact of Omitted Variables, Confounders, and Missing Data on Estimates of Teacher Effects**

In VAM, analysts rely on observational, rather than experimental, data. Reliance on such data can lead to inaccuracy in estimates of teacher effects due to (1) differences between schools or classrooms that are not fully controlled in the analysis (such differences "confound" the results) and (2) shortcomings of the data collected within schools.

***Impact of Absence of Controlled Comparisons Across Schools.*** When differences between schools are not experimentally controlled, influences on student learning by factors other than teachers, such as other characteristics of the school in which the teacher works, may not be properly accounted for. For instance, if students attending different schools differ in ways that are likely to affect both achievement and growth in achievement and if the composition of the school's students (e.g., the proportion of students eligible for free and reduced-price lunches) affects these outcomes, bias in estimates of teacher effects can occur.

Some recent work on this topic suggests that variations in individual student characteristics have little influence on estimated teacher effects, but our own exploration suggests that the composition of the school had a great impact on estimates of teachers' effectiveness. We conducted a limited investigation of performance in mathematics — three grades in one school district were examined — and found that the composition of the school does affect growth in some settings. Thus, if variations in the composition of the school are not taken into account, these omitted variables may produce bias in applications of VAM. Because true teacher effects might be correlated with the characteristics of the students they teach, current VAM approaches cannot separate effects caused by the composition of the school from teacher effects.

Also difficult to disentangle from the effect of the students' current teachers are other characteristics of schools (i.e., characteristics other than the composition of the student body), of districts, or of prior teachers. If these variables are omitted from the analysis, their effects are subsumed by the estimated teacher effects. Alternatively, if such effects are included in models and if teachers of differing effectiveness cluster at the school or district level, part of the true teacher effects will be attributed to schools or to districts. Both approaches may result in biased estimation of the true teacher effects. Analysts must decide which potential error is more acceptable.

***Impact of Missing Data.*** Longitudinal student achievement data will inevitably be incomplete. Information regarding the performance of individual students, as well as data linking students to teachers, may be lacking. Estimates of teacher effects may be sensitive to both the nature of missing data and the analytic approach used to address the problem. For example, if the test scores of low-performing students are missing, the scores of high-performing students will have a disproportionate impact on estimates of teacher effectiveness, possibly making teachers appear more effective than is, in fact, the case. Little is currently known about the effects of missing data on VAM estimates of teacher effects, but the potential for bias is large because the factors that contribute to missing links and missing test scores are common: Students are mobile, with large proportions transferring among schools every year.

**Effects of Using Achievement Tests as an Outcome**

VAM uses measures of student achievement to define and estimate teacher effects, but these achievement measures are limited in several ways. Changes in the timing of tests, the weight given to alternative topics, or the methods used to create scores from students' responses (the "scaling" of the test) could affect conclusions about the relative achievement or growth in achievement across classes of students. Such changes would, in turn, change estimates of teacher effects. In some cases, the effects could be substantial. For example, in a middle school in which curriculum is differentiated, a test emphasizing advanced content may favor teachers instructing the most able students, while a test emphasizing more basic content may boost the estimated impact of those teaching less advanced students.

**Effects of Sampling Error**

Sampling error is another potential source of error in VAM estimates. Estimates of teacher effects have larger sampling errors than estimates of school effects because of the smaller numbers of students used in the estimation of individual teacher effects. Thus, some estimates of interest will be too unreliable to use. Even so, for some purposes, such as identifying teachers who are extremely effective or ineffective, the estimates might be sufficiently precise. However, for other purposes, such as ranking teachers, the uncertainty in the estimates is likely to be too large to allow anything to be said with any degree of confidence.

**Recommendations**

Using VAM to estimate individual teacher effects is a recent endeavor, and many of the possible sources of error have not been thoroughly evaluated in the literature. The goal of this study was to identify possible sources of error and bias and evaluate what is known at this point. To improve the quality and usefulness of VAM in the future, the authors recommend that researchers

- develop databases that can support VAM estimation of teacher effects across a diverse sample of school districts or other jurisdictions
- develop computational tools for fitting VAM that scale up to large databases and allow for extensions to the currently available models
- link estimates of teacher effects derived from VAM with other measures of teacher effectiveness as a means of validating estimate effects
- conduct further empirical investigation on the impact of potential sources of error in VAM estimates
- determine the prevalence of factors that contribute to the sensitivity of estimated teacher effects
- incorporate decision theory into VAM by working with policymakers to elicit decisions and costs associated with those decisions and by developing estimators to minimize the losses.

**The Bottom Line**

The current research base is insufficient to support the use of VAM for high-stakes decisions, and applications of VAM must be informed by an understanding of the potential sources of errors in teacher effects. Policymakers, practitioners, and VAM researchers need to work together so that research is informed by the practical needs and constraints facing users of VAM and so that implementation of the models is based on the kinds of inferences and decisions the research currently supports. If teachers are to be held accountable for the performance of their students, they deserve the best measurement of their effects on students that we can provide.

---



---

This product is part of the RAND Corporation **research brief series**. RAND research briefs present policy-oriented summaries of individual published, peer-reviewed documents or of a body of published work.

This research brief describes work done for **RAND Education** documented in *Evaluating Value-Added Models for Teacher Accountability* by Daniel F. McCaffrey, Daniel M. Koretz, J.R. Lockwood, and Laura S. Hamilton, MG-158-EDU, 2004, 154 pages, (**Full Document**). MG-158 is also available from RAND Distribution Services (phone: 310-451-7002; toll free: 877-584-8642; or email: order@rand.org).



RB-9050-EDU (2004)

Ed Issues "Merit Pay"

# Merit Pay and the Evaluation Problem: Why Most Merit Pay Plans Fail and a Few Survive

RICHARD J. MURNANE
*Harvard University*

DAVID K. COHEN
*Michigan State University and Harvard University (on leave)*

*Richard J. Murnane and David K. Cohen use the framework of microeconomics to account for the short lives of most merit pay plans. They demonstrate that teaching is not an activity that satisfies the conditions under which performance-based pay is an efficient method of compensating workers. They then show that merit pay plans survive in a few school districts, in part because the districts are special and in part because the merit pay plans are quite different from conventional notions of performance-based pay.*

Designing a compensation system that provides strong incentives for employees to pursue organizational goals is a challenge every organization faces. Merit pay for teachers is often suggested as a compensation system that will enable public school systems to meet this challenge. Yet the promise of merit pay is dimmed by knowledge of its history; most attempts to implement merit pay for public school teachers over the last seventy-five years have failed.

The first part of this paper uses microeconomics, the intellectual home of merit pay, to explain the failures of most merit pay plans. We show that merit pay, even taken on its own terms, does not provide a solution to the problem of motivating teachers. The second part of the paper investigates why merit pay survives in a very few school districts in the United States. The analysis is based on interviews we conducted with a great many teachers and administrators in six school districts with enduring merit pay plans. We explain that in these special districts, merit pay contributes to the solution of problems quite different from the problem of motivating teachers.

*Harvard Educational Review* Vol. 56 No. 1 February 1986

0017-8055/86/0200-0001$01.25/0

## Compensation of Public School Teachers

More than 99 percent of public school teachers in the United States work in districts that employ uniform salary scales.[1] Under such contracts, a teacher's salary is determined exclusively by educational credentials and years of teaching experience. All teachers with the same credentials and experience receive the same salary, irrespective of subject specialty or perceived performance. Typically, each school district sets its own salary scale or negotiates it with the local teachers' union through collective bargaining.

The limitations of uniform salary scales have been well documented; there is no financial reward for superior performance and no financial penalty, short of dismissal, for inferior performance (Hanushek, 1981). Many critics of uniform salary schedules argue that improving the quality of education offered by public schools requires a change from uniform salary schedules to a compensation scheme that bases a teacher's compensation on performance, as measured either by gains in student test scores or by supervisors' evaluations of the teacher's actions in the classroom. Such performance-based compensation plans are typically called merit pay.

Merit pay is an old idea. In 1918, 48 percent of U.S. school districts sampled in one study used compensation systems that they called merit pay (Evendon, 1918, as reported in Johnson, 1984). Little is known about these early plans, except that most did not last. In 1923 the National Education Association (NEA) reported that 33 percent of sampled districts used merit pay (NEA, 1923, p. 52), and a subsequent NEA survey reported that 18 percent of districts surveyed awarded merit pay (NEA, 1928, pp. 230-240).

Interest in merit pay waned during the 1940s and early 1950s as the vast majority of public school districts in the United States adopted uniform salary schedules. Between 1939 and 1953 the number of school systems in cities with populations of more than 30,000 that used merit pay fell from 20 to 4 percent (Porwoll, 1979, p. 26).

Sputnik rekindled interest in merit pay by raising questions about the effectiveness of American schools. During the 1960s approximately 10 percent of U.S. school districts had merit pay plans, most of which fared no better than their predecessors. By 1972 the number of districts using merit pay had fallen to 5.5 percent (Porwoll, 1979). A 1978 survey of the 11,500 U.S. school districts with enrollments of 300 or more found only 115 with merit pay plans (that is, 4 percent of the districts that responded to the survey and 1 percent of the districts to whom the questionnaire was sent). Moreover, the majority of districts that reported having tried and dropped merit pay indicated that their plans lasted less than five years (Porwoll, 1979, p. 41).

Thus, the history of merit pay suggests that while interest in paying teachers according to merit endures, attempts to use merit pay do not. Moreover, teacher union resistance cannot account for the demise of most merit pay plans, for most plans predated unions or failed in nonunion districts. We must search for other

[1] The 99 percent figure was derived from data presented by Calhoun & Protheroe (1983).

explanations. We believe that the most powerful ideas for understanding why merit pay plans fail can be found in the literature of economics. Specifically, we turn to economic analyses of employment contracts, a growing field within microeconomics.

## Why Most Merit Pay Plans Fail

### *The Contracts Literature: A Framework for Analysis*

One branch of microeconomics, which we will call the contracts literature, examines the costs and benefits associated with using different types of employment contracts to compensate workers engaged in particular kinds of production activities. The following assumptions underlie this literature:

1. Workers' preferences are not completely consonant with the employing organization's goals. Workers prefer to work less hard than the organization would like if there are no adverse consequences for them.
2. Monitoring the output or actions of individual workers is costly.
3. Imperfect monitoring will induce workers to attempt behavior that makes them appear productive relative to other workers but in fact is contrary to the goals of the organization. Williamson (1975, p. 9), an important contributor to the contracts literature, labels this behavior "opportunistic" and defines it as "self-interest seeking with guile."

As seen from the perspective of the contracts literature, the type of employment contract an organization should adopt depends on the type of work employees perform. This is because the cost of evaluating workers' output, the cost of evaluating workers' actions, and the potential for opportunistic behavior all depend on the nature of the production activity.

The perspective provided by the contracts literature is helpful in analyzing merit pay for three reasons. First, this literature takes seriously the evaluation problem. It explicitly acknowledges that evaluating worker performance is costly for management and that imperfect evaluations—defined as less than perfect knowledge of all worker actions—may elicit unpredicted and potentially destructive responses from workers. It is this evaluation problem that has plagued most attempts to introduce merit pay into public education.

Second, the contracts literature emphasizes the importance of trade-offs between the gains from providing incentives for employees to work hard and the costs of various ways of evaluating workers' contributions. Implicit in this emphasis on trade-offs is the often neglected recognition that a merit pay system that brings about modest increases in teachers' effort levels might not be worthwhile if the costs of the measures taken to evaluate teacher performance are extremely high.

Third, the contracts literature focuses attention on the nature of the production activity in which workers are engaged. It explains why an analysis of the production activity provides the best clues to the responses that particular compensation plans will elicit. We will argue that compelling explanations for the failure of most merit pay plans must focus on the nature of teachers' work. In the following

sections we use the framework provided by the contracts literature to explain why neither "new style merit pay" nor "old style merit pay" is an effective strategy for motivating teachers to achieve high performance levels.

*New Style Merit Pay: A Piece-Rate Compensation System*

"New style merit pay" (Bacharach, Lipsky, & Shedd, 1984), also called "payment by results" (Coltham, 1972), bases individual teachers' merit pay bonuses on their students' test score gains. The attractiveness of this strategy is that the evaluation problem is solved by actually measuring certain dimensions of each teacher's output, thereby avoiding the subjective quality of evaluations conducted under old style merit pay, in which bonuses are based on supervisors' evaluations of teachers' performance. There are only a few documented cases of school districts that have used new style merit pay, although merit pay plans that compensate teachers on the basis of student test score gains have recently been supported by several state legislatures.[3] In this section we show that new style merit pay is very much like what economists know as a piece-rate compensation system, and that teaching does not satisfy the conditions under which this type of compensation is efficient.

Approximately 30 percent of U.S. workers in manufacturing are employed under piece-rate contracts, the most common form of payment by results (Pencavel, 1977; Seiler, 1984). Piece-rate contracts work well when the actual contribution of the individual worker to the firm's output can be measured at relatively low cost. Commercial laundries' contracts with workers who iron shirts provide an example. The number of shirts ironed is a relatively accurate measure of the worker's contribution to the firm. Consumer complaints provide a check on quality. Multiple dimensions of output can be managed by providing a schedule of piece rates for different types of clothing.

Piece-rate contracts do sometimes elicit opportunistic behavior. For example, workers may neglect the maintenance of the machines on which they work since they are not rewarded for machine maintenance (see Pencavel, 1977). For many types of work, however, the costs of such opportunism are outweighed by the advantages that piece-rate contracts have over contracts that attempt to control opportunism by monitoring worker actions. In particular, piece-rate contracts provide a strong incentive for workers to find the most rapid way to iron shirts. High productivity results in immediate rewards; a drop in output results in immediate penalty.

Why haven't merit pay plans that compensate teachers on the basis of their output, as measured by student test score gains, become popular? One reason concerns the nature of the incentives that such a compensation system provides. Any explicit list of pay rates for specific levels of student test score gains (economists would refer to such a list as a payment algorithm) creates a specific price—a piece rate—for each student's test score gain in each subject area. For example, an algorithm that bases compensation solely on gains in average reading scores implicitly places a zero price on student gains in other subject areas. Moreover, it places an equal weight on each student's gain. If teacher time is viewed as a private good

[3] See U. S. Department of Education (1984, p. 45) for a reference to legislation that provides state financial support to school districts that adopt new style merit pay plans.

(time spent with one student reduces time available for other students), then this algorithm creates incentives for teachers to allocate time so that the last minute of time spent with any child yields the same expected test score gain. This means that there are incentives for teachers to minimize the time they spend with children whose test scores will not respond to modest increases in attention.

There is limited evidence that teachers do respond to payment by results by allocating their time to specific subject areas and individual children. For example, in the middle of the nineteenth century in England, elementary school teachers worked under a payment-by-results plan that based their compensation on the number of children who acquired a set of narrowly defined skills. This led to a narrowing of the curriculum to exclude all nontested subjects, including many that were perceived to be important—for example, history and geography—but were difficult to test (Coltham, 1972, p. 24).

Other evidence comes from the experiments sponsored by the Office of Economic Opportunity in the early 1970s, in which private firms provided reading instruction to public school children, with the firms' compensation dependent on student test score gains. In at least one of the sites, teachers concentrated their time on children in the middle of the test score distribution, neglecting those at the top who would advance well on their own (test score gains above a threshold were not rewarded), and those at the bottom, whose test scores would not respond to modest additional amounts of teacher time (Gramlich & Koshel, 1975).

Several readers of early drafts of this paper argued that the evidence presented above, in fact, provides support for the usefulness of new style merit pay as a strategy for motivating teachers. They pointed out that the evidence demonstrates that teachers do change their behavior in order to respond to the incentives they face. All that is needed, these readers argue, is to fine-tune the payment algorithm, for example, by giving weight to skill development in more subject areas and perhaps by weighting achievement gains of some children more than those of others. We believe that the proposed technical solution of fine-tuning the payment algorithm neglects two critical problems: the lack of consensus about the appropriate weights, and the nature of teachers' work. We consider each problem in turn.

Most policy debates about public education avoid the divisive topic of weights, which is, at its core, a discussion about whose education, or what sort of education, matters the most. Instead of explicitly debating what the weights should be, it is common in public education to delegate decisions on resource allocation to teachers and administrators, with the inoperable admonition that they provide every student with the opportunity to fulfill his or her potential. Such delegation is not consistent with the design of contracts that pay teachers on the basis of their output, for teachers' different decisions about weights mean that they are each trying to produce a somewhat different mix of outputs.

If the public schools' lack of consensus on weights were the only problem in paying teachers on the basis of their students' progress, one would expect to see more extensive use of such compensation schemes in private schools, where, according to Erickson (1982), family choice leads to greater agreement on school goals. The limited available evidence suggests, however, that performance-based pay for teachers is relatively rare in private schools. In 1983 only 7 percent of Catholic high schools reported that they used any form of merit pay, and none of those

schools based pay differentials on student test score gains (National Catholic Education Association, 1985).

Why aren't teachers paid on the basis of their students' test score gains, even in organizations where there is relatively high consensus on goals, union power is negligible, and management can unilaterally decide how teachers will be compensated? We believe that the answer lies in the nature of the work in schools. Even where there is a high level of consensus on goals, the goals are multidimensional—for example, raise the average reading level in each class, teach all students to embrace democratic values, help each student realize his or her own potential, and eliminate drugs and violence from the school. While it may be reasonable to attribute progress toward certain goals, such as raising reading scores, to individual teachers working behind closed classroom doors, it is not possible to measure each teacher's contribution to the attainment of other school goals. For example, eliminating violence and drugs from a school requires that teachers open their classroom doors and work *as a team* to monitor students' actions outside the classroom. If teachers really do work as a team, it is not possible to measure each teacher's contribution to the group output—in this case, a lower level of drugs and violence in the school (Alchian & Demsetz, 1972). Consequently, individual teachers' contributions to achieving this school goal cannot play a role in determining their compensation under new style merit pay.

If teachers' pay is based solely on success in raising reading scores, there are strong incentives for teachers to keep their classroom doors closed and neglect the teamwork that contributes to the accomplishment of other school goals. Moreover, the strategy used by some firms to combat this form of opportunism—hiring workers to perform the tasks neglected by piece-rate workers (maintaining the machinery, for instance, in our laundry example)—does not work well in schools. Teachers, who work with students every day in class and know students' names and personalities, are likely to be more effective in eliminating drugs and violence from a school than are specialized security officers.

School principals as well as teachers realize that much of the important work in schools must be done by teachers working together—for example, some maintain quiet in halls and libraries while others teach. Compensation algorithms that reward only those dimensions of performance for which each teacher's contribution can be measured could create perverse incentives, inducing teachers to abandon hall and library duty, for instance. This may explain why paying teachers on the basis of their students' test scores is extraordinarily rare in American education.

It is important to note that our discussion of the problems posed by merit pay rests on the nature of teachers' work and the incentives that piece-rate compensation schemes provide. This is quite different from the typical objection to new style merit pay, which emphasizes the inadequacies of standardized tests. While it is true that standardized tests of, say, students' reading skills often do not provide an accurate measure of students' skills, and consequently of the fruits of teachers' work, the inadequacy of tests is not the fundamental problem with new style merit pay. Even if tests were developed that provided accurate measures of students' skills in particular subject areas, incentives to allocate time strategically to particular students and particular subject areas and to neglect aspects of the job not measurable by standardized tests would still remain.

*Old Style Merit Pay*

The significance of teamwork and the presence of school principals who have direct supervisory functions suggest the feasibility of basing teachers' compensation on principals' evaluations. In fact, such old style merit pay is the common model. In this section we explore the extent to which the contracts literature helps us understand why most experiments with old style merit pay have failed.

The lessons from the contracts literature (see Alchian & Demsetz, 1972; Williamson, 1975) regarding the conditions under which it is efficient to base the compensation of individual workers on supervisors' assessments of their performance can be easily summarized. Merit pay is efficient when the nature of the activity in which workers are engaged is such that supervisors can provide relatively convincing answers to these two questions posed by workers:

1. Why does worker X get merit pay and I don't?
2. What can I do to get merit pay?

Unloading boxes from a truck is often suggested as an activity where supervisors can answer workers' questions about performance-based pay differentials. Supervisors can state that worker X was paid more than other workers because he carried two boxes at a time, while other workers carried one at a time. Workers are likely to accept this answer because they recognize that carrying two boxes at a time is, in fact, productive. They also recognize that the nature of the activity gives worker X few possibilities for opportunistic behavior—that is, for actions that make him appear productive but in fact do not contribute to the work at hand. Supervisors can answer workers' second question by stating that they too can earn higher pay by carrying two boxes at a time. Workers are likely to find this answer acceptable because the required action is something they can do if they so choose.

Teachers' work is, by its nature, very different from work such as unloading a truck. As is true for workers in any field, some teachers are more effective than others—hence the call for merit pay. Most analysts agree, however, that effective teaching cannot be characterized as the consistent use of particular well-defined techniques.[3] In other words, there is no analog to carrying two boxes on every trip.

One consequence of the imprecise nature of the activity of teaching—where this expression denotes the loose relationship between particular teacher actions and student learning—is that supervisors cannot answer convincingly when teachers ask why teacher X received merit pay and they did not. As one of the administrators we interviewed commented: "I know who the good teachers are. They're so and so, so and so, and so and so. Why are they good teachers? Well, I don't know, they are just good teachers; but I know who they are." Many teachers who are denied merit pay find this answer unsatisfactory. One reason is that they are aware that the nature of teaching, with its closed classroom doors and its network of relationships among teachers and between teachers and parents, provides great potential for opportunistic behavior. In other words, there are many things that a teacher could do to impress a principal and to suggest that he or she was more

[3] See Wise et al. (1984, p. 10) for a discussion of the claims and refutations concerning the role of specific teacher actions in fostering student learning.

effective than his or her colleagues. Examples might include using friendships with parents to spread rumors about other teachers' incompetence, and refusing to share materials that could help other teachers. Thus, teachers have reason to question whether merit pay is awarded to teachers who are in fact the most productive or to those who are most facile in impressing supervisors.

A second consequence of the imprecise nature of teaching is that supervisors cannot answer convincingly the teacher's second question, What can I do to earn merit pay? In other words, supervisors cannot suggest specific actions that the teacher can undertake which both teacher and supervisor recognize will enhance the teacher's effectiveness. Without an unequivocal answer to this second question, teachers may have little incentive to change their behavior in pursuit of higher income. What is worse, teachers may learn that concealing their problems and playing up to evaluators is what the organization rewards—dramatically complicating managers' evaluation problem.

In effect, the lesson from the contracts literature is that the problems with old style merit pay are more fundamental than careless implementation or inadequate training of evaluators—to name but two of the explanations often given for the failure of merit pay plans. The problem lies in the nature of the teaching activity itself. Specifically, it is the lack of a blueprint for effective teaching that prevents supervisors from providing convincing answers to teachers' two primary questions about merit pay.

What the contracts literature does not reveal is exactly what problems arise under merit pay that have led most school districts to drop this type of compensation system after a brief trial. While our research was not designed to address this question, we did learn some interesting facts about the problems caused by merit pay, particularly from teachers and administrators in two districts that have had merit pay for more than twenty years and have altered their plans several times to deal with perceived problems. The comments of these participants are informative in understanding what happens when supervisors cannot answer teachers' questions about why some teachers receive merit pay and others do not.

One theme that ran through our interviews was a perception on the part of administrators that merit pay could easily backfire, since teachers who received evaluation ratings lower than they felt were fair might respond by working less hard. This theme is exemplified in the story one former principal told about a fine teacher whose work he rated "excellent." Unfortunately, excellent was the second highest rating in the system, and the teacher firmly believed she deserved the top rating, "outstanding." She responded to the principal with, "If that's all you care, then that's all you'll get," and, indeed, he reported that her work "fell off."

Another theme we discerned from our field notes was that past experiences conditioned teachers' expectations about their evaluations. No teacher expected to be given a rating lower than the one he or she had received in the previous rating period. Rarely do discussions of merit pay focus on the *repetitive* nature of the evaluation process. Yet teachers see their "merit" ratings in terms of what they and others have been told by supervisors in the past. Being demoted is difficult for anyone, but it would be particularly hard if supervisors could not pinpoint what was wrong and explain how the situation could be remedied.

Several administrators cited negative consequences that arose from giving a teacher a lower rating than the teacher had received in the past. One principal mentioned a teacher with no better than adequate performance to whom he gave a rating lower than the rating given by his predecessor. The teacher, who had planned to retire at the end of the year, was so infuriated by this rating that she postponed her retirement for two years. The principal was left with a teacher he did not want; moreover, the teacher had become angry and recalcitrant as a result of her evaluation rating. The key point here is that even an evaluation system that produces valid and reliable performance ratings is not enough to guarantee the success of merit pay. If teachers feel that the ratings are unjust, and evaluators cannot convince them to the contrary, their reactions to the ratings may undermine the education students receive.

A third theme in our interviews was that merit pay tended to interfere with school principals' efforts to build effective instructional teams in their schools. Several school principals commented that, prior to the introduction of merit pay, they often gave teachers ratings higher than they actually deserved and then encouraged them to live up to the high ratings. The principals reported that this was the most effective strategy for stimulating many teachers to improve their performance, because it built teachers' confidence and established trust in the principals. While this evaluation strategy produced ratings that were not objectively valid, the principals felt it promoted teacher morale and better teaching performance. Principals felt that this approach allowed them to focus on the specific problems a teacher was struggling with, whereas more objective evaluations produced an adversarial atmosphere and could create incentives for teachers to conceal problems.

Many of these same principals worried that the use of merit pay would restrict their ability to pursue the strategy of encouraging teachers through the use of high ratings. One reason they worried was that the school district administration was pressing principals to be objective in their ratings and to standardize ratings across schools. A second cause for concern had to do with school board complaints of an excessive number of high ratings; administrators were being pressed to lower ratings and to provide a stronger defense of the top ratings they did give. At the same time, principals felt pressure from teachers to explain why they had not been given the top rating while the teacher in the next classroom had. As one principal stated, "Merit pay turns my job from being a coach into being a referee." He further implied that his teachers no longer saw him as a helpful coach but as a critical referee —and this threatened his ability to motivate the teachers to higher levels of effort.

Some readers may conclude that one of the *benefits* of merit pay is that it pressures principals into actually evaluating teachers objectively, one of the most important parts of their job. There is some truth to this argument; certainly, many of the principals we interviewed felt this pressure. There is more to be said, however. The principal's primary job is to ensure that the children who pass through his or her school learn as much as possible. Yet the principal doesn't teach students; teachers do that work. The principal's success, therefore, depends to a large extent on his or her success in encouraging teachers to work hard and work together. When a principal gives a lower evaluation to a teacher than he or she had previously received, the teacher may lose some money, but the principal may lose

the cooperation needed to make the school work. Our field notes contain many stories from principals describing the distressing consequences of giving lower ratings than the teachers expected.

Many principals saw merit pay as making their job more difficult by increasing both the tensions surrounding the formal evaluation process and the intensity with which teachers asked why they did not get the top rating and what they could do to receive a better rating—questions that principals could not answer convincingly. In fact, the general thrust of the principals' comments, with a few notable exceptions, is quite consistent with the survey evidence indicating that low morale and "problems of administration" are the primary reasons school districts drop merit pay (see Calhoun & Protheroe, 1983).

Our evidence leads us to emphasize the importance of the imprecise nature of teachers' work as a factor contributing to the demise of old style merit pay. One of the readers of an early draft of this paper commented that, if we were correct and the problem were not simply poor public sector management, we should expect that old style merit pay would not be common in for-profit educational institutions. While an in-depth exploration of this proposition was beyond our resources, we did attempt to respond to this comment by learning about the compensation policies of the Stanley H. Kaplan Educational Center, a large, nationwide for-profit firm specializing in preparing students to take standardized tests such as the SAT.

Stanley Kaplan does monitor the performance of its teachers closely, in part by observing them in the classroom and, to an even greater extent, by soliciting student evaluations of each teacher's performance. In fact, Kaplan's students are quick to complain when the quality of instruction does not justify the cost of the course. Kaplan uses the feedback from students in deciding which teachers to dismiss, but does not use this information in determining individual teacher's compensation. In fact, teachers who work for Kaplan are paid in much the same manner that public school teachers are paid. All teachers are paid according to a salary scale that bases compensation on experience, that is, on the number of courses taught. There are no bonuses for superior performance.

We asked the personnel director of Kaplan why the firm does not use performance-based pay. Her answer included these points: all Stanley Kaplan teachers are effective; those who are not are dismissed. There are some teachers who are superstars, and the firm has considered paying bonuses to them. This plan was rejected because of management's perception that the positive impact of bonuses on the performances of the superstars would be more than offset by negative effects on the performances of effective teachers who do not receive bonuses, do not know why they were passed over, and cannot be told how to become superstars.

In the context of this paper, the Kaplan evidence can be interpreted as implying that even when management feels it can make relatively accurate, fine-tuned distinctions among teachers, it would not be able to convince the merely good teachers of the superior performance of some of their coworkers. As a result, the responses to the pay differentials would not further the goals of the organization. Thus, the imprecise nature of teaching prohibits evaluators from answering the hard questions teachers pose about old style merit pay and leads a successful profit-making firm to base compensation on experience. It is important to add that

Stanley Kaplan uses evaluation aggressively, even without merit pay, both to dismiss ineffective teachers and to offer useful advice to effective teachers. This approach is obviously similar to that taken by administrators in many public schools.

## Why Some Merit Pay Plans Survive

If merit pay is not an effective strategy for improving teachers' performance, why do merit pay plans survive in a few districts? Are the districts atypical? Are the provisions of the merit pay plans atypical? Did merit pay in these districts help to solve problems other than that of motivating teachers?

We began our search for the answers to these questions by identifying school districts that have used merit pay for a number of years. Two Educational Research Service publications were helpful in this regard. The first (Porwoll, 1979) identified 115 school districts in the United States that used merit pay in 1978. The second (Calhoun & Protheroe, 1983) reported the results of a survey that inquired whether each of these 115 districts was still using merit pay in 1983, and if not, why not. The 47 districts that reported in the 1983 survey that they were still using merit pay formed the population from which we selected districts for study.

Within the population we looked first for urban districts with ongoing merit pay plans. Since many urban districts are thought to have particularly serious problems with poor teaching quality and low teacher morale, an analysis of enduring merit pay plans in such districts might provide important insights into the factors that contribute to the success of performance-based contracts. However, we found no urban districts with long-lived merit pay plans. In fact, we could not find even one documented case of a large, once-troubled school district that had successfully used merit pay to improve its performance. On the contrary, one of the striking aspects of the list of districts with enduring merit pay was the large percentage of very small districts serving relatively homogeneous student populations. Moreover, these districts tended to use very small amounts of money as merit pay bonuses.

We then looked for districts that had used merit pay for at least five years and had either used pay differentials of at least $1,000 or served more than 10,000 students. We found seven districts that met these criteria. We spent several days in six of these districts interviewing teachers and administrators with the goal of learning how each merit pay plan worked and how teachers and administrators reacted to these plans.

### *Characteristics of the Six Districts*

The six districts we studied vary in size from 2,500 to 60,000 students. Three are located in the Southwest, one in the Northeast, one in the Mid-Atlantic region, and one in the North Central region of the country. Two districts have collective bargaining; the union role in the other four is insignificant.

Part of the reason merit pay plans persist in the six districts has to do with their unusual working conditions. All of the six districts are considered to be among the best in their geographical areas—places where teachers like to work and where high housing prices reflect, in part, the desirability of the public schools. In eval-

uating the role merit pay plays in contributing to these districts' accomplishments, it is important to focus first on attributes other than merit pay that these districts have in common.

All of the districts have salary schedules, to which merit pay is added, that are above average for their geographical areas. The high salaries and good working conditions permit these districts to be selective in choosing applicants for teaching positions. None of these districts adopted merit pay as a response to the idea that there was not enough money to pay all teachers well so they would at least pay a few good teachers well. In fact, several administrators made comments such as, "No merit pay system would ever work without salaries at a point that teachers can live on."

None of these districts use merit pay as a strategy to give negative signals to teachers perceived to be ineffective. However, using evaluation practices that are in principle unrelated to merit pay, they do dismiss teachers judged to be incompetent and are pressured by parents to do so. These practices have not been resisted by teachers' unions in the two districts with relatively powerful unions. The union leaders in these districts stated that they made sure due process was observed but that it was not in the union's interest to protect incompetent teachers. One lesson to be learned from examining the characteristics of school districts with long-lived merit pay plans is that attractive working conditions may be a prerequisite for the survival of merit pay.

*Characteristics of the Enduring Merit Pay Plans*

Working conditions do not provide the whole answer to why merit pay survives in a few districts. In fact, merit pay has been dropped by a great many districts that appear similar to the six we studied. Thus, to explain the survival of merit pay in our districts we must look at the plans themselves.

The six merit pay plans that we analyzed differ in many respects. However, in every case the plan incorporates a strategy for dealing with the two questions, already noted, that many teachers ask about merit pay. The strategies consist of varying combinations of four themes: extra pay for extra work, making everyone feel special, making the program inconspicuous, and legitimation through participation. These strategies represent adaptations of the merit pay idea that eliminated those conflicts between merit pay and the nature of teachers' work that we discussed above. However, they turn merit pay into something else. In fact, we regard each of these adaptations as evidence supporting the theme developed in the first part of this essay, namely, that teachers' work does not satisfy the conditions under which performance-based compensation is an effective means of motivating workers to high performance levels. In the analysis that follows, we stress the ways in which each district changed one or more crucial aspects of the merit pay idea. While these districts still refer to their plans as merit pay, economists would not view them as examples of performance-based compensation.

*Extra pay for extra work.* One common element in the long-lived merit pay plans is that the definition of performance is altered so as to reduce emphasis on classroom teaching and increase emphasis on completion of tasks outside the classroom. For example, the numerical rating system used by one district to determine merit pay awards gives school and community service the same weight as class-

room performance. Another district requires that a teacher complete six outside activities to be eligible for merit pay. As one teacher commented, "This isn't merit pay; it's how you get the yearbook done."

A complementary practice is to make the *teacher* responsible for documenting that he or she is worthy of merit pay. As part of the merit pay application process in several districts, teachers had to prepare lengthy documents describing their accomplishments and providing evidence in the form of testimonials from colleagues and parents. One teacher commented, "When I finished this last time, I had a volume no less than three inches thick of evidence, arguments, and materials."

These practices, which we call extra pay for extra work, provide one set of relatively convincing answers to the two questions teachers raise about merit pay. Administrators can clearly state that teacher X received merit pay because he or she devoted time to organizing a variety of activities and to documenting his or her accomplishments, both in and out of the classroom. If another teacher wants merit pay, he or she can do these same things.

This approach to merit pay relieves administrators of the impossible task of discerning and defending differences in the quality of teachers' classroom instruction. But this approach also means that the school districts using it have effectively given up any effort to relate financial rewards to the core of the teacher's job—namely, classroom instruction. The use of this approach underscores our earlier argument that merit pay is ill-suited to teaching.

*Make everyone feel special.* A second theme is to quietly award merit pay to almost all teachers. This strategy is most pronounced in one district in which a numerical rating system is used to determine whether teachers receive no award or an award of $500, $1,000, $1,500, or $2,000. Only teachers who had worked in this district for at least six years were eligible for merit pay. Eligible teachers could choose not to participate in the merit pay plan, and this choice spared them from documenting their accomplishments and undergoing the merit pay review process. Only a very few teachers chose not to participate in this merit pay plan, however.

Teachers whom we interviewed in this district were unaware of the distribution of actual awards but typically were pleased that they each received a substantial award. In fact, every teacher who participated in the voluntary merit pay program (over 90 percent of eligible teachers in the district) received an award; 85 percent of the teachers received either $1,500 or $2,000. We suspect that the bunching of the ratings at the top of the scale and the relatively small monetary differential between the top two awards is important in minimizing ill feeling on the part of teachers in schools headed by hard-grading principals. In this district, if a principal is a hard grader, the effective teacher receives an annual bonus of $1,500 instead of $2,000.

In effect, the "make everyone feel special" strategy deals with teachers' potentially destructive questions about merit pay by reducing the number of teachers who ask. We find it interesting that this theme was particularly evident in the two districts in our sample that have had merit pay for more than twenty years. But if this approach deals effectively with teachers' questions, it does so by rewarding everyone, cutting off questions by cutting off the reasons for asking them. One idea behind merit pay, however, is to use differential financial rewards to improve worker performance. If most teachers receive the top reward or an amount close

to it, then there will be little difference in the incentives and thus little chance that the differences will affect teacher performance. Once again, this approach to implementing merit pay provides additional evidence of the poor fit between this type of compensation scheme and teachers' work.

*Make merit pay inconspicuous.* In several districts, the design of the merit pay system is such that the incentives are of little interest to a large percentage of the teachers. For example, in one district, eligibility for merit pay requires ten years of service, completion of six activities outside the classroom, and satisfactory performance evaluations. The reward for fulfilling these requirements is $600 (somewhat more if coupled with advanced degrees). Only 40 percent of the teachers in this district who do fulfill the length of service requirement choose to participate in the voluntary merit pay plan. In another district, teachers can apply for one of four different award levels, each level having different requirements. While the award levels are sizable—$1,000 for level one, $4,000 for level four—the requirements are so demanding that only 12 percent of the teachers apply for any level (two-thirds of these teachers receive awards). The level four requirements include a master's degree and thirty hours of graduate credits, superior teaching skills as demonstrated, for example, by "representing the district at the state or national level as a resource person, chairperson, or committee member," and superior professional contributions, as demonstrated, for example, by serving "in an official capacity in the management of the professional associations or organizations related to a specific field of study."[4] For the vast majority of the teachers in this school district, the financial awards do not justify the extra work.

In all six districts, merit pay has a low profile. In part, this stems from the perception that merit pay is something almost any teacher could earn but that the financial rewards do not justify the extra work. Another element is that, in five of the six districts, teachers are urged not to discuss with colleagues either who receives merit pay or the amount of the awards. In these districts, where most teachers like their jobs, the primary effect of secrecy seems to be to reduce teachers' interest in merit pay and thereby to reduce the number of teachers who ask the hard questions about why some teachers get merit pay and others do not. But whatever the reasons for the low profile, this common approach to implementation is further evidence that our districts tended to turn merit pay into something else. If the aim of a differential compensation system is to stimulate better performance, then it would be important for workers to know who did the better job, and why. But the districts we visited built barriers against the acquisition of this information.

*Legitimation through participation.* The final attribute of merit pay uncovered in our districts concerns the process by which the programs were designed. In all of the districts, teachers played a significant role in the design of the merit pay plans. Moreover, in each of the two districts that have had merit pay for more than twenty years, the system has been revised several times in response to teacher complaints. We believe that teachers' participation in the design and redesign of the plans contributes to the plans' longevity. One reason for this effect is that the

[4] These quotations are taken from the school district's description of its performance-based compensation plan.

process of participation reveals information about teachers' preferences that is extremely difficult to collect unless teachers volunteer it. This information, moreover, is critical in enabling supervisors to predict teachers' responses to incentives. Participation gives teachers a reason to volunteer information and a mechanism for doing so. A second reason that participation contributes to the longevity of merit pay plans is that it creates the impression that merit pay is not a system thrust upon teachers but rather one they helped to create. Seen as such, teachers may still ask why some teachers get merit pay and others do not, but the intensity with which they ask is diminished. Teachers recognize that if many of them find the program objectionable, they can change it.[5]

One could view teachers' participation as an example of workers' self-protective behavior. In this case, however, workers' objections to merit pay were shared by management. We therefore view worker participation as yet another effort by management and workers to redefine a compensation system that has great potential for doing damage.

## Is There a Role for Merit Pay?

The schemes we studied, then, were not merit pay, at least if this term denotes performance-based compensation. But the six districts nonetheless seemed quite convinced that their compensation plans were useful. We wondered why. Did these plans have some desirable effect on the scholastic performance of students in these districts? Did the plans help the districts to resolve problems that they faced?

On the first question, merit pay in the districts we studied does not appear to have affected the quality of teaching; neither administrators nor teachers offered any evidence that merit pay had a significant impact on the way teachers teach. This is not surprising given the attributes of the enduring plans. This conclusion is also compatible with the theme developed above: the nature of teachers' work is such that basing individual teachers' pay on assessments of their performance is unlikely to motivate teachers to work harder.

If merit pay does not motivate teachers to work harder, why do a very few districts retain it? Our interviews with teachers and administrators suggest that merit pay has helped these six districts solve problems quite different from the problem of motivating teachers. These problems include (1) how to support good teachers who differ in their relative needs for income and free time, (2) how to encourage meaningful dialogue between teachers and administrators about difficult issues, such as the quality of the evaluation process, (3) and how to build community support for the public schools.

In the districts we visited, merit pay contributes to solutions to these problems in the following ways:

1. Extra pay for extra work provides opportunities for teachers with greater financial needs to augment their incomes by spending time on school-related activities.

[5] The importance of voice as a mechanism for improving the performance of organizations is elegantly developed by Hirschman (1970).

2. The ongoing discussions of how merit pay works, what its problems are, and what changes are needed provide forums for meaningful dialogue between teachers and administrators concerning difficult issues, such as the nature of the evaluation process.
3. The merit pay plans contribute to the perception that teachers are accountable. As one teacher commented on why members of the community supported merit pay for teachers, "The people out there who are paying taxes want to make sure that in the area of teacher pay, those who are doing the real work are the ones who get the rewards, above and beyond the standard." This perception of accountability increased the willingness of communities to pay teachers well. One administrator remarked that merit pay "has meant a lot of money for a lot of teachers that would otherwise not have been provided, knowing the Board of Education."

We do not mean to imply that merit pay is necessary to solve the problems of satisfying teachers' varied needs, encouraging dialogue, and promoting community support for the schools. In fact, a far greater number of school districts solve these problems without merit pay. For example, many districts meet some teachers' needs for additional income through extra pay for extra work without calling this merit pay, or through small grants competitions.[6] Many districts use the collective bargaining process to promote meaningful dialogue between teachers and administrators (see Freeman & Medoff, 1984). Others promote community support through volunteer programs, public/private partnerships, and outreach activities.

What we do want to suggest is a different way of looking at merit pay. This is useful because, if the past is any guide to the future, the current, perhaps waning wave of interest in merit pay will not be the last time that educators feel pressure to adopt this type of compensation plan. In thinking about merit pay in the future, it is useful to ask whether it can play a modest role in satisfying teachers' varied needs, encouraging meaningful dialogue between teachers and administrators, and promoting community support for the schools. In most school districts, the answer to this question will be an emphatic "no." But in a few districts the answer could be a tentative "yes." We hope that our work provides some clues to the types of districts where an answer of "maybe" makes sense and what types of merit pay plans hold some promise.

There is one final theme in our evidence that seems to apply to districts both with and without merit pay: improving teachers' performance through evaluation. If evaluation is to contribute to the goal of helping teachers improve, it must be carried out by skilled and knowledgeable supervisors in an atmosphere that rewards honesty and cooperation. When teachers who conceal their failings receive higher pay than those who do not, the atmosphere for useful evaluation and advice is poisoned. If supervisors are to engage in a productive dialogue with teachers, they must act in a way that is consistent with the sustained nature of their relationship with teachers. Evaluation is a repetitive sequence which creates expectations, memories, and sensitivities that can either contribute to improved performance or, if treated insensitively, undermine it. It was the goal of merit pay's advocates to put the power of money into the evaluation process as a way to improve teach-

[6] See, for example, *Small Grants for Teachers* (n.d.).

ers' performance. That goal is misguided. But the broader lesson—that school administrators must work to create relationships with teachers in which evaluations contribute to improvement, change, and cooperative problem solving—is one that must not be forgotten even after the pressures for merit pay dissipate.

## References


Alchian, A. A., & Demsetz, H. (1972). Production, information costs, and economic organization. *American Economic Review, 62,* 777-795.

Bacharach, S. B., Lipsky, D. B., & Shedd, J. B. (1984). *Paying for better teaching.* Ithaca, NY: Organizational Analysis and Practice.

Calhoun, F. S., & Protheroe, N. J. (1983) *Merit pay plans for teachers: Status and descriptions.* Arlington, VA: Educational Research Service.

Coltham, J. B. (1972). Educational accountability: An English experiment and its outcome. *School Review, 81,* 15-34.

Erickson, D. A. (1982). Disturbing evidence about the "one best system." In R. B. Everhart (Ed.), *The public school monopoly* (pp. 393-422). Cambridge, MA: Ballinger.

Evendon, E.S (1918). *Teachers' salaries and salary schedules in the United States, 1918-19.* Washington, DC: National Education Association.

Freeman, R. B., & Medoff, J. L. (1984). *What do unions do?* New York: Basic Books.

Gramlich, E., & Koshel, P. (1975). *Educational performance contracting.* Washington, DC: Brookings Institution.

Hanushek, E. (1981). Throwing money at schools. *Journal of Policy Analysis and Management, 1,* 19-41.

Hirschman, A. O. (1970). *Exit, voice, and loyalty.* Cambridge: Harvard University Press.

Johnson, S. M. (1984). Merit pay for teachers: A poor prescription for reform. *Harvard Educational Review, 54,* 175-185.

National Catholic Education Association. (1985). *The Catholic high school: A national portrait.* Washington, DC: Author.

National Education Association. (1923). *Report of the salary committee, teachers' salaries and salary trends in 1923.* Washington, DC: Author.

National Education Association. (1928). Practices affecting teacher personnel. *Research Bulletin, 6.* Washington, DC: Author.

Pencavel, J. H. (1977). Work effort, on-the-job screening, and alternative methods of remuneration. In R. Ehrenberg (Ed.), *Research in labor economics* (Vol. 1, pp. 225-258). Greenwich, CT: JAI Press.

Porwoll, P. J. (1979). *Merit pay for teachers.* Arlington, VA: Educational Research Service.

Seiler, E. (1984). Piece rate vs. time rate: The effect of incentives on earnings. *Review of Economics and Statistics, 66,* 363-375.

*Small grants for teachers.* (n.d.). Pittsburgh: Allegheny Conference on Community Development.

U.S. Department of Education. (1984). *The nation responds: Recent efforts to improve education.* Washington, DC: Author.

Williamson, O. E. (1975). *Markets and hierarchies: Analysis and antitrust implications.* New York: Free Press.

Wise, A. E., Darling-Hammond, L., McLaughlin, M. W., & Bernstein, H. T. (1984). *Teacher evaluation: A study of effective practices.* Santa Monica: Rand Corporation.

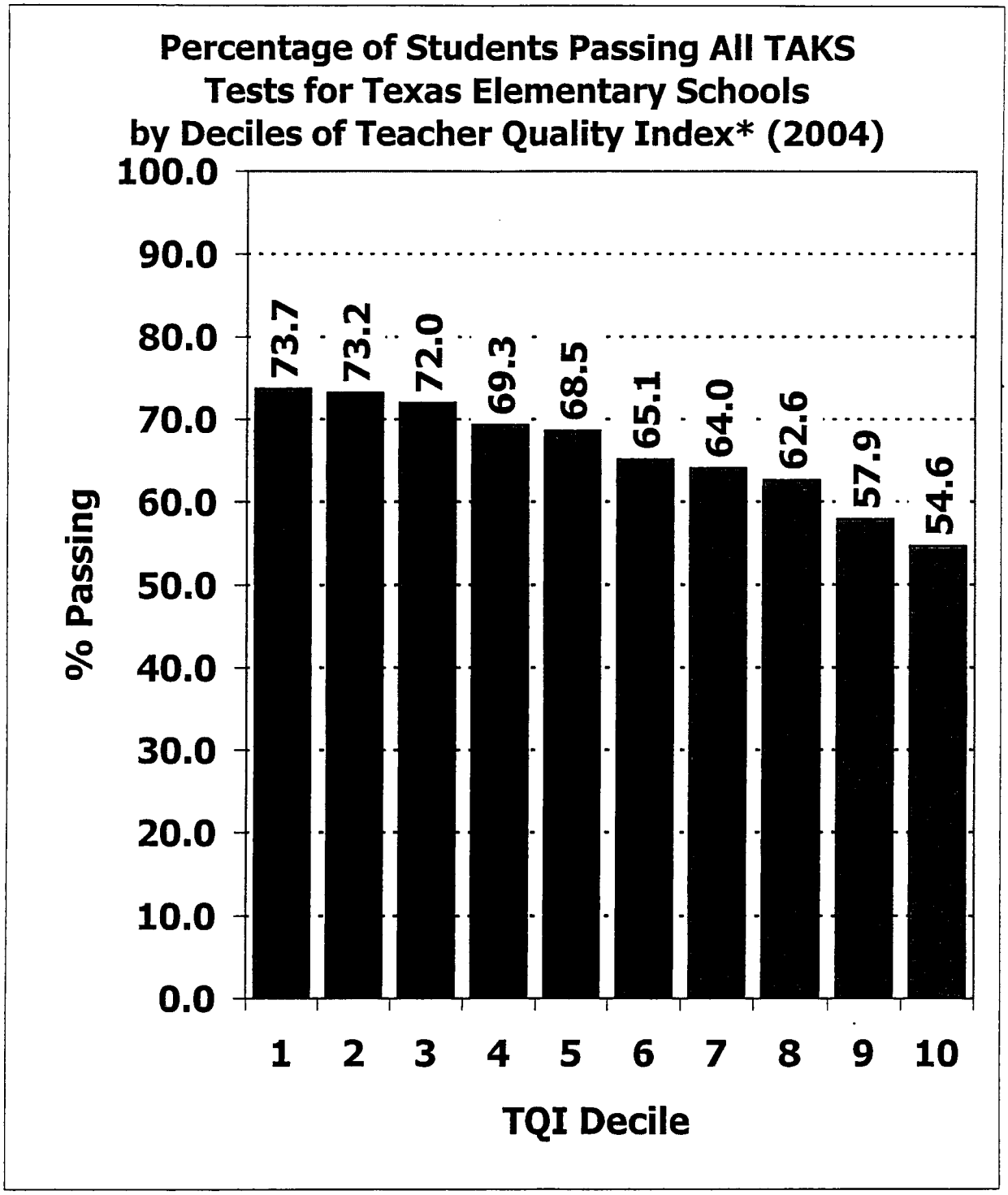

The research on which this paper is based was supported by the Institute for Finance and Governance, School of Education, Stanford University. We would like to thank Katherine Jamentz, Patrick Murnane, and Niall Nelson for able research assistance; and Chris Argyris, Estelle James, Susan Johnson, Charles Lindblom, Giandomenico Majone, Richard Nelson, Robert Pollak, David Stern, Dvora Yanow, and participants in workshops at Boston University, Harvard University, and the University of Pennsylvania for helpful comments on earlier drafts. We would especially like to thank Edward Pauly, who read several drafts of the paper and provided many important ideas.

# Distribution of Teacher Quality in Texas Public Schools (2004)

Presented to the
Senate Education Committee

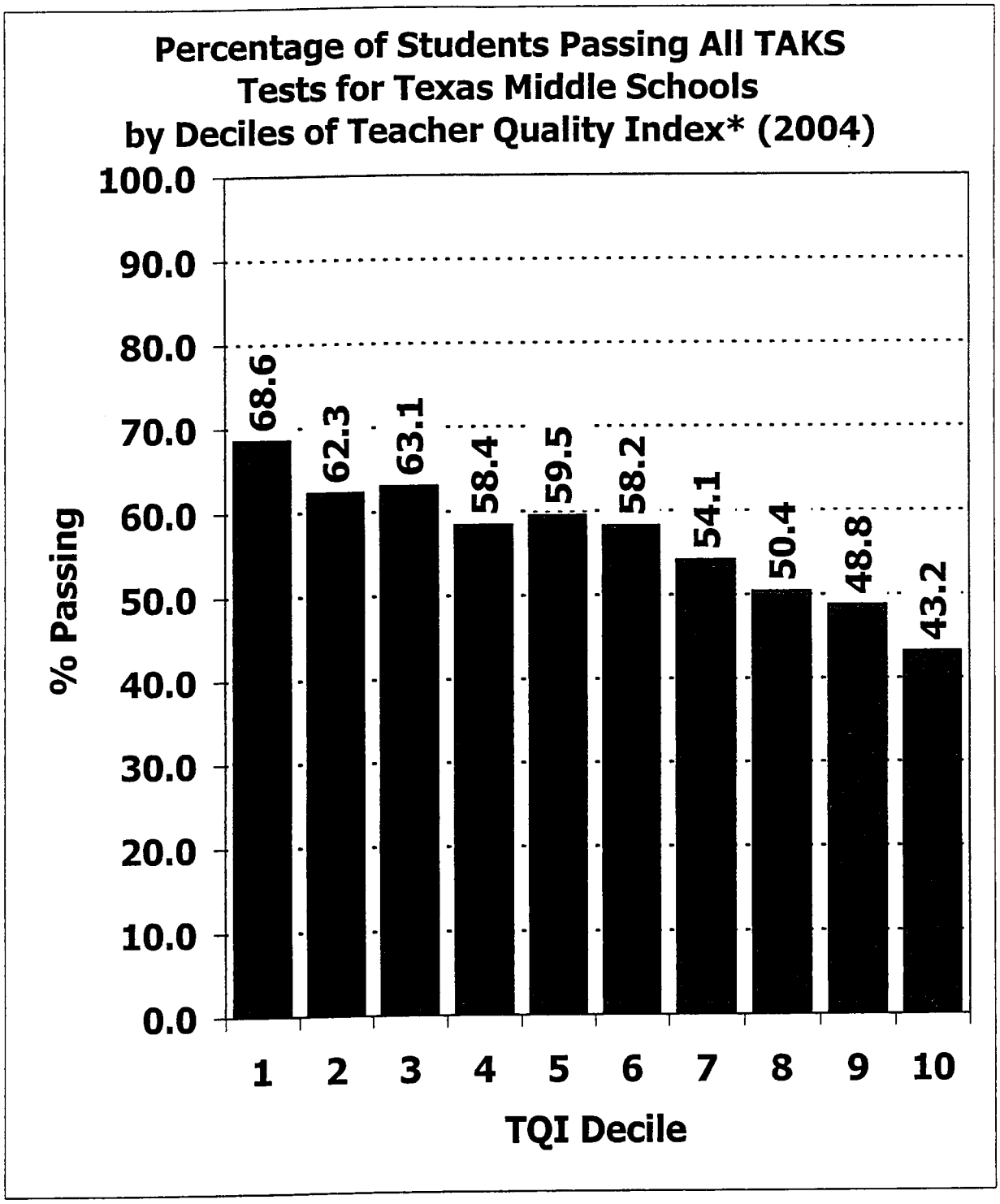
March 31, 2005

Dr. Ed Fuller
Department of Educational Administration
The University of Texas



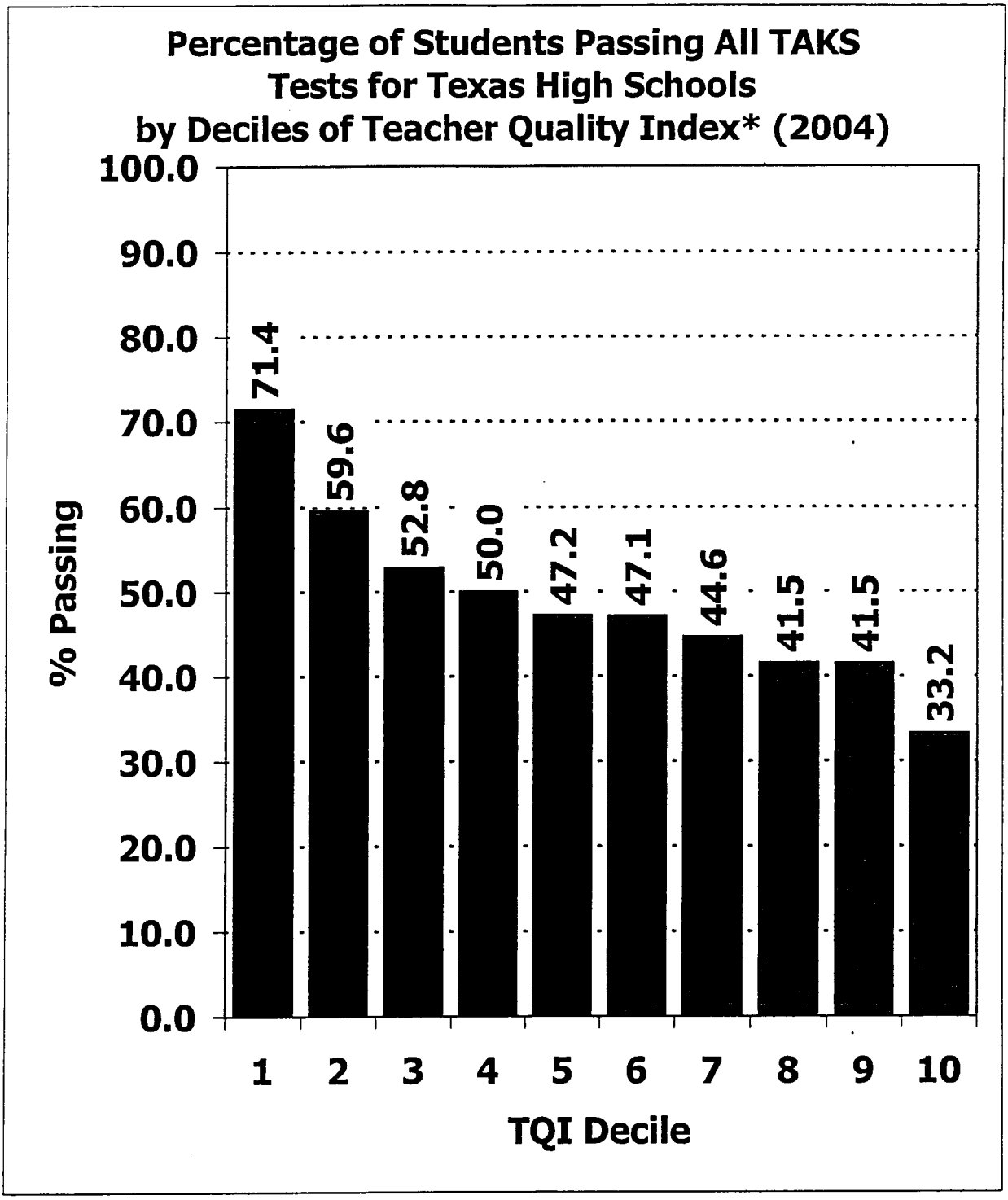

The Teacher Quality Index (TQI) is based on three components of teacher quality at the school level: the percentage of self-contained, English, mathematics, science, and social studies teachers assigned out-of-field, the percentage of novice teachers (teachers with less than 3 years of experience), and the teacher turnover rate. All data are from the 2003-04 academic year, with the exception of the turnover rate which measures the percentage of FTEs leaving a school from 2003-03 to 2003-04.

SOURCE:




The Teacher Quality Index (TQI) is based on three components of teacher quality at the school level: the percentage of self-contained, English, mathematics, science, and social studies teachers assigned out-of-field, the percentage of novice teachers (teachers with less than 3 years of experience), and the teacher turnover rate. All data are from the 2003-04 academic year, with the exception of the turnover rate which measures the percentage of FTEs leaving a school from 2003-03 to 2003-04.

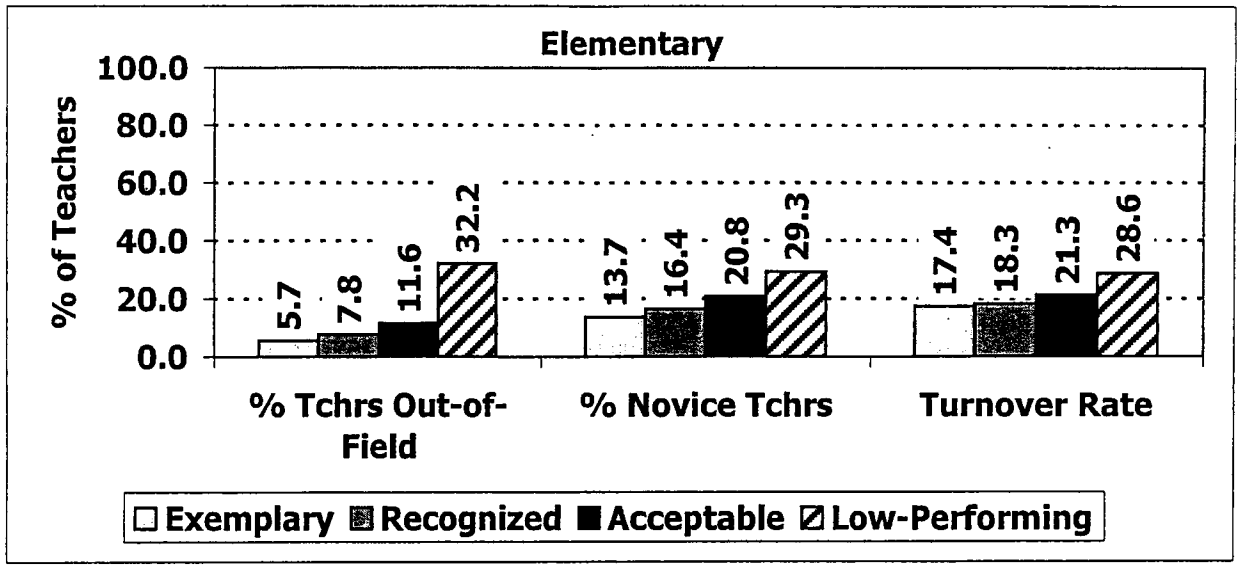
SOURCE:

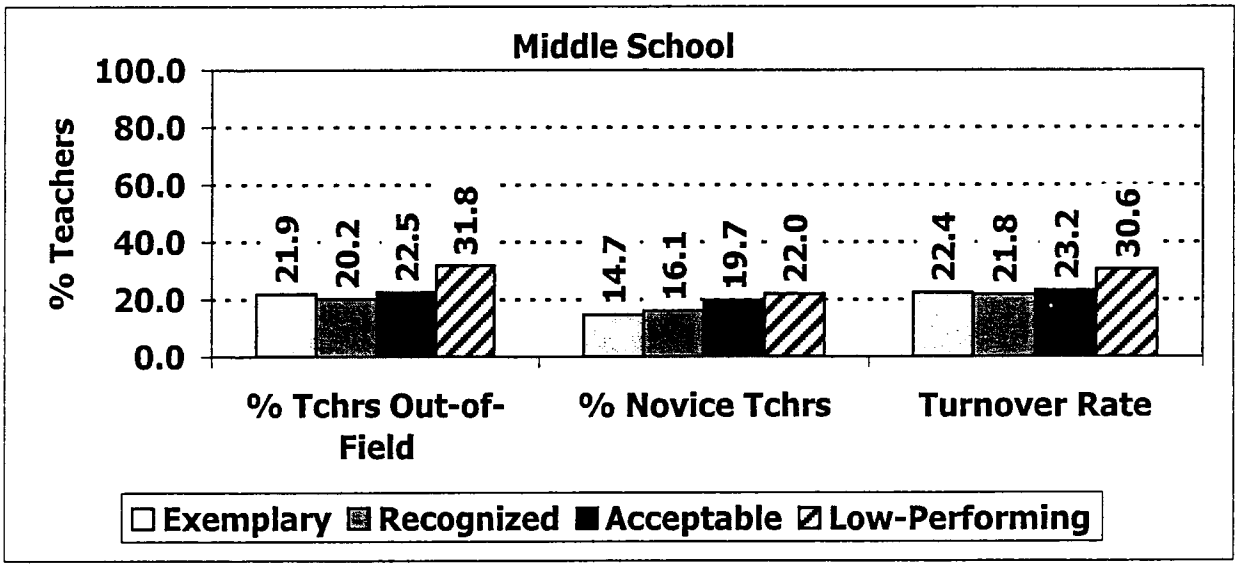


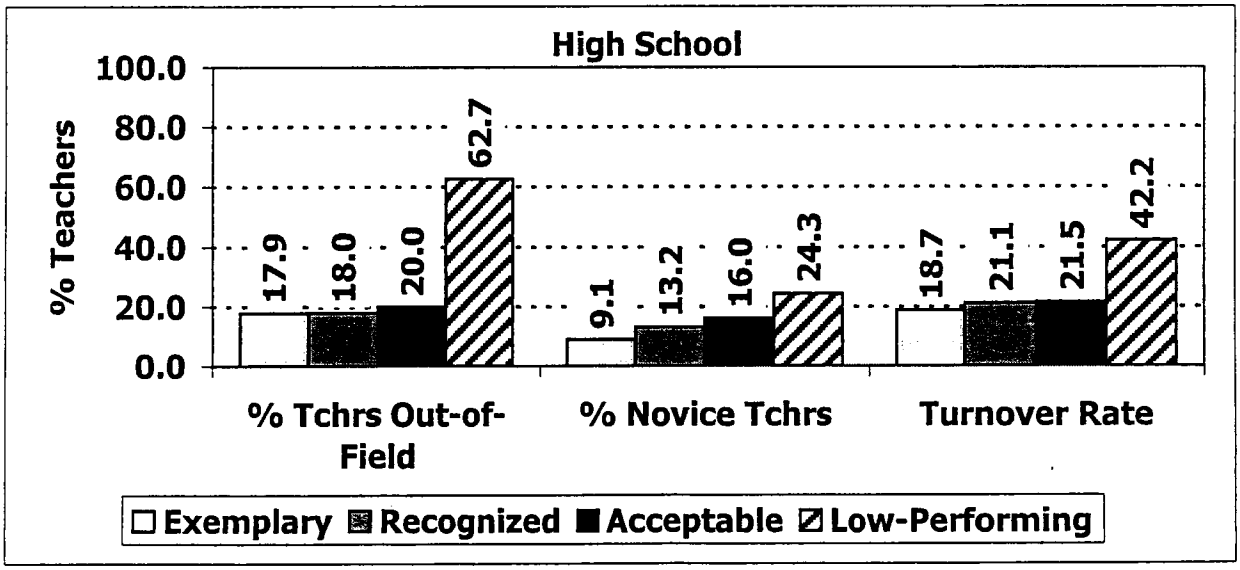

The Teacher Quality Index (TQI) is based on three components of teacher quality at the school level: the percentage of self-contained, English, mathematics, science, and social studies teachers assigned out-of-field, the percentage of novice teachers (teachers with less than 3 years of experience), and the teacher turnover rate. All data are from the 2003-04 academic year, with the exception of the turnover rate which measures the percentage of FTEs leaving a school from 2003-03 to 2003-04.

SOURCE:

## Teacher Quality Indicators for Texas Schools by Accountability Rating (2004)



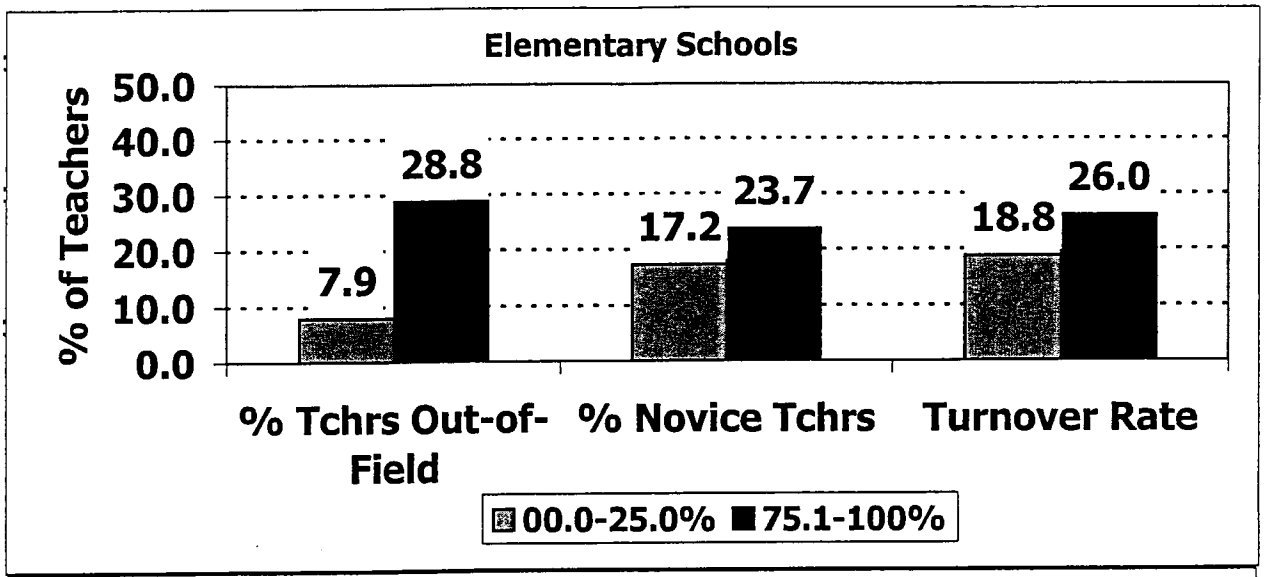



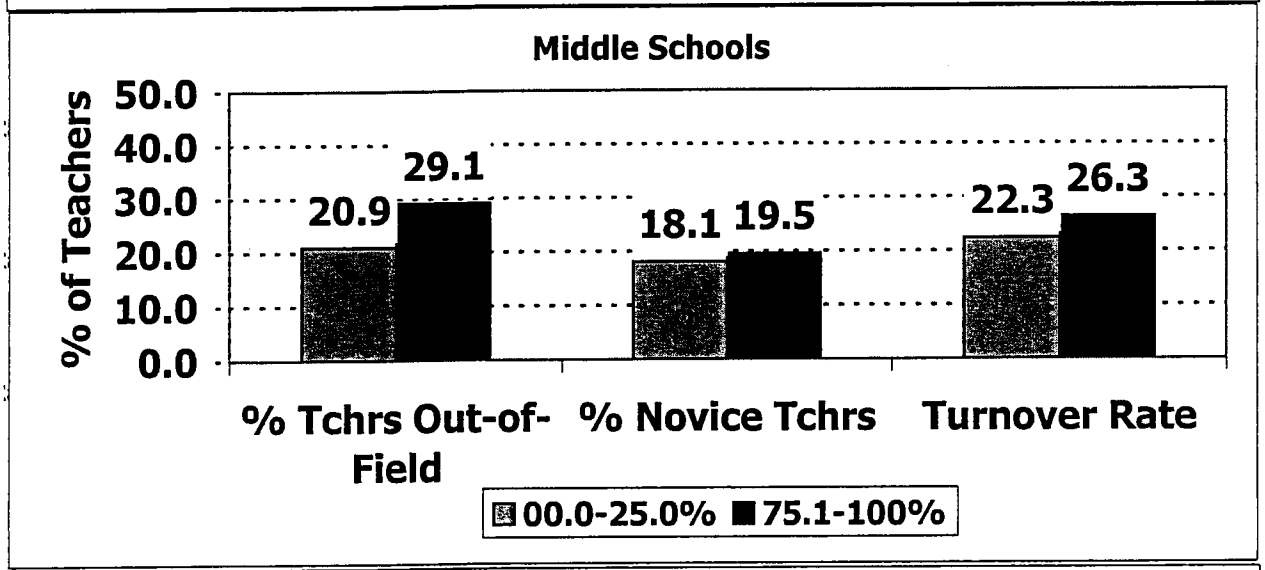



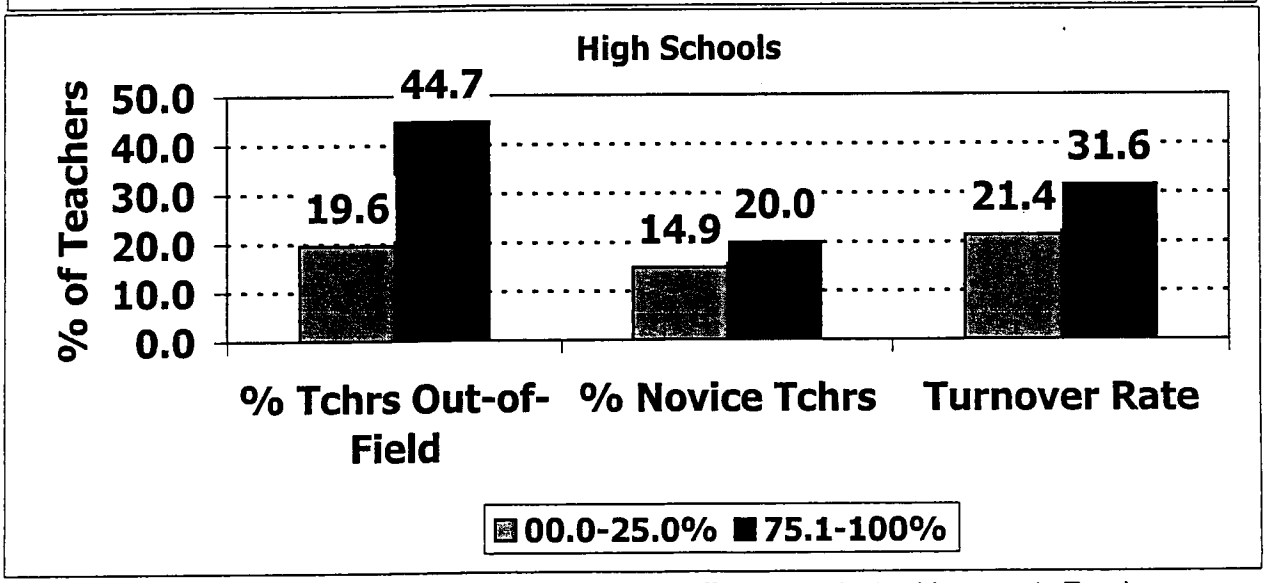

All three measures have been found to have a significant effect on student achievement. Teachers assigned out-of-field, novice teachers, and teacher turnover all negatively affect student achievement. All three measures are based on teacher ful-time equivalents (FTEs) as found in PEIMS.

SOURCE:

# Teacher Quality Indicators for Schools Serving High and Low Percentages of African American Students (2004)



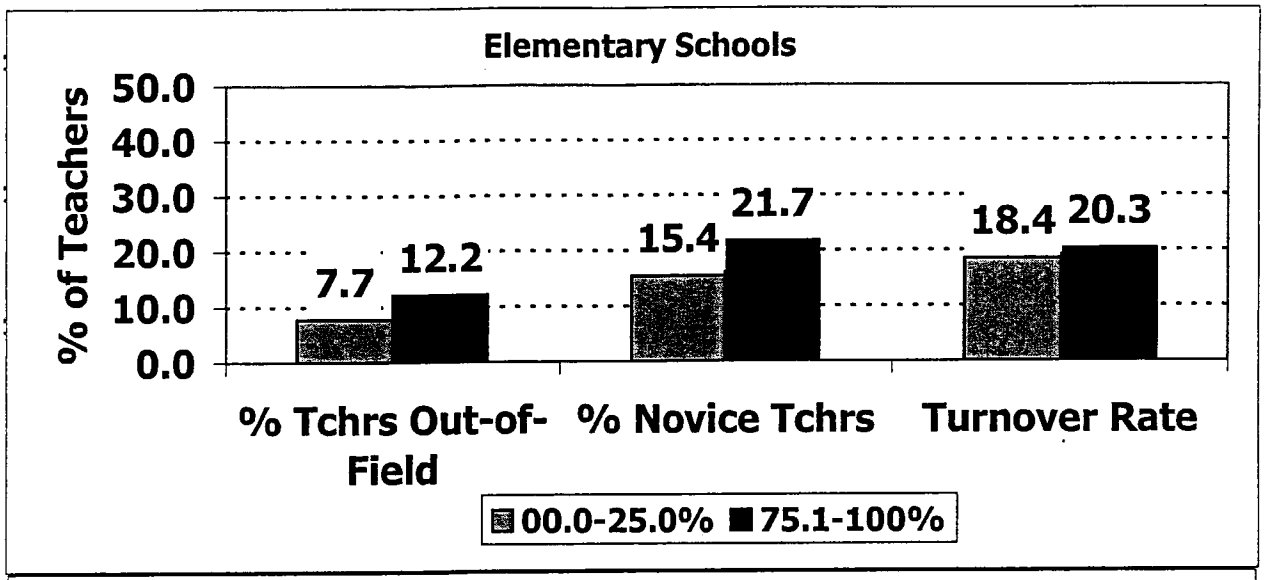



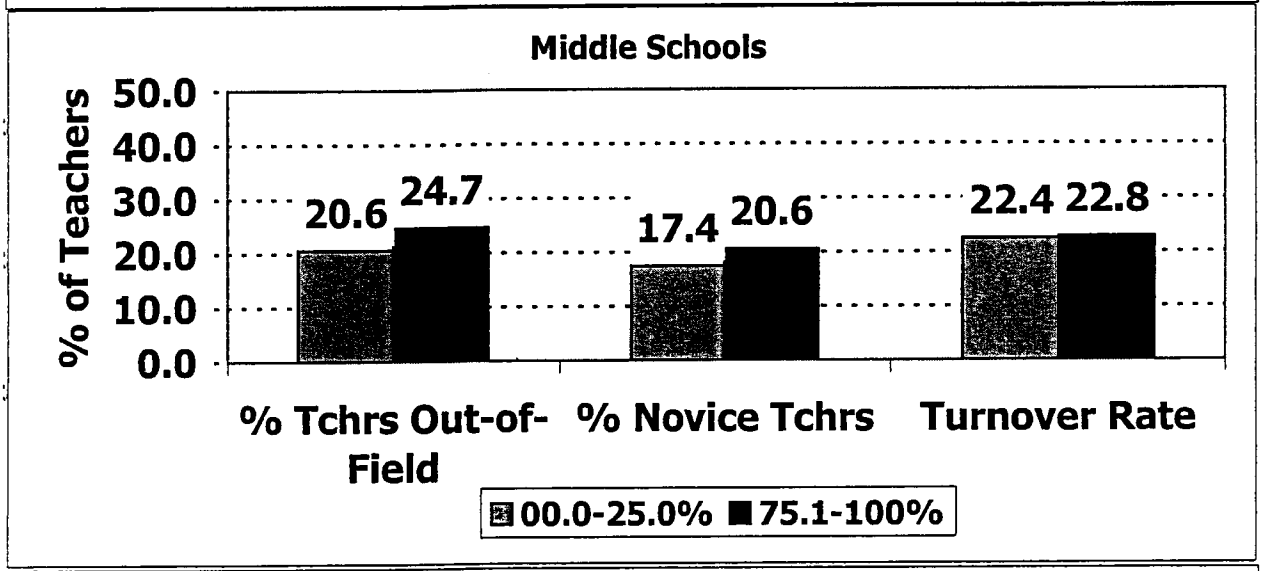

All three measures have been found to have a significant effect on student achievement. Teachers assigned out-of-field, novice teachers, and teacher turnover all negatively affect student achievement. All three measures are based on teacher ful-time equivalents (FTEs) as found in PEIMS.

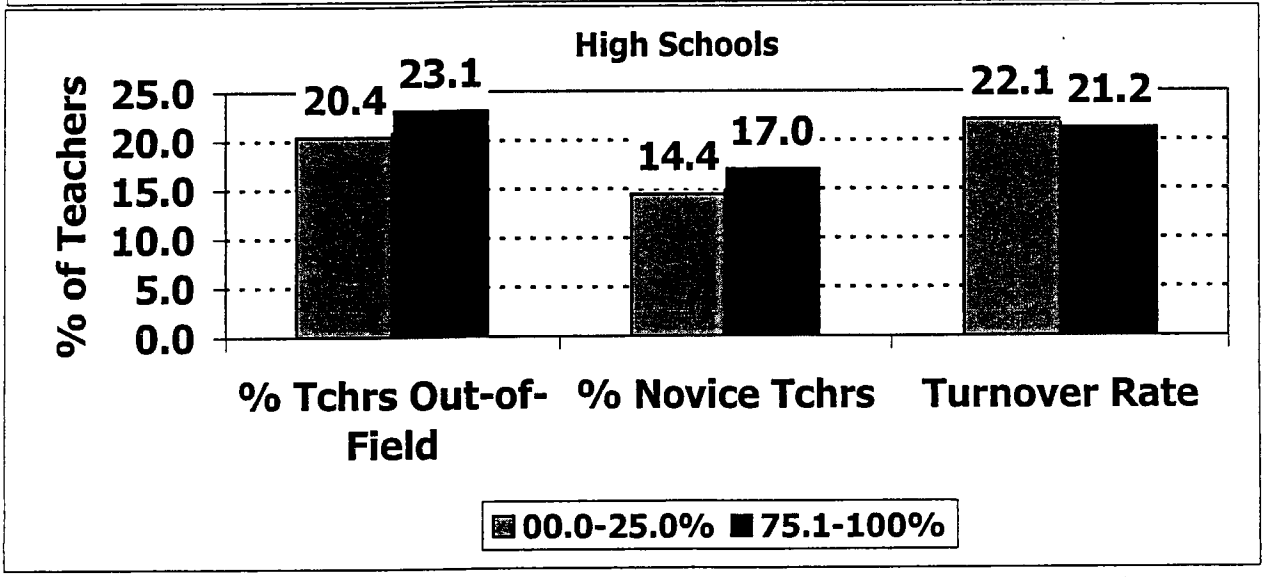
SOURCE:
AEIS and PEIMS, TEA

Dr. Ed Fuller

# Teacher Quality Indicators for Schools Serving High and Low Percentages of Hispanic Students (2004)



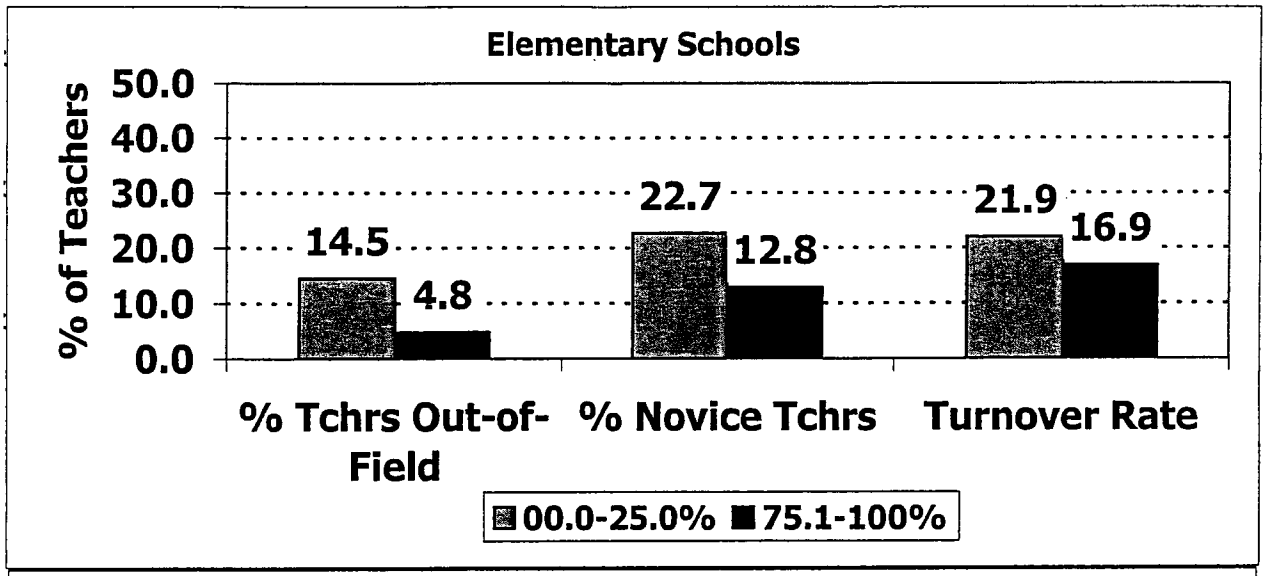



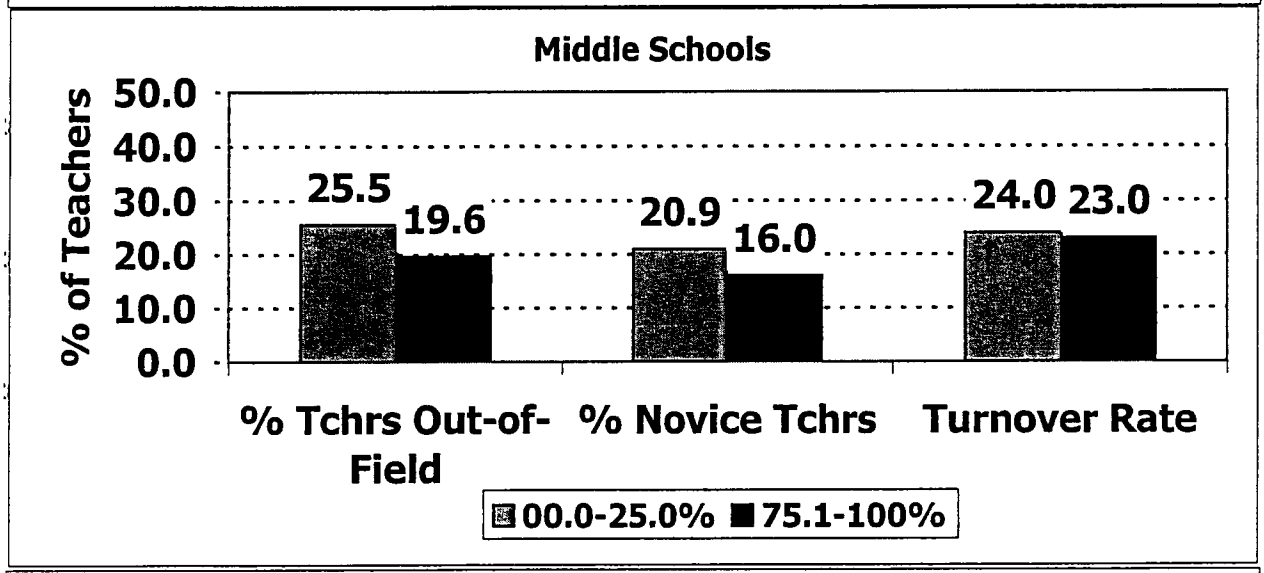

All three measures have been found to have a significant effect on student achievement. Teachers assigned out-of-field, novice teachers, and teacher turnover all negatively affect student achievement. All three measures are based on teacher ful-time equivalents (FTEs) as found in PEIMS.

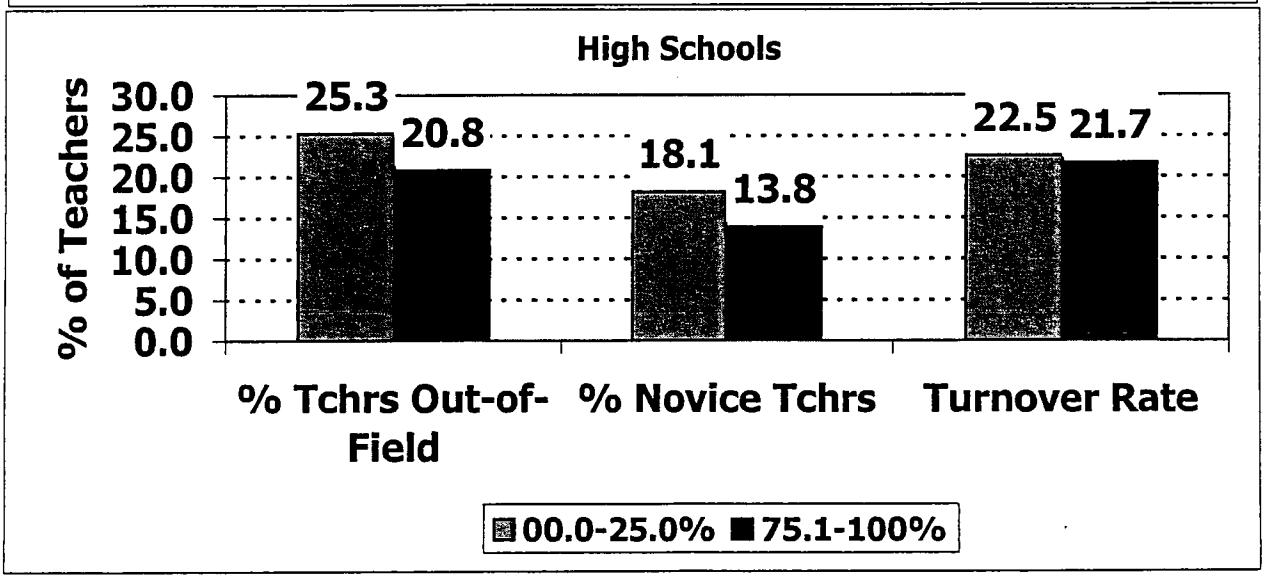
SOURCE:
AEIS and PEIMS, TEA

Dr. Ed Fuller

# Teacher Quality Indicators for Schools Serving High and Low Percentages of White Students (2004)



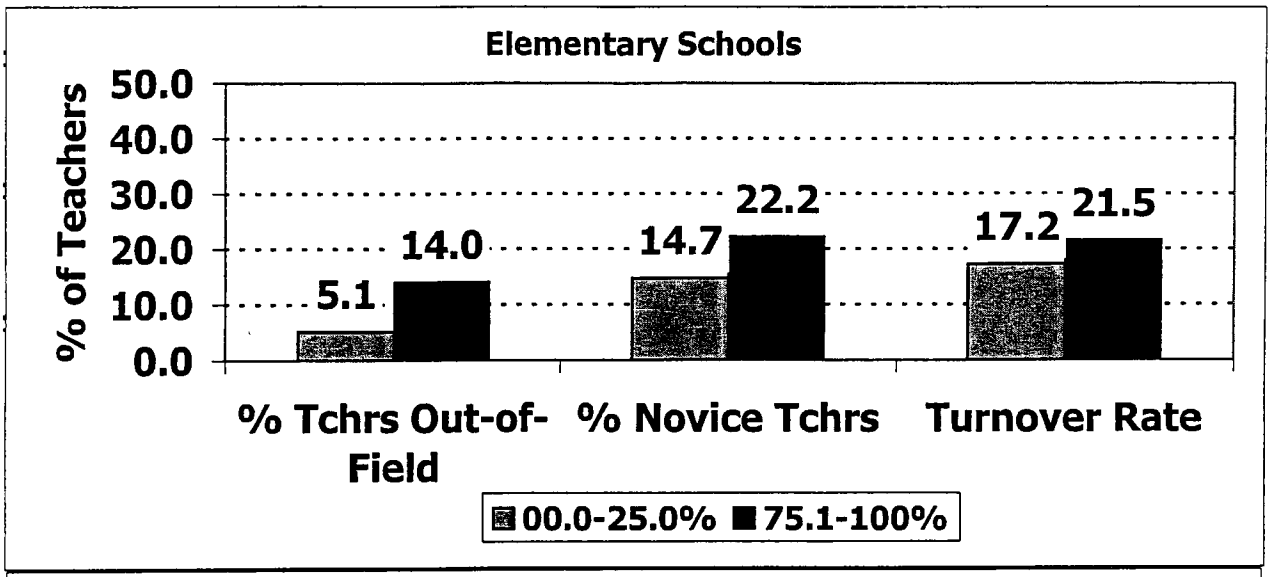



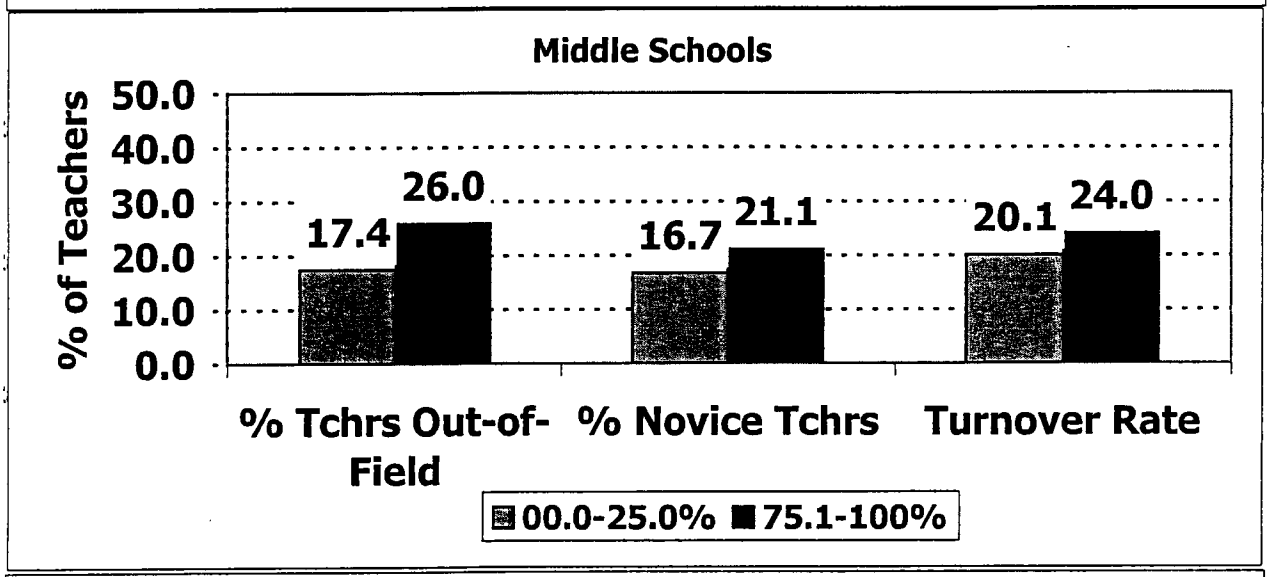

All three measures have been found to have a significant effect on student achievement. Teachers assigned out-of-field, novice teachers, and teacher turnover all negatively affect student achievement. All three measures are based on teacher ful-time equivalents (FTEs) as found in PEIMS.

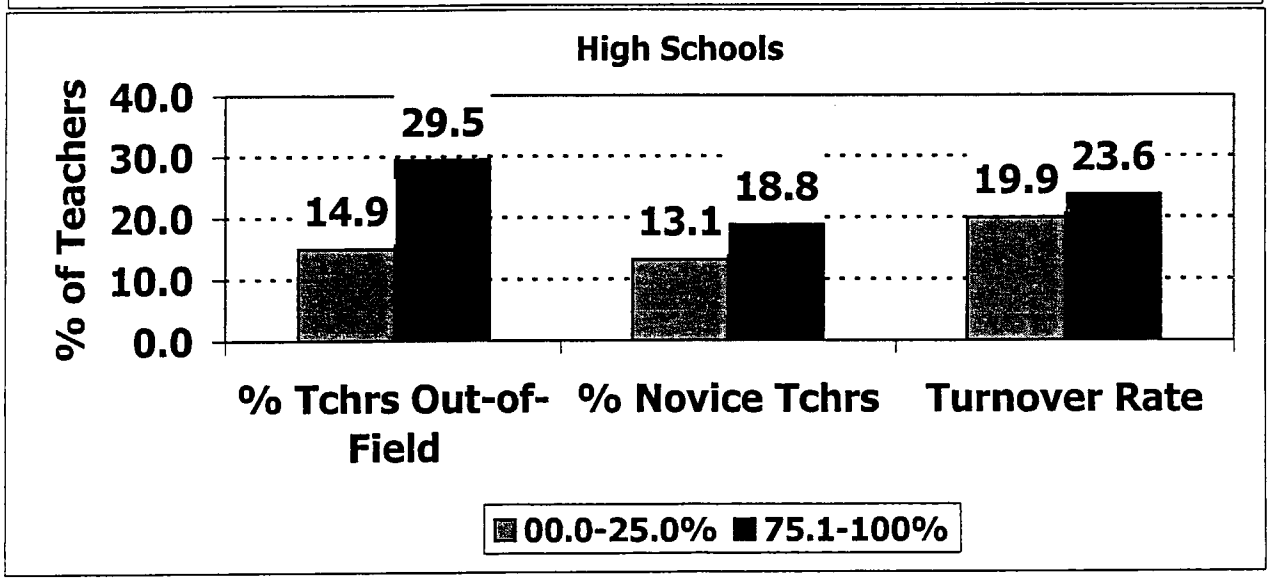
SOURCE:
AEIS and PEIMS, TEA

Dr. Ed Fuller

# Teacher Quality Indicators for Schools Serving High and Low Percentages of Economically Disadvantaged Students (2004)



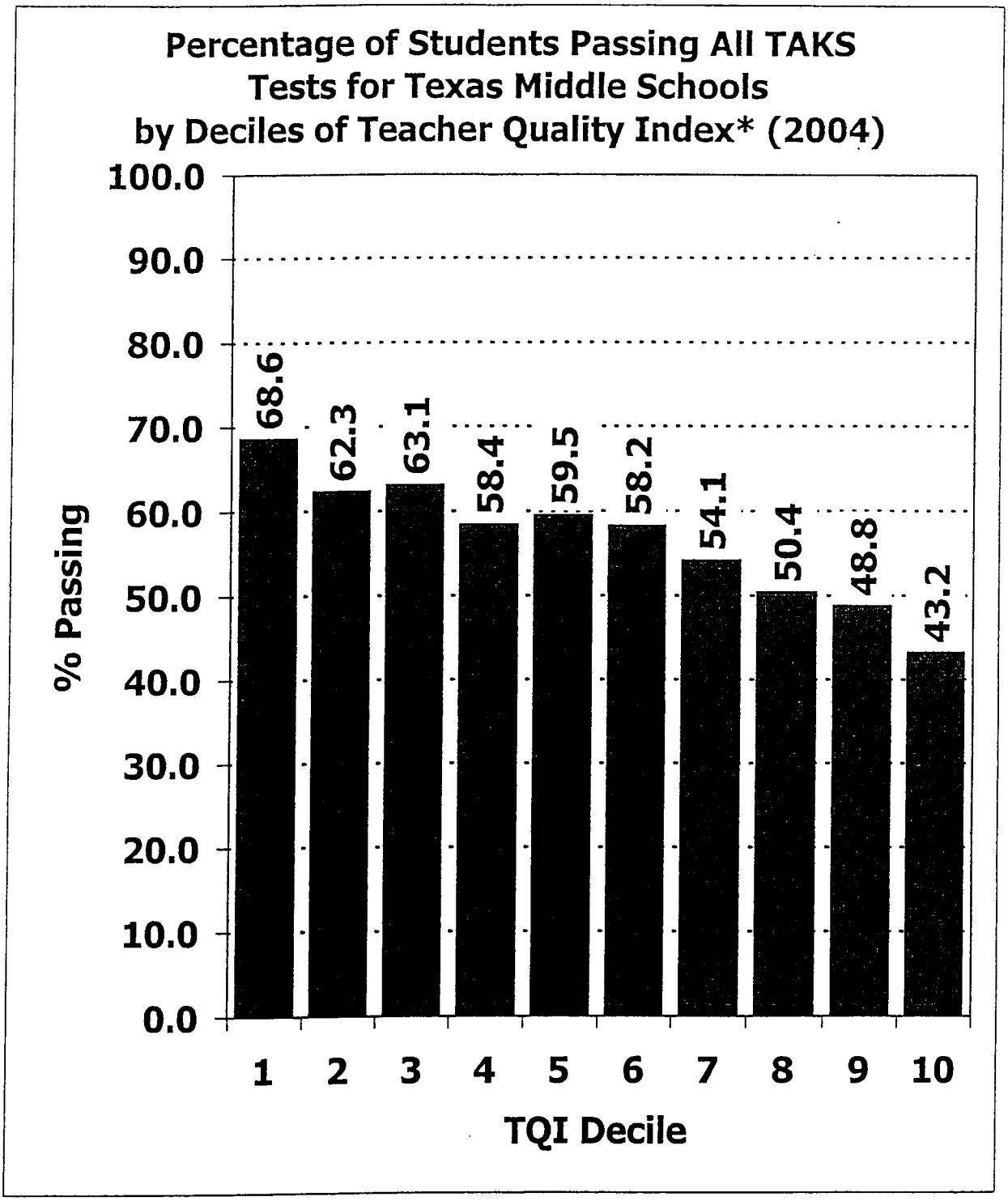




All three measures have been found to have a significant effect on student achievement. Teachers assigned out-of-field, novice teachers, and teacher turnover all negatively affect student achievement. All three measures are based on teacher ful-time equivalents (FTEs) as found in PEIMS.

SOURCE:
AEIS and PEIMS, TEA

Dr. Ed Fuller




The Teacher Quality Index (TQI) is based on three components of teacher quality at the school level: the percentage of self-contained, English, mathematics, science, and social studies teachers assigned out-of-field, the percentage of novice teachers (teachers with less than 3 years of experience), and the teacher turnover rate. All data are from the 2003-04 academic year, with the exception of the turnover rate which measures the percentage of FTEs leaving a school from 2003-03 to 2003-04.

SOURCE: