D R A F T

Not for Quotation

TEACHER PERFORMANCE INCENTIVES, COLLECTIVE BARGAINING AGREEMENTS, AND STUDENT OUTCOMES

Randall Eberts, Executive Director
W. E. Upjohn Institute for Employment Research

Kevin Hollenbeck, Senior Economist
W.E. Upjohn Institute for Employment Research

Joe Stone, W. E. Miner Professor
Department of Economics
University of Oregon

Version 1.2

December 1999

Prepared for
"Devising Incentives to Promote Human Capital,"
A National Academy of Sciences Conference
Irvine, California
December 17-18, 1999

Comments are invited and should be sent to the first author. The research assistance of Kristine Kracker and Noyna DebBurman is gratefully acknowledged, as is the excellent clerical assistance of Phyllis Molhoek and Claire Black. The views expressed do not necessarily represent the views of The W.E. Upjohn Institute for Employment Research or The University of Oregon. The usual disclaimer applies.

W.E. Upjohn Institute for Employment Research
300 S. Westnedge Ave.
Kalamazoo, MI 49007

1. Introduction

Public discontent with the performance of U.S. public elementary and secondary schooling in raising student achievement is deep. International comparisons of test scores place the U.S. in the lowest echelons, and comparative analyses of student achievement across regions and demographic groups point out sizeable gaps. Urban schools appear to be the worst performers. In response, reformers have advocated incentive-based or market-driven educational reforms to improve school quality such as merit pay, policies to weaken collective bargaining, enhanced choice through charter schools or voucher programs, and school-based performance awards. Some incentive-based policies have been tried in many places for many years; others are quite new. In either case, the evidence to date is weak, at best, that any of these kinds of policies have been effective in raising student achievement.

Why don't we see better results from the various reform proposals? Some would say that the implementation of these reforms has not gone far enough. Merit pay increases are not large enough to compensate teachers for the risk involved or to induce them to excel in the classroom; implementation of vouchers has not been sufficiently widespread to create a competitive environment to prompt more innovations and accountability in public schools; and charter schools have not been in operation long enough to overcome initial startup problems.

Teacher unions may also thwart the success of market-based reforms.[1] There is strong sentiment that the goals of teacher unions, as manifested in collective bargaining agreements, are not aligned with the goal of improving student performance. Critics argue that contract provisions to limit management prerogatives and improve the workplace environment of teachers divert resources and attention away from students and create little incentive for teachers to work to improve the

[1]For example, Ballou and Podgursky (1997), pp. 107-08, present this argument.

educational environment. For some time, proposals have been offered to weaken unions as a means to increase competition, give managers more discretion, and reconstitute an incentive structure in the compensation system.

Another possibility for the weak evidence on the effectiveness of incentive-based reforms is precisely because they have been successful. Since public schools are used as the comparison group for these evaluations, there may be little difference in performance between the two types of schools, simply because the reforms have improved outcomes for students in both private and public schools.

A fourth explanation may lie in the inherent nature of the educational process. Education involves multiple stakeholders, disparate and conflicting goals, complex and multi-task jobs, team production, uncertain inputs, and idiosyncratic elements contingent on the attributes of individual students, efforts and attitudes of fellow teachers, and classroom environments. The complexity of the process may tend to mitigate student achievement effects of reforms based on individual incentive pay.[2] There is a large literature on incentives that demonstrates that even in private-sector workplaces, only a small proportion of jobs base compensation on explicit contracts that reward individual performance. Rather, private sector companies prefer to reward individuals based on discretionary subjective measures of performance or to follow bureaucratic rules that establish job grades and promotion criteria.[3] Furthermore, private sector businesses reward workers more through promotions and group-based merit systems than through individual merit rewards (Prendergast, 1999).

[2]See Murnane and Cohen (1986).

[3]Others have cited problems with successfully implementing a merit pay system even in private business. A study in the early 1980s on the topic found that the practices of merit pay in private industry is neither as common nor effective as many believe (Lawler, 1983).

The purpose of this paper is to review evidence on the effectiveness of various incentive-based or market-driven educational reforms. We review these reforms in the context of incentive systems as outlined in Prendergast (1999). Specifically, we consider merit pay, charter schools, and school-based performance systems. The next section of the paper considers merit pay. Merit pay reforms attempt to incorporate individual pay-for-performance systems into school compensation systems. Many of these systems have been introduced in non-union schools; a few have been successful in gaining union approval. Very little empirical evidence has been produced concerning the influence of merit pay on student achievement in either case. We have recently acquired student achievement data from a high school that is operating a merit pay system within a unionized district, and we hope to soon contribute to the empirical literature on effectiveness of merit pay systems. In this paper, we describe its structure and operation, as a basis for evaluating the program.

In section three, we examine the literature on the effects of collective bargaining on student outcomes in order to discern the extent to which union goals are aligned with student performance goals. The basic question is whether or not unions negotiate contracts that allocate resources and foster work relations that are compatible with a positive educational environment.

The fourth section explores a relatively recent variation on a market-based system that introduces choice to students and parents, and thus creates a competitive environment for public schools. Charter schools are publicly-supported organizations, authorized by a state-sanctioned body, governed by a board of local citizens (not elected), and held accountable to explicit goals that constitute their charter. Michigan's charter school movement is one of the earliest and most established. Recently an evaluation mandated by the state legislature was completed. However, it did not attempt a rigorous analysis of the effect of charter schools on student achievement. To

augment that evaluation, we estimate the difference in student test score levels and gains between charter schools and traditional public schools in Michigan.

The final section of the paper considers explicit group incentives—namely, school-based performance awards. Our interpretation of the evidence leads us to consider school-based performance systems as the most promising of the reform proposals. We reach this conclusion for several reasons. First, the incentive system is more compatible with the nature of the educational process than individualized incentive systems such as merit pay. Second, it is more consistent with the risk preferences of teachers and their response to incentives. Third, it does not directly threaten collective bargaining compensation agreements that exist in large part to avoid subjective evaluations of performance. Fourth, the generally weak evidence on the benefits of extant market-based systems leads to serious consideration of reforms that do not rely solely on the power of compensation, which may be too weak on the margin to induce innovations and accountability in public schools. Whereas school-based performance awards may be the best means for inculcating incentives into elementary and secondary schools, they still may result in unintended consequences. Thus, administrators and policymakers must continuously monitor and modify these systems to overcome problems as they arise.

2. Performance-based Compensation Systems for Individuals

2.1 Theoretical Perspective

This first substantive section of the paper uses the framework set out by Prendergast (1999) to place the delivery of educational services in the context of incentive-based compensation systems (merit pay), as well as to show how the nature of education complicates and confounds the

effectiveness of such systems.[4] The simple, static principal-agent model that Prendergast explicates rewards agents for taking on additional risk through a pay-for-performance contract with higher (mean) wages. In his model, the performance measures used are noisy and the efficacy of the incentives depends on the risk aversion of the agents. As is well known, incentives may result in unintended, sometimes perverse, consequences. Prendergast uses the term "dysfunctional behavioral responses;" Murnane and Cohen (1986) call it "opportunistic behavior."

Institutional factors that may result in such dysfunctional responses include poorly defined or poorly measured outcomes leading to a reliance on subjective evaluations that may be flawed, multi-tasking by job incumbents, team production, and multiple principals/stakeholders. An obvious condition that must hold for pay-for-performance schemes is that performance is measured. But measurement may be costly or difficult. Occupations where "output" has an important quality dimension that is subjectively measured (e.g., design, or arts and entertainment), or whose output is a dimension of quality of life (e.g., most medical or related occupations), are most likely to rely on subjective evaluations of performance.

Subjective evaluations may be flawed because (a) evaluators may be subject to a moral hazard problem if they are being compensated based on their performance, and they can claim part of the evaluatee's performance as their own (theft); (b) individuals being evaluated may engage in non-productive activities to curry favor with their evaluators; or (c) evaluators may end up with distributions of ratings that are compressed because of a reluctance to give very high or very low ratings[5] (Prendergast, 1999, pp. 29-31). Jobs in which the incumbents perform many different tasks

[4]Much of the argument presented here was also presented in Murnane and Cohen (1986).

[5]In their case studies, Murnane and Cohen (1986) found that school principals reported causing morale problems by subjectively rating very good teachers as "Excellent," when there was a higher level category of "Outstanding." These problems would be avoided by giving fewer or many more "Outstanding" ratings.

also strain an incentives-based compensation contract. First, multiple tasks imply multiple performance measures, some of which may have measurability problems. Second, if performance measures are skewed in how they relatively weight the tasks, then the agent may respond by investing too much effort into the tasks that receive the most weight in the performance measurement system.

Team production introduces the 1/n problem, in which each individual's contribution (and reward) is diluted by the size of the team. Furthermore, if the individuals' contributions to the team are costly to observe or measure, then team-based incentives may lead to free riders. The problems for an incentive-based compensation system when there are multiple stakeholders come from a potential for misalignment of organizational goals. In effect, the principal-agent arrangement becomes a "principals"-agent problem, and multiple principals may have different, and conflicting, goals that they want followed. For example, training directors and production supervisors may conflict with each on how to reward an individual's (paid) time spent in training activities.

Each of these four constraints on the effectiveness of incentives-based compensation—need for reliance on subjectively measured outcomes, multiple tasks undertaken by incumbent workers, team production, and multiple stakeholders—characterize the teaching and learning process in schools. Learning outcomes can be and are assessed through standardized tests, which are amenable to performance-based contracts (particularly if value-added measures are available). However, there are many additional dimensions to student learning and development that are either not assessed, or are assessed but not with standardized instruments, so that ultimately evaluations are inherently subjective.

Schools (at all levels of the K-12 system) typically have literally dozens of learning processes or programs going on simultaneously. These might include core academic subjects; non-core

academic subjects such as art, physical education, music; acquisition of technology skills; career development; special education; extracurricular offerings; gifted and talented programs; human growth and development; and remediation or developmental education. At some levels of education, e.g., elementary grade levels, teachers may individually be involved in virtually all of these programs. Even at secondary school levels, which are typically organized by academic discipline, teachers get involved in areas such as career development, extracurricular leadership, and special education (as mainstreaming becomes the status quo). But the myriad of programs or processes that occur in schools is only one source of multi-tasks. Even within a teacher's discipline, multiple tasks comprise the teaching and learning process–curriculum development and planning, instruction, and assessment, for example. Furthermore, good teaching requires attention to students' learning styles, which may mean multiple modes of instruction.

Education, to some extent, requires team production. For example, many elementary and middle schools are organized into teams of teachers. The notion, at the elementary level, is to rely on teachers' comparative advantages in core academic areas so that each teacher in a team is responsible for the subjects with which they are most comfortable. But even apart from explicit team teaching, departmentalized secondary schools result in team production because students' performances on standardized tests depend on learning in several courses taught by different teachers.

Finally, school governance and control is characterized by many different stakeholders with differing, and sometimes, conflicting goals. Administrators who are accountable for direct student achievement may be most responsive to levels of test scores. School board members who are accountable for resource decisions may be most interested in changes (value added) over time in test scores. Parents may be most concerned about postsecondary education attendance rates, whereas

employers may be most concerned about "soft," employability skills such as problem-solving, attendance, and attitude.

Another characteristic of most school districts is that they have very little control over their revenue streams. As noted, incentive-based contracts allocate part of the production risk to the employees in return for higher rewards (wages). But since school administrators have very little revenue to share, they cannot offer very sizeable increases in compensation were teachers willing to accept the risk inherent in a merit-pay system.[6]

In short, while empirical evidence and common sense show clearly that economic actors respond to incentives, we argue that there are several wedges between performance measures and the actions of teachers that tend to mitigate against individual level, incentive-based compensation schemes. In developing this paper, we found little empirical evidence on the effects of merit pay on student achievement. Most of the literature on merit pay systems documents the institutional experiences in districts. Those experiences, for the most part, have been rather short-lived and usually negative. For example, a major study of merit-based pay (Hatry, Greiner, and Ashford; 1994) found that most (75%) merit-pay programs that had been in existence in 1983 and had been studied by the researchers, were no longer operational in 1993.[7] An interesting self-described limitation of the Hatry et al. (1994) study is that they did not examine student achievement. They note,

> We would especially have liked to have performed an in-depth analysis of the impact of incentive programs on student achievement. However, very few of the participating districts had attempted any systematic evaluation of the effects of their

[6]We do note below that collective bargaining seems to impose added costs (inefficiency) to a district. Perhaps these costs could be identified, captured, and re-allocated as merit pay. Also market-based alternatives to traditional public schools may be less reliant on state or local government funding, and therefore be in more control of their revenue.

[7]Murnane and Cohen (1986) also emphasized the short-lived, ineffective nature of merit pay systems.

> incentive plans on student achievement, even though a basic assumption behind incentive plans is that teachers can indeed significantly affect learning. (pp. 7-8).

In a dissertation study involving one district in Pennsylvania, Tulli (1991) found no correlation between gains in student achievement and teachers awarded merit bonuses under this district's plan. The author noted, however, that this district heavily weighted inputs such as attendance, participation in professional development, and supervision of extra activities relative to student outcomes.

2.2 A Case Study of a Merit Pay System Within a Collective Bargaining Environment.

To add to the empirical literature on merit pay systems, we have acquired data from a particular high school that implemented a merit pay system in 1996. What is interesting about this school is that it is the only school building in its local public school system that is non-unionized. The Michigan district where this high school is located has an enrollment of about 9000 students and has 15 school buildings; three of which are high schools. Community High School is an alternative education facility that has an enrollment of approximately 500 students pursuing a high school diploma. The school also has an adult basic education enrollment of about 100 students pursuing GED preparation, a high school degree, or English as Second Language (ESL) programs. There are approximately two dozen faculty members at Community High School.

This school implemented its merit pay system at a time of great transition in adult education and K-12 educational funding in Michigan. In 1995, the state legislature took two actions that influenced adult and alternative education institutions in the state. First, in the State budget, the legislature virtually "zeroed out" all state funds for adult education and transferred those funds into the state economic development agency to be used for customized training. Second, the State passed

a comprehensive public education finance reform that moved the state from a system primarily funded at the local level by property taxes to a system which is State-funded, primarily through sales taxes. The new state finance system funds "pupils" (students under age 21 in elementary or secondary schooling and pursuing a program leading to a high school diploma) at a foundation allowance that is nearly identical for all districts. Given that the State funds alternative education pupils at approximately $6,000 per year (1999 $), Community High School decided to disengage from the adult education consortium it had been in prior to 1995, and to focus primarily on alternative education for high school students. In the same year that the legislature revised the school finance system, it also passed a law that allows school districts to operate pilot programs that may be exempt from collective bargaining requirements.

The district decided to operate Community High School as a "pilot" program with a performance-based compensation scheme for its teachers, who collectively decided to remain separate from the local district's education association (union). Alternative education settings are characterized by students who have often not succeeded in traditional school settings, and usually experience attendance problems and intermittent dropping out and re-enrollment episodes. Consequently, the performance-based incentives are targeted on student retention.

Description of the performance-based compensation system. Teachers are paid a base wage that depends solely on their education (not experience or tenure) for each 60-minute class that they teach. Teachers with a Master's degree or higher receive a 5 percent higher base pay. Two supplements may be earned and added to the base pay. A retention bonus is paid if 80 percent or more of the students assigned to the class (as of the end of the 2nd week of the quarter) are still

enrolled and attending at the end of the quarter.[8] The bonus is the same for all teachers, no matter what their educational background, and is approximately 12.5 percent of the base for Bachelor's degree teachers (12 percent for Masters plus).

The second supplement is based on student evaluations. Students rate the following 15 factors on a 1 - 5 scale:

- Objectives, requirements, and expectations for the class were clearly started
- The instructor was well-informed and had current knowledge of subject matter
- The instructor was well-prepared for each class
- The instructor presented the material clearly
- The material was presented in an organized manner
- The instructor used class time well
- The instructor was interested in students and was willing to listen to them
- The instructor encouraged student participation and welcomed questions, discussion, and different points of view
- The instructor encouraged a high level of student attendance in the class
- The methods of evaluating student progress and performance were clearly stated
- The instructor encouraged students to think critically
- The instructor was enthusiastic
- The instructor was aware of the varying levels and abilities of students
- the materials used were appropriate for the subject taught
- This class was worthwhile

Teachers who receive an average rating of 4.65 or higher[9] on the 5-point scale for all 15 items in all of their classes (weighted by class enrollment) in each quarter for four (4) consecutive quarters receive the performance bonus, which increases their base pay by about 5 percent **and** increases their retention bonus by 10 percent.[10] To give the reader a sense for the size of these bonuses, during

[8]The initial enrollment in the class for purposes of calculating retention is capped at 20, so to earn the retention bonus, teachers must have 16 students or 80 percent of the initial enrollment, whichever is less. Sometimes actual class sizes exceed 20, and in these cases, the retention bonus is still earned if the ending enrollment is 16 or more. This was seen as an incentive for teachers to allow larger enrollments in their classes, when it was warranted by scheduling and overall enrollment concerns.

[9]This was the average rating in the previous school year.

[10]Hatry et al. (1994) found a range of merit pay awards in their study from at most 25 percent of salary to 5 percent or less. See also Lawler (1983).

school year 1998-99, the base pay for a teacher with a Bachelor's degree was $816 per class ($22,848 for nine months; 4 quarters with 7 classes).[11] With the performance bonus and retention bonuses in all classes, the per class pay would be $979 ($27,412 for nine months; 4 quarters with 7 classes).[12]

Impacts on student outcomes. Do the performance incentives affect student outcomes at this school? To answer this question, our intention is to analyze data from students at this school and at a very similar school in the county that relies on a traditional experience/education compensation scheme. In particular, we have obtained data for the two school years prior to and after the implementation of the performance incentive system (i.e., 1994/95 and 1995/96; 1997/98 and 1998/99) for both high schools. Our plan is to perform a difference-in-differences analysis of several outcomes including retention, attendance, grades, and graduation rates.

In summary, merit-based pay schemes have been tried in many districts over a long period of time, but have made little in-road.[13] Bureaucracies remain the status quo organizational model in education. Bureaucracies, as defined by Prendergast (1999), are "entities that use rules or regulations to determine resource allocation rather than individual incentives." In many states, the employment contract between the bureaucracy (principal) and teachers (agents) is laid out explicitly through collective bargaining at the local level. The next section of the paper examines incentives within a collective bargaining environment.

[11]Community High School is on an 8-period per day schedule, and the average teaching load is 7 classes.

[12]Many teachers have more than 6 classes per term. With at least 6, the teachers receive full benefits equivalent to the unionized teachers in the district.

[13]Johnson (1986) documents the ebb and flow of enthusiasm for merit pay plans during the last century, citing evidence of heightened interest in the 1930s and 1960s.

3. Collective Bargaining and Incentives

In a paper that examines ways to redesign teacher salary systems for educational reform, Firestone (1994) summarizes the concerns of teachers about being evaluated with systems that link pay to performance. These concerns are closely aligned with the issues raised earlier about the problems with performance measures and multi-tasks.

- Teachers do not believe that evaluative methods, whether in the form of subjective evaluations or student test scores adequately determine the quality of their performance.
- They doubt that current methods can adequately capture the diversity of educational goals that they deem desirable nor address the variation in their circumstances, such as different subject matter, teaching conditions, and student abilities.
- Teachers fear that favoritism will predominate in systems that rely on the observations of principals or department heads.
- Teachers are concerned that pay systems will reward activities that take time away from their instruction and preparation time.
- Individual incentive programs interfere with collegial relationships between teachers by discouraging team efforts.
- Pay systems force teachers to pursue pedagogical approaches that run counter to their view of good teaching.

One view of collective bargaining agreements is that they are a response by teachers to establish explicit rules about various types of conduct, working conditions, and protection from arbitrary treatment. These provisions appear to be in line with teachers' perceptions about desirable working conditions (e.g., class size), protection from arbitrary sanctions (reduction in force rules), and importance of certain activities (preparation time). Results from the few empirical studies of teacher collective bargaining and student outcomes show that many of these rules are positively related to student test score gains and that teachers covered by these rules exert more effort toward

these activities. In this respect, the preferences of teachers and the goals of their supervisors (e.g., school administrators) and stakeholders (parents and their representatives, board members) are fairly closely aligned. However, we argue, with supportive evidence, that the work rules and employment restrictions in teacher collective bargaining agreements lead employers to deviate from the least-cost combination of inputs, as predicted in the literature on bureaucratic rules (Prendergast, 1999, p. 38, citing Milgrom and Roberts (1988) and Tirole (1992)).

3.1 Working Conditions, Classrooms, and Student Achievement

Collective bargaining agreements establish rules that affect the working conditions of teachers and thus the school environment. If improvements in working conditions are in line with factors that positively affect student achievement, then collective bargaining can lead to improved student outcomes. Three studies offer evidence of the effect of collective bargaining on inputs into the educational process. Using a nationally representative data from the Sustaining Effects Survey of elementary schools, Eberts and Stone (1984) report that teachers covered by collective bargaining have smaller classes and spend more paid time in preparation. They find that unionized teachers have 4 percent more paid time to prepare for classes and that the student-teacher ratio is 12 percent lower. The study also finds that unionized teachers spend less time in instruction, but this reduction is offset by increases in preparation, administrative duties, and consultations with parents.

The tendency for teachers covered by collective bargaining to experience smaller classes is supported by Kleiner and Petree (1988) and Hoxby (1996). Kleiner and Petree rely on state-level data to examine the relationship between the percentage of districts covered by collective bargaining within a state and the student-teacher ratio. They find that as the percentage of districts with a

collective bargaining agreement increases from 0 to 100 percent, the student-teacher ratio falls by 7 percent. Hoxby's estimates, based on district-level data, show a similar reduction in the student-teacher ratio for unionized teachers. Using a difference-in-differences specification, she estimates class size to be 6 percent smaller in union districts. Introducing instrumental variables into the specification yields a 9 percent reduction in class size.

Collective bargaining affects the classroom environment not only by limiting class size but also by influencing instructional modes. Only two studies we know of have considered the effect of unions on instructional modes, and their general conclusion is that teacher unions, like other unions, tend to standardize the workplace. Eberts and Stone (1984, p. 148-150) find that districts in New York state that bargain class-size provisions are more likely to adopt traditional classroom instruction as opposed to other arrangements. In their national study of elementary students, Eberts and Stone (1984) examine union-nonunion differences in instructional arrangements. They find that fourth graders in unionized districts on average spend 42 percent less time with a specialist, 62 percent less time with an aide, 26 percent less time with a tutor, and 68 percent less time in independent, programmed study. Argys and Rees (1995) add to the evidence of standardization by showing that students of low and high academic ability are in larger classes in schools where the teachers are unionized as compared to nonunion schools.

Unions also affect the involvement of teachers in the educational process and their relationship with principals (a "voice effect"). Eberts and Stone (1984, pp. 150-153) find that elementary teachers covered by union contracts place more importance on participation in student assignment and teacher assignment than do teachers not covered by contracts. The differences are statistically significant. Unions affect the actual level of teacher participation in educational decisions in two ways. First,

those teachers who think participation in decision-making is more important report that they are more active in making such decisions. Second, teachers covered by collective bargaining are more actively involved in decisions regarding teacher assignment than those not covered, controlling for differences in the importance teachers place on making those decisions.

Eberts and Stone (1984) also consider the effect of collective bargaining on the ability of principals to provide leadership, particularly with respect to instructional standards. In the national Sustaining Effects Study, elementary teachers were asked to rate the leadership abilities of their principals by responding to several questions about the principal's role in the mathematics program. Teachers in union districts responded that their principals provided more instructional leadership, devoted more time to activities related to math curriculum development and devoted more time to needs assessment, program planning, and evaluation of the math program (pp. 57, 159). These attributes have been cited as important to improving student outcomes.[14]

3.2 Student Achievement

A number of studies have found that the factors discussed above, namely class size, allocations of teacher time to preparation and instruction, and instructional leadership, are positively related to gains in student achievement.[15] A much smaller set of studies has examined the direct effect of collective bargaining on student outcomes. Stone (1998) summarizes and critiques the results of

[14]See Eberts and Stone (1984, 1986) for evidence and other work that addresses this issue.

[15]The large literature on educational production functions and whether or not inputs into the schooling consistently and significantly affect student outcomes is hardly without controversy. Hanushek (1986) raises the issue of whether inputs matter, and reanalysis of the same literature by Hedges, Laine, and Greenwald (1994) draw the opposite conclusion. As Ladd (1996) points out, those studies based on sounder methodologies, such as analysis of student-level data using pre- and post-tests and controlling for school- and home-based resources, generally show that school inputs do affect student test score gains.

seven such studies. While none of the studies are based on the ideal experiment of randomly assigning students to union and nonunion districts, several studies do include extensive control variables for students and school attributes and for nonrandom assignment of students. Eberts and Stone (1984, 1987) use detailed student, teacher, and school data from a national sample of 14,000 elementary students to examine their relative achievement in union and nonunion schools. Individual student gains in math were assessed by administering a pre-test at the beginning of the academic year and a post-test at the end of the year. Eberts and Stone estimate a value-added specification of an education production function in which each student's post-test score was regressed on the pre-test score and a host of other control variables. They found that students in districts covered by a collective bargaining contract scored roughly 1 percent higher on the post-test, or about 3.3 percent higher as a percentage of the average gain from the pre-test to the post-test. The union effect is statistically significant.

Milkman (1989) follows a similar approach for twelfth-grade high school students. He uses a standardized math test administered in the tenth grade as the pre-test for his specification of an educational production function. His results are similar to those of Eberts and Stone. Students in union districts scored 2 percent higher than students in nonunion districts. In a separate but similar study, Milkman (1997) finds that minority students in union districts score about 1.4 percent higher than similar minority students in nonunion districts. Grimes and Register (1990) and Argys and Rees (1995) also find small but significantly positive union effects on student achievement.

Studies using aggregate state data find larger positive effects of unions than those using student-level data. Kleiner and Petree (1988) find that SAT and ACT scores are 6 to 8 percent higher in states with 100 percent union representation versus states with no representation. Their

specification includes a model with fixed state effects, but few individual or school control variables. Nelson and Rosen (1996) also look at aggregate state data and include more detailed control variables than Kleiner and Petree but without fixed effects. They find that students in states with more than 90 percent union coverage score on average 4.5 percent higher on SAT tests than students in states with fewer than 50 percent union representation.

The positive effect of unions on student achievement is not enjoyed by all students. The student-level studies by Eberts and Stone, Milkman, and Argys and Rees find that an inverted U-shaped effect of collective bargaining on student achievement gains. For students of average ability, as measured by pre-test scores, those in union districts score higher on post-tests than those in non-union districts. The opposite is true for low-achievers and high-achievers. Below- and above-average students in nonunion districts score higher than those in union districts. Eberts and Stone interpret these results to provide additional evidence of the tendency of unions to standardize the workplace.

These results for below- and above-average students are consistent with those presented in Hoxby's (1996) detailed district-level of study of the effect of unions on high school drop-out rates. She finds that the presence of collective bargaining, where at least 50 percent of the teachers are union members, increases high school drop-out rates by 2.3 percent. She interprets this result to infer that unions reduce student achievement. However, since students at the lower end of the test-score distribution are more likely to drop out of school, Hoxby's results are also consistent with the results of Milkman and Argys and Rees (and indirectly with Eberts and Stone) that show that below-average students in union districts experience less academic success.[16] Argys and Rees further explore the

[16]Stone (1998) provides this explanation to reconcile Hoxby's results with those of the six other studies that show positive union effects on student achievement.

effects of unions on students with different academic abilities and find that these effects appear to work primarily through various class characteristics including the sizes of the classes to which the students in the tails of the test-score distribution are assigned. Interestingly, in one of their specifications, inclusion of class characteristics actually erased the negative effect of unionization on low-ability students.

Therefore, empirical studies of the effect of collective bargaining on student achievement find little if any support for the argument that unions on average reduce academic success. Unions, by negotiating rules to standardize the workplace through class size provisions and more traditional instructional models, may affect students with different abilities but even these effects may be small. Thus, the codification of bureaucratic rules through collective bargaining agreements does not appear to have significant negative effects on student achievement.

This is not to say that unions do not have other detrimental effects. Unions increase the cost of education by between 8 and 15 percent and distort the least-cost combination of inputs (Stone, 1998; Eberts and Stone, 1991). These findings are consistent with the prediction that bureaucracies establish rules to avoid counterproductive activities, such as influence activities, at the expense of inefficient allocations. Eberts and Stone (1991) calculate that work rules and employment restrictions in teacher collective bargaining agreements require employers to deviate from the least-cost combination of inputs that in turn increases cost by about 8 percent, compared with non-union districts. Thus, if we accept the findings that collective bargaining agreements increase (or at least do not decrease) student achievement, then union contracts increase cost without reducing technical productivity. Eberts and Stone suggest that this observation may explain why there is little systematic evidence that work rules and employment restrictions reduce productivity in public schools.

4. Organizational Incentives in Education

4.1 Background

The nature of production and employment relationships in education appear to mitigate against the adoption of principal-agent, incentive-based contracts with teachers, so the operation of schools relies on bureaucracies, which typically operate in a unionized setting. The question that we turn to now is the extent to which group or organizational incentives can be used to motivate bureaucracies in a way that significantly improves student achievement. One type of reform is the introduction or strengthening of competitive options to public schools. Proponents argue that market pressures, primarily through student choice of schools, will improve school performance. The premise is that students and parents will choose schools based on quality factors or quantifiable performance outcomes, and those schools that have structured themselves to deliver these outcomes efficiently will survive and grow, and those that do not will transform themselves to be more productive or go out of business. Indeed, within the public-school system, Hoxby (1996) finds that students have higher dropout rates in counties where the Herfindahl measure of school-district concentration is high than in counties where it is low, raising the possibility, at least, that choice and schooling alternatives can improve student achievement indirectly through competitive discipline. Of course, these effects may be restricted, or most pronounced in large urban counties with either a single district or few districts. The other type of group incentive is school-based performance awards (SBPA's), to which we turn in a subsequent section of the paper.

The available empirical evidence about the effectiveness of market-based options is rather weak.[17] In particular, examinations of private school options, charter schools, and voucher programs

[17]Our results below are consistent with this proposition.

generally show no positive outcomes or rather weak evidence, at best.[18] Studies of the difference in performance, typically measured as student scores on standardized tests, between private (or schools of choice) and public schools find little evidence of superior performance by schools presumably driven by market forces. Evaluations of the Milwaukee Parental Choice Program, for example, are mixed but the weight of the evidence tends toward the conclusion that the benefits of school choice are small. Rouse (1998) finds test score gains for only math but not for reading. These results are consistent with the findings reported by Witte (1997), but Greene, Peterson, and Du (1997) finds that students attending choice schools did better in both subjects. Rouse concludes that "this program cannot shed light on whether vouchers provide an incentive for the public schools to improve and therefore increase the quality of education provided to all low-income children."

Neal (1997), in assessing the literature on the benefits of private schooling, concludes that "many results are quite fragile." Estimates of the achievement gains associated with private schooling often vary considerably across studies that employ the same data sources, with one exception. Catholic secondary schooling raises graduation rates. Based upon recent nationally representative data, Figlio and Stone (1999) find that with adequate correction for selection bias, neither religious nor nonreligious private schools improve either test scores or high-school completion, though minority students in large central cities do appear to fare better in private schools. This paper strives to add to the empirical evidence on the impact of market-based alternative schools by examining standardized tests administered to all students in Michigan.

[18]Neal (1997) finds rather strong positive outcomes for urban, parochial schools. However, Rothstein, Carnoy, and Benveniste (1999) suggest that there is little difference between urban public schools and private schools.

4.2 An Examination of Student Achievement in Michigan Charter Schools

Proponents of school reform have argued that charters and vouchers can provide adequate market pressure to improve school performance. The charter school movement has gained considerable attention in recent years. Since Minnesota passed the first charter school law in 1991, more than 30 states have put such laws in place (Nathan, 1999). Approximately 1,100 charter schools are in operation, enrolling over 250,000 students. Arizona, California, and Michigan have led the movement, accounting for half of those schools and students.[19]

We will focus on charter schools in Michigan. The Michigan experience with charter schools offers an interesting case in which nonunion publicly supported schools are created within a strong union environment. Furthermore, charter schools also are supposed to adhere and be held accountable to an explicit set of specific goals. The analyses attempts to provide some insight into the performance of the two types of publicly-funded schools. Michigan's law was passed in 1993, and now 170 charter schools enroll approximately 50,000 students, or about 3 percent of Michigan's K-12 student enrollments (Horn and Miron, 1999). According to Michigan law, the primary purpose of charter schools, referred to as public school academies (PSAs) is to:

- Improve pupil achievement;
- Stimulate innovative teaching methods;
- Create new professional opportunities for teachers;
- Achieve school-level accountability for educational performance;
- Provide parents and pupils with greater choices among public schools;
- Create competition among public schools to use state funds more effectively, efficiently, and equitably (Horn and Miron, 1999, p. 18).

This set of purposes reflects the intent of the original proponents of charter schools (Hassel, 1999).

[19]Hassel (1999) describes the process by which Michigan legislators passed the charter school bill. Hassel notes that Michigan is an excellent example of a state that produced strong charter school laws within a strong union environment. However, his analysis of the 30 states that passed charter school laws shows no statistically significant difference between the passage of such laws and the extent and strength of collective bargaining in the state.

Michigan charter schools generally adhere to the following rules and practice. These rules are presented in the evaluation funded by the state of Michigan and performed by Horn and Miron of Western Michigan University's Evaluation Center.

- Each school is authorized for a particular mission and has identified goals and purposes unique to that mission.
- Teachers must be certified just as they are at other public schools.
- Schools may not screen students, but they may limit the number of students they serve. If more students apply than can be enrolled, a random selection process is used.
- Charter schools are free to choose their own core curriculum, and are not required to provide services to meet the needs of all students, such as those with special needs.
- Charter schools are subject to all laws and regulations that apply to public schools.
- Charter schools receive the state foundation grant on a per-pupil basis. This level of funding cannot exceed the amount received by the local "host" district. Charter schools cannot charge tuition, but can raise funds through legal foundations, receive grants, etc. (Horn and Miron, 1999, p. 3).

State law requires that public educational institutions authorize charter schools. In a recent study of charters in Michigan, 109 of the 137 existing charter schools at that time had been authorized by universities and only 12 by local public school districts. Charter schools are governed by a board, which is approved by the authorizing entity. Board members are public officials and are subject to all applicable laws. However, unlike regular school board members, they are not elected by parents or by any other specified group and instead are officially appointed by the authorizer.

The evaluation of Michigan charter schools, conducted by Horn and Miron, offers insights into their structure and how closely they are able to adhere to the principles upon which the charter school law was established. To state succinctly, charter schools are intended to identify a specific set

of goals, align their resources to pursue those goals, offer ways to evaluate the performance in meeting the goal, and thus hold teachers and administrators more accountable for educational achievement. The evaluation cited several shortcomings with respect to the current practice of charter schools and recommended the need for improvement in the following areas:

- Vague mission statement and justification of the need for the charter school;
- Real or potential conflicts of interest among employees and board members;
- Congruence between curriculum and the philosophy/mission of the school;
- Assessment and evaluation procedures for students and employees;
- Limited innovations being developed and applied in the charter schools (Horn and Miron, 1999, p. 101).

Analysis of MEAP test scores. Of course, some charter schools did better than others in achieving these objectives. In order to provide a general perspective of the effectiveness of charter schools relative to their public school counterparts, we examine the difference in student outcomes, as measured by Michigan's educational assessment tool, the Michigan Educational Assessment Program (MEAP). The MEAP test is a convenient measure of the educational outcomes of students, since it is administered to all public school students, including charter school students. However, the MEAP test may not be an ideal measure with which to compare student outcomes across schools. First, the MEAP is not a norm-referenced standardized test, so that it is difficult to compare MEAP scores from one year to the next. Second, and related to the first, the subject areas of the MEAP test differ each year, making it impossible to use as a pre- and post-test to measure the educational gains of individual students. Comparing test score levels may be biased against those schools that target disadvantaged students. But adequate controls for the composition of students and other factors outside of the school's control are difficult to obtain. Unfortunately, reporting of variables such as

percentage eligibility for free or reduced price lunch is poor, especially for charter schools,[20] so it is difficult to put together a comprehensive set of factors to control for student composition. Third, some of the charter schools are so new that they have not enrolled students long enough to attribute the activities of the schools to the outcomes of their students in any meaningful way. Fourth, the MEAP test does not provide an adequate assessment of what the school is trying to accomplish. These arguments have been echoed by public school administrators and teachers, and underscore the problems of using standardized measures for evaluation purposes.

Nonetheless, the MEAP test is one of the few ways to compare the performance of all public schools. With greater attention given to accountability of schools, many states are adopting such instruments to assess students and some states, notably South Carolina and Kentucky, use these tests, along with other factors, to allocate state resources to schools. Furthermore, according to the WMU evaluation, many charter schools used the MEAP as evidence of the success of their program and some charter schools list the MEAP test as their only evidence of student achievement (Horn and Miron, 1999, p. 83).[21]

We examine the MEAP test scores for all fourth and fifth grade students in charter schools and in the public school districts within which the charter schools are located. By pairing charter

[20]The data set has a student-level indicator variable for free or reduced price lunch eligibility. However, this information is obtained only if the district administrator of the MEAP test consults administrative data and reports it accordingly. This was voluntary and very few districts complied. The data set also has a building-level variable for the percentage of enrolled students who are eligible for free or reduced price lunches. At charter schools, Horn and Miron (1999) found that some principals did not report this variable "because they did not have hot lunches in their schools and therefore they do not collect the information." Furthermore, for schools that do report the data, it pertains to the entire enrollment, and not just the classes that take the test. Finally, a district can only know about students who apply for free or reduced price lunch eligibility, and there may be significant levels of non-applicants who would be eligible.

[21]Researchers and evaluators will use other measures of student outcomes, such as drop out rates (e.g., Hoxby, 1996). However, since most charter schools include only K-8 grades, drop out rates are not meaningful, and are not recorded.

schools with their host districts, we attempt to create the local "market" for educational services in which both the charter schools and the public school districts compete. We identify 42 areas, which are mainly urban areas.

A limited number of control variables are included in our empirical models. We have included a few student characteristics: gender, race/ethnicity, and building-level eligibility for free or reduced price lunch. Of the 42 charter schools included in the sample, however, 29 report that no students in their school receive free or reduced price lunches. Therefore, if students are indeed eligible but are not recorded as such, then the results are biased against showing a favorable performance on part of the school.

Table 1 displays building-level data for buildings that housed 5th grade students who took the MEAP tests in 1996/97 in the 42 "host" districts and charter schools. The means give a picture of the differences in student and school characteristics between traditional public schools[22] and charter schools. Charter schools are smaller than traditional public schools, enroll fewer nonwhite students, have smaller classes, pay their teachers less, and have lower per pupil expenditures.[23]

Several models have been estimated to examine the effect of attending a charter school on test score levels. Since we paired the charter school with its "host"[24] public school district, we use fixed effects to control for factors in the district that are common to both types of schools. This approach controls for the average difference in student characteristics across districts, but it does not control for differences in student characteristics between charter schools and public schools within each

[22]We use the term traditional public schools to distinguish them from charter schools, which are also public schools.

[23]In the evaluation, Horn and Miron (1999) report that although many charter schools formed during the first few years targeted minority students, the trend in more recent years has been the opposite. The percentage of white students has risen from 35 percent in 1995 to about 60 percent in 1999.

[24]Again, the word "host" simply means geographic proximity. The host is not the chartering agency.

geographical region. In some models, we include student characteristics to control for these factors, and find that it does not make much difference in the estimated effect of charter schools.

The dependent variables in the analyses of test score levels are the MEAP test scores of fourth and fifth graders for two school years, 1996/97 and 1997/98. Fourth graders take a math and reading test, while fifth graders take a science and writing test. The log of test scores are regressed on three groups of variables. The first set of variables is student characteristics. The percentage of students in each school eligible for free or reduced price lunch program is also included in order to account in some way for differences in socioeconomic factors.

The second set of variables relates to school environment. We include the building-level pupil/teacher ratio and enrollment. The third set of variables includes average teacher salary and expenditure per pupil. Since charter and traditional public schools are located within the same area, there are no cost of living differences across school types. As a result, differences in teacher salaries reflect factors other than cost of living, such as the experience and educational qualifications of the teachers or compensating differentials with respect to work environments. Teachers who prefer the charter school environment may be willing to work there for lower pay. Expenditures per pupil provide a proxy for the amount of school-based resources available at the building level. The coefficients on the variables in these three sets of variables generally have the expected signs and are statistically significant, with and without the fixed effects included. (See tables 2 and 3). However, the explanatory power of the models is highly limited. Even the model with all factors included and the fixed effects explains no more than 10 percent of the variation in the log of the test scores.

The primary focus of the empirical work is to estimate the difference in the level of student test scores between PSAs and traditional public schools, controlling for as many factors as available.

The results displayed in tables 2 and 3 show that in all specifications students attending PSAs have lower test scores than students in traditional public schools. The magnitude of the results vary by grade, year, and subject matter. The coefficient is negative and statistically significant in 30 out of the 40 specifications presented in the tables. The differentials are approximately 2.5 percentage points, 2.0 percentage points, 4.0 percentage points, and 8.0 percentage points for math, reading, science, and writing, respectively. The results are robust across specifications, even when school characteristics, such as class size and expenditures per pupil are included.

The differentials are generally larger for those subjects that arguably are more dependent upon school-based instruction than home-based instruction. For example, the charter school differential for fourth grade math and fifth grade science and writing test scores are statistically significant whereas the differential for reading is not statistically significant under most specifications.[25] One could argue that parental help with reading could mitigate the effects of inferior school-based instruction. Unfortunately, we do not have any measures of home-based activities to control for this effect.

Because we have student date of birth on the file as well as a few other student characteristics, we were able to identify a large number of cases in which we could match the 5th grade test results in 1997/98 to the 4th grade test results in 1996/97. Then we reestimated the models in tables 2 and 3 with previous year test results (in different subjects) as controls.

Specifically, we matched 4th and 5th graders in consecutive years. We matched them by date of birth, gender, race/ethnicity (five categories), and building. We kept only those students that were unique to each cell. That is, we discarded any cell in which we had multiple entrants. Therefore, the

[25]The estimated effects for fourth grade math and fifth grade science and writing are all approximately .40 of a standard deviation.

resulting group of students were in the same building both years. We had nearly a 50 percent match: 13,000/33,000. We ran the same models for the matched sample that we ran for the total sample and got almost identical results (results not reported here). We then ran the same models with the matched group including the log of the appropriate 4th grade test as the pretest in each regression. Fourth grade math was the pretest for 5th grade science; 4th grade reading was the pretest for 5th grade writing. The results are shown in table 4. The pretest is strongly related to the 5th grade test for science, and even more so for writing. The coefficient on PSAs is roughly the same, except that the negative coefficients for science are not statistically significant as they were without the pretest. For writing, the several coefficients on PSA are both negative and statistically significant. These results are consistent with previous results that suggest that students in PSAs do not have higher test scores, nor higher test score gains.

Since charter schools are relatively new, the lower test scores of their students may be explained by the inexperience of staff and the inefficiencies of starting up a new venture. One could also argue that students have not been enrolled long enough in charter schools to make a difference in their performance. However, results (not reported here) show that test scores administered for the same charter schools a year later (i.e., 1997/98 vs. 1996/97) yield results that are similar to those found for the previous year–students in charter schools have lower test score levels than students in traditional public schools.

Analysis of educational management organizations. Some proponents of charter schools argue that school performance would improve if schools would follow business practices more closely. In Michigan, the majority of charter schools are managed by for-profit businesses.[26] We re-

[26]Horn and Miron (1999) note that around 70 percent of Michigan's charter schools are operated by for-profit educational management organizations (EMO's).

estimated the models reported in tables 2 and 3 by replacing the charter school dummy variable with two variables, one indicating whether or not the charter school is managed by a for-profit company and the other indicating whether or not it is self-managed. The two variables are mutually exclusive and together represent the entire group of charter schools that were in the previous analysis. Results shown in table 5 indicate that charter schools managed by for-profit companies have lower test scores relative to traditional public schools than do charter schools not managed by EMO's. The results are not consistent with the position that "market" alternatives necessarily yield better student performance, as measured by test score levels.

Analyses of test score "gains". The use of test scores to compare the performance of schools is problematic, particularly when only a limited number of variables are available to control for student characteristics and home- and school-based resources. As mentioned earlier, test score gains for individual students cannot be computed, because the MEAP tests cover different subjects each consecutive school year and because it is not possible to track individual students across years.

An alternative, albeit inferior, approach is pursued. First, the distribution of student test scores in each building for each subject is calculated for each year. Second, test scores of fourth grade students at the 25^{th}, 50^{th}, and 75^{th} percentiles are subtracted from test scores of fourth grade students at the same percentiles in the subsequent year. The test score gains are obviously not for the same student, but for students that occupied the same place in the test score distribution. Third, the differences, or gains in test scores for students in each building, are regressed against a charter school dummy and other factors. These control variables include the test score in the base year and the difference in between the 75^{th} and 25^{th} percentile test score for each building and for each year. The latter variable helps to control for differences in the spread of the test scores across buildings.

Pupil/teacher ratio for the building, building enrollment, and percentage of students on free and reduced price lunch are also included in the regression.

Two tests, 4th grade math and 5th grade science, are used in estimating the difference in test scores gains between charter schools and traditional public schools. The results are displayed in tables 6 and 7. Test score "gains" are insignificantly different or are lower for charter schools in all specifications but one. The exception is the simplest specification for 4th grade math tests in which only the charter school (PSA) dummy variable is included in the regression. In this case, charter schools have a 1-2 percentage point greater test score gain at each of the three points in the building distributions relative to traditional public schools. However, when control variables are included in the regression, the coefficients turn negative and are not statistically significant except for those in specification D. In that specification, math test scores "gains" for 4th grade students in the 25th and 50th percentiles are 1 to 2 percentage points lower than for students in regular public schools. The difference in test score gains for students in the 75th percentile was not statistically significant.

Results for 5th graders also show lower test score "gains" for students in charter schools. Test score gains were lower for the 50th and 75th percentiles in charter schools in specification D than for those in traditional public schools. Unlike the 4th grade results, in none of the specifications was the coefficient on the PSA variable positive and statistically significant.

The control variables are generally statistically significant and the full regression model explained as much as 42 percent of the variation in the percentage gain in test scores. The percentage gain test score is negatively related to the test score level in the base year, and these coefficients are statistically significant in all cases. Students generally achieved higher test score gains in schools with

larger enrollment, a lower pupil/teacher ratio, and a lower percentage of students on the free or reduced price school lunch program.

The analysis of test scores suggests that charter schools, during their five years of operation in Michigan, have not improved test scores for their students. However, the results presented here on the effectiveness of charter schools are far from conclusive. The test score data present several problems as previously mentioned. Furthermore, it is not possible to construct a random assignment experiment nor to control for selection bias. Yet, if there is a selection bias, it may run in favor of charter schools. If charter schools enrolled students who are academically challenged, then sample selection would be biased against levels and gains in test scores for students in charter schools. However, charter schools enroll a lower percentage of students on the free or reduced price lunch program and a lower percentage of minorities. Furthermore, the active choice by parents to send their children to charter schools would suggest that these students have home support for education, which would suggest a bias in favor of higher levels and gains in test scores.

It is also interesting to note that charter schools run by for-profit businesses, which are presumably subject to even greater market incentives, have lower test score levels than traditional public schools. These levels are typically lower than the test scores of students in non-for-profit charter schools. Charters may have the ability to introduce competition and new innovations in the provision of education, but the evidence so far suggests that charter schools have fallen well short of this promise. Taken at face value, the evidence leads to the conclusion that competition for students; supposedly clear statements of educational goals; adherence by teachers, principals, and parents to these goals; and accountability to the authorizing agency does not seem to improve school performance.

5. School-Based Performance Awards

Using market-based alternatives as a threat to traditional school systems seems not to be effective.[27] The other way to motivate organizations is through shared, group incentives. In the private sector, compensation supplements such as profit sharing or gainsharing are used (Kruse 1993). Also firms use benchmarking to compare their performance along a number of dimensions to key competitors. Firms may pay bonuses if benchmarks are met or exceeded. Nonprofits do not have the residual, profits, to distribute; so they often rely on the benchmarking model. For example, a number of government programs use performance standards based on benchmarks or averages, and reward organizations that exceed those standards. In education, most states have instituted curriculum standards in their educational systems and have implemented standardized testing programs to assess buildings or districts against those standards. Heckman, Heinrich, and Smith (1997) argues that if market alternatives are adequate, then there is no need to rely on a system of performance standards. However, prior evidence and our evidence on the weak performance of market alternatives[28] may suggest that performance standards or other means of school-based performance awards (SBPA's) may be the only viable means of providing incentives or holding schools accountable.[29]

Kelley (1999) describes school-based performance awards in several districts and localities. In an earlier paper, Kelley (1997) examined the Kentucky SBPA, part of an ambitious state-level reform that was instituted in 1992. The Kentucky system relied on the state's biannual test that was

[27]Hoxby (1994) does find that increased choice among public schools seems to be effective.

[28]Evidence does seem to suggest that market alternatives that serve particular groups of students, or particular types of private schools may be effective. See Neal (1997) and Figlio and Stone (1999). The latter concludes that minority students in urban districts may benefit from greater choice.

[29]Note that Heckman et al. (1997) point out difficulties that may result from performance standards. In short, they find that performance standards in a bureaucracy lead to focused effort, not necessarily productive effort.

administered in grades 4-5, 7-8, and 11-12 (the KIRIS). Student scores were classified as novice, apprentice, proficient, and distinguished. The performance target that was set for each school building was a 10 percent increase over a (1992) baseline result. Schools were classified as (1) in crisis–performance lower than baseline by at least 5 percentage points; (2) in decline–performance lower than baseline, but by less than 5 percentage points; (3) improving–performance over baseline, but by less than 10 percentage points; (4) successful–met the target; and (5) reward–exceeded the target and received financial awards. The financial awards were given to the schools, and the teachers decided how to distribute them. Kelley's study found that in reward schools, the curricula had been carefully aligned with the state assessment, the teachers were supportive of the accountability system, and principals were skilled facilitators (rather than strong leaders). In low-performing schools, curricula had not been aligned with the assessment, teachers viewed the test results as more reflective of their student's ability than their teaching skills, and principals tended to be nurturers, rather than skilled facilitators.

Clotfelter and Ladd (1996) evaluates Dallas' performance-based system and find positive effects on student achievement as measured by a standardized test. Using a production function approach similar to the models described above, they find increases on the order of 10 to 12 percent in the pass rate on the math and reading portions on the state-administered Texas Academic Assessment System (TAAS) standardized test in Dallas relative to the other major urban districts in Texas. Since this study did not use a true control group, attribution to the performance-based system is not assured, and in fact, the authors present the puzzling finding that similarly sized impact occurred in the year prior to implementation suggesting that other factors in the Dallas system may be changing relative to the other cities. The authors also point out that gains or levels should be

regression-adjusted for student characteristics in order to be fair, and that excessively large rewards will result in dysfunctional behaviors such as cheating or data manipulation.

6. Summary

In summary, we argue that the nature of schools and the teaching and learning process tend to mitigate against the effectiveness of merit pay (or individual-based incentives). Following a strategy of weakening collective bargaining institutions is also unlikely to improve student achievement, particularly as measured by means or medians. (The evidence does seem to suggest that unionization is detrimental to students at either end of the distribution of achievement.) What about group or organizational incentives? One type of incentive would be to subject traditional public schools to the forces of the market through charter schools or vouchers. However, our empirical evidence on the impact of charter schools on student achievement as measured by a standardized test in Michigan seems to buttress the existing evidence that market-based alternatives to traditional public schools are not directly effective.

Another type of incentive is explicit school building performance awards. These incentives are less subject to the problems associated with subjective evaluations, multi-task job descriptions, and team production. However, it is still the case that there are multiple outcomes and multiple stakeholders and we should expect unintended consequences. Moreover, local circumstances vary considerably. Consequently, we agree with Hanushek and others' (1994) prescription that administrators and policymakers should evaluate and be ready to revise their performance award systems.

Table 1. Means of Selected School Characteristics (standard deviations in parentheses)

Variable	Charter Schools		Traditional Public Schools	
Nonwhite	56.6%	(38.3)	66.1%	(35.4)
Free or Reduced Price Lunch	14.9%	(27.6)	56.7%	(27.5)
Enrollment/building	185	(88)	467	(207)
Pupil/teacher ratio	18.6	(4.7)	25.4	(4.4)
Average teacher salary	$29,985	(7735)	$46,328	(8316)
Ave. expenditure per pupil	$3,720	(1339)	$4,376	(846)
Sample size	34*		538	

*8 of the 42 charter schools in the sample did not have 5th grade.

Table 2. Coefficient Estimates from a Model Explaining 1996/97 MEAP Test Score Levels (*t*-statistics in parentheses)

Subject:	Math					Reading				
Specification:	A	B	C	D	E	A	B	C	D	E
PSA=1	-0.026 (-4.38)	-0.026 (-4.41)	-0.031 (-5.10)	-0.026 (-3.48)	-0.040 (-12.84)	0.001 (0.129)	-.0029 (-0.362)	-0.0055 (-0.67)	-0.0015 (-0.145)	-0.026 (-5.97)
Nonwhite=1		-0.017 (-17.16)	-0.017 (-16.64)	-0.017 (-16.37)	-0.012 (-13.38)		-.0223 (-16.38)	-0.022 (-15.81)	-0.022 (-15.50)	-0.011 (-8.98)
Female=1		0.00002 (0.035)	0.00002 (0.026)	0.00001 (0.015)	-0.0002 (-0.31)		0.020 (22.42)	0.020 (22.14)	0.020 (21.96)	0.019 (20.41)
Free lunch eligibility fraction		-0.033 (-15.86)	-0.033 (-15.45)	-0.032 (-13.77)	-0.036 (-19.07)		-0.05 (-19.18)	-0.06 (-18.97)	-0.06 (-17.68)	-0.05 (-19.88)
Building enrollment			-0.0040 (-3.50)	-0.0038 (-3.11)	0.0064 (6.95)			-0.0014 (-0.90)	-0.0011 (-0.66)	0.020 (15.60)
Pupil/teacher ratio			-0.0042 (-1.52)	-0.0094 (-2.43)	-0.019 (-6.39)			-0.0005 (-0.143)	-0.0035 (-0.66)	-0.013 (-3.26)
Average teacher salary				0.0073 (2.23)	0.0079 (2.96)				0.0040 (0.89)	-0.010 (-2.63)
Expenditure/pupil				-0.0021 (-0.59)	0.0025 (0.96)				-0.0002 (-0.05)	0.010 (2.83)
District fixed effects?	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No
R-square	0.070	0.086	0.087	0.086	0.054	0.042	0.073	0.074	0.074	0.045
Sample size	35,816	35,326	34,956	34,605	34,605	35,816	35,326	34,956	34,605	34,605

Table 2 . (Continued)

Subject:	Science					Writing				
Specification:	A	B	C	D	E	A	B	C	D	E
PSA=1	0.0026 (0.464)	-0.039 (-6.47)	-0.047 (-7.38)	-0.040 (-5.58)	-0.045 (9.26)	0.041 (2.57)	-0.047 (-2.78)	-0.074 (-4.08)	-0.045 (-2.22)	-0.099 (-7.39)
Nonwhite=1		-0.029 (-19.17)	-0.029 (-19.16)	-0.028 (-18.42)	-0.015 (-11.52)		-0.029 (-6.81)	-0.033 (-7.69)	-0.031 (-7.18)	-0.015 (-4.13)
Female=1		0.005 (5.37)	0.005 (5.55)	0.005 (5.50)	0.005 (5.22)		0.077 (27.89)	0.077 (27.87)	0.076 (27.64)	0.075 (27.02)
Free lunch eligibility fraction		-0.42 (-14.41)	-0.05 (-15.10)	-0.04 (-13.16)	-0.05 (20.12)		-0.10 (-13.44)	-0.10 (-13.23)	-0.09 (-9.92)	-0.13 (-17.23)
Building enrollment			-0.004 (-2.06)	-0.0057 (-2.99)	0.019 (13.43)			-0.037 (-7.08)	-0.042 (-7.91)	-0.020 (-5.14)
Pupil/teacher ratio			-0.009 (-2.23)	-0.020 (-3.57)	-0.057 (-13.30)			0.0009 (0.079)	-0.077 (-4.95)	-0.121 (-10.03)
Average teacher salary				0.003 (0.636)	0.0016 (0.43)				0.072 (5.51)	0.094 (8.83)
Expenditure/pupil				-0.022 (-4.32)	-0.012 (-3.09)				-0.091 (-6.37)	-0.041 (-3.75)
District fixed effects?	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No
R-square	0.076	0.094	0.097	0.096	0.062	0.058	0.088	0.093	0.094	0.075
Sample size	31,684	31,497	31,126	30,962	30,962	31,684	31,497	31,126	30,962	30,962

Source: MEAP scores and student characteristics obtained from the Michigan Department of Education and financial data obtained from the Michigan Department of Education. MEAP scores, enrollment, pupil/teacher ratio, average salary, and expenditure/pupil are expressed in logs.

Table 3. Coefficient Estimates from a Model Explaining 1997/98 MEAP Test Score Levels (*t*-statistics in parentheses)

Subject:	Math					Reading				
Specification:	A	B	C	D	E	A	B	C	D	E
PSA=1	-0.016 (-2.95)	-0.019 (-3.49)	-0.025 (-4.63)	-0.017 (-3.02)	-0.030 (-10.09)	0.005 (0.694)	-0.001 (-0.199)	-0.006 (-0.835)	0.003 (0.425)	-0.019 (-4.61)
Nonwhite=1		-0.019 (-20.33)	-0.019 (-20.08)	-0.019 (-19.57)	-0.012 (-14.49)		-0.017 (-13.08)	-0.017 (-12.92)	-0.016 (-12.36)	-0.008 (-6.79)
Female=1		-0.0006 (-0.977)	-0.00097 (-1.57)	-0.0011 (-1.69)	-0.00055 (-0.867)		0.0143 (16.61)	0.0139 (16.13)	0.0137 (15.73)	0.0146 (16.50)
Free lunch eligibility fraction		-0.042 (-21.50)	-0.043 (-21.40)	-0.04 (-18.84)	-0.044 (-27.69)		-0.055 (-20.33)	-0.056 (-20.18)	-0.057 (-19.01)	-0.057 (-25.86)
Building enrollment			-0.011 (-9.84)	-0.010 (-8.61)	0.006 (6.69)			-0.008 (-5.38)	-0.008 (-4.74)	0.013 (10.36)
Pupil/teacher ratio			-0.003 (-1.35)	-0.004 (-1.62)	-0.009 (-3.75)			0.005 (1.41)	0.004 (1.10)	-0.005 (-1.44)
Average teacher salary				0.005 (2.17)	-0.011 (-5.12)				0.011 (3.10)	-0.006 (-1.94)
Expenditure/pupil				-0.001 (-0.507)	0.011 (4.87)				0.003 (0.690)	0.016 (5.17)
District fixed effects?	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No
R-square	0.076	0.104	0.102	0.099	0.059	0.050	0.075	0.072	0.072	0.044
Sample size	34,648	34,435	33,992	33,316	33,316	34,648	34,435	33,992	33,316	33,316

Table 3. (Continued)

Subject:	Science					Writing				
Specification:	A	B	C	D	E	A	B	C	D	E
PSA=1	0.013 (2.66)	-0.034 (-6.61)	-0.036 (-6.71)	-0.030 (-4.87)	-0.030 (-6.58)	0.003 (0.190)	-0.072 (-4.64)	-0.081 (-5.07)	-0.053 (-2.84)	-0.034 (-2.55)
Nonwhite=1		-0.031 (-23.21)	-0.030 (-22.90)	-0.029 (-21.79)	-0.018 (-15.42)		-0.029 (-7.36)	-0.027 (-6.83)	-0.026 (-6.56)	-0.017 (-4.83)
Female=1		0.005 (5.17)	0.005 (5.17)	0.005 (5.19)	0.005 (4.99)		0.091 (34.27)	0.092 (34.23)	0.091 (33.82)	0.092 (33.47)
Free lunch eligibility fraction		-0.053 (-19.25)	-0.054 (-19.32)	-0.054 (-17.96)	-0.067 (-29.30)		-0.092 (-11.32)	-0.092 (-11.05)	-0.082 (-9.21)	-0.17 (-24.73)
Building enrollment			-0.011 (-6.83)	-0.015 (-8.81)	0.014 (11.43)			-0.035 (-7.37)	-0.039 (-7.63)	-0.007 (-1.91)
Pupil/teacher ratio			-0.000 (-0.014)	-0.003 (-0.834)	-0.029 (-8.62)			-0.050 (-4.99)	-0.057 (-5.04)	-0.089 (-9.11)
Average teacher salary				0.008 (2.07)	-0.005 (-1.46)				0.009 (0.852)	0.008 (0.853)
Expenditure/pupil				-0.018 (-5.11)	0.006 (1.79)				-0.041 (-3.92)	0.007 (0.773)
District fixed effects?	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No
R-square	0.091	0.119	0.122	0.124	0.076	0.066	0.104	0.109	0.110	0.084
Sample size	32,363	32,122	31,803	31,192	31,192	32,363	32,122	31,803	31,192	31,192

Source: MEAP scores and student characteristics obtained from the Michigan Department of Education and financial data obtained from the Michigan Department of Education. MEAP scores, enrollment, pupil/teacher ratio, average salary, and expenditure/pupil are expressed in logs.

Table 4. Coefficient Estimates from a Model Explaining 1997/98 MEAP Test Score Levels for 5th Graders Using 4th Grade Tests as a Control (*t*-statistics in parentheses)

Subject:	Science					Writing				
Specification:	A	B	C	D	E	A	B	C	D	E
PSA=1	0.039 (5.78)	-0.005 (-0.752)	-0.006 (-0.783)	0.002 (0.254)	-0.006 (-0.971)	-0.023 (-1.07)	-0.062 (-2.76)	-0.062 (-2.62)	-0.032 (-1.23)	-0.006 (-0.342)
Pretest Scores	0.630 (62.2)	0.601 (59.33)	0.603 (59.34)	0.601 (58.77)	0.618 (60.73)	0.868 (36.45)	0.794 (33.19)	0.804 (33.52)	0.802 (33.21)	0.809 (33.70)
Nonwhite=1		-0.020 (-10.28)	-0.020 (-10.22)	-0.020 (-10.0)	-0.008 (-5.17)		-0.021 (-3.47)	-0.019 (-3.10)	-0.019 (-3.14)	-0.026 (-5.29)
Female=1		0.003 (2.78)	0.003 (2.84)	0.003 (2.72)	0.003 (2.51)		0.068 (17.95)	0.069 (18.04)	0.068 (17.74)	0.069 (17.87)
Free lunch eligibility fraction		-0.0005 (-12.84)	-0.0005 (-13.22)	-0.0005 (-12.23)	-0.0005 (-15.51)		-0.0006 (-4.45)	-0.0006 (-4.50)	-0.0004 (-3.15)	-0.0009 (-9.78)
Building enrollment			-0.013 (-5.94)	0.016 (-6.87)	0.002 (0.958)			-0.039 (-5.72)	-0.042 (-5.96)	-0.050 (-9.39)
Pupil/teacher ratio			0.010 (2.16)	0.008 (1.66)	-0.004 (-0.966)			-0.032 (-2.21)	-0.040 (-2.48)	-0.045 (-3.27)
Average teacher salary				0.0006 (0.114)	-0.011 (-2.73)				0.016 (1.05)	0.030 (2.40)
Expenditure/pupil				-0.010 (-2.13)	0.006 (1.40)				-0.031 (-2.08)	-0.009 (-0.703)
District fixed effects?	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No
R-square	0.292	0.308	0.312	0.313	0.292	0.158	0.181	0.187	0.188	0.176
Sample size	13,077	13,074	12,962	12,804	12,804	12,763	12,761	12,652	12,500	12,500

Source: MEAP scores and student characteristics obtained from the Michigan Department of Education and financial data obtained from the Michigan Department of Education. MEAP scores, enrollment, pupil/teacher ratio, average salary, and expenditure/pupil are expressed in logs. Pretest for science is math; pretest for writing is reading.

Table 5. Coefficient Estimates from a Model Explaining 1996/97 MEAP Test Score Levels that Identifies PSAs Managed by ESOs (*t*-statistics in parentheses)

Subject:	Math			Reading		
Specification:	A	B	C	A	B	C
PSA managed by Business = 1	-0.032 (-5.21)	-0.027 (-3.50)	-0.042 (-11.76)	-0.0096 (-1.14)	-0.0075 (-0.719)	-0.031 (-6.49)
PSA not managed by Business = 1	-0.026 (-3.35)	-0.023 (-2.69)	-0.036 (-6.18)	0.011 (1.01)	0.012 (1.023)	-0.0083 (-1.05)
Nonwhite = 1	-0.017 (-16.66)	-0.017 (-16.38)	-0.012 (-13.40)	-0.022 (-15.87)	-0.022 (-15.57)	-0.011 (-9.05)
Female = 1	0.00002 (0.024)	0.000007 (0.011)	-0.0002 (-0.309)	0.020 (22.14)	0.020 (21.94)	0.019 (20.40)
Free lunch eligibility fraction	-0.03 (-15.48)	-0.03 (-13.73)	-0.04 (-19.07)	-0.06 (-19.13)	-0.06 (-17.82)	-0.05 (-20.04)
Building enrollment	-0.0038 (-3.36)	-0.0037 (-3.05)	0.0065 (7.00)	-0.0010 (-0.616)	-0.0008 (-0.482)	0.020 (15.74)
Pupil/teacher ratio	-0.0040 (-1.47)	-0.009 (-2.40)	-0.019 (-6.29)	-0.00008 (-0.022)	-0.0028 (-0.536)	-0.012 (-3.05)
Average teacher salary		0.0071 (2.16)	0.0077 (2.89)		0.0030 (0.68)	-0.010 (-2.82)
Expenditure/pupil		-0.0021 (-0.615)	0.0025 (0.963)		-0.0006 (-0.137)	0.010 (2.84)
District fixed effects?	Yes	Yes	No	Yes	Yes	No
R-square	0.087	0.086	0.054	0.074	0.074	0.045
Sample size	34,956	34,605	34,605	34,956	34,605	34,605

Table 5. (Continued)

Subject:	Science			Writing		
Specification:	A	B	C	A	B	C
PSA managed by company =1	-0.056 (-8.21)	-0.049 (-6.26)	-0.056 (-9.64)	-0.102 (-5.28)	-0.073 (-3.29)	-0.125 (-7.74)
PSA not managed by company =1	-0.026 (-3.00)	-0.024 (-2.63)	-0.024 (-3.22)	-0.007 (-0.289)	0.0036 (0.14)	-0.053 (-2.50)
Nonwhite =1	-0.029 (-19.11)	-0.028 (-18.44)	-0.015 (-11.56)	-0.033 (-7.65)	-0.031 (-7.20)	-0.015 (-4.16)
Female =1	0.0054 (5.56)	0.0054 (5.51)	0.005 (5.24)	0.077 (27.88)	0.076 (27.65)	0.075 (27.04)
Free lunch eligibility fraction	-0.05 (-15.30)	-0.04 (-13.33)	-0.05 (-20.29)	-0.12 (-13.46)	-0.09 (-10.10)	-0.13 (-17.37)
Building enrollment	-0.0030 (-1.64)	-0.005 (-2.69)	0.019 (13.65)	-0.034 (-6.58)	-0.041 (-7.57)	-0.019 (-4.92)
Pupil/teacher ratio	-0.0074 (-1.90)	-0.019 (-3.44)	-0.056 (-13.09)	0.0047 (0.432)	-0.074 (-4.80)	-0.119 (-9.85)
Average teacher salary		0.0023 (0.49)	0.0012 (0.320)		0.070 (5.35)	0.093 (8.74)
Expenditure/pupil		-0.022 (-4.38)	-0.012 (-3.12)		-0.092 (-6.44)	-0.041 (-3.77)
District fixed effects?	Yes	Yes	No	Yes	Yes	No
R-squared	0.097	0.096	0.063	0.093	0.095	0.075
Sample size	31,126	30,962	30,962	31,126	30,962	30,962

Source: MEAP scores, student characteristics, financial and management data obtained from the Michigan Department of Education. MEAP scores, enrollment, pupil/teacher ratio, average salary, and expenditure/pupil are expressed in logs.

Table 6. Coefficient Estimates from a Model Explaining Differences in MEAP Mathematics Test Score Levels at Distributional Quartiles (*t*-statistics in parentheses)

25th Percentile

Specification:	A	B	C	D
PSA=1	0.019 (2.15)	-0.003 (-0.444)	-0.0008 (-0.117)	-0.021 (-2.62)
Test score in 1996/97		-0.0013 (-16.37)	-0.0011 (-12.40)	-0.0012 (-13.87)
75th percentile test score minus 25th percentile test score in 1996/97			0.0010 (7.75)	0.0008 (6.47)
Pupil/teacher ratio 1996/97				-0.026 (-2.83)
Building enrollment in 1996/97				0.0009 (0.254)
Free or reduced price lunch eligibility fraction in 1996/97				-0.03 (-4.09)
R-square	0.006	0.33	0.39	0.42

Table 6. (Continued)

50th Percentile

Specification:	A	B	C	D
PSA=1	0.017 (2.56)	-0.0007 (-0.126)	-0.0009 (-0.166)	-0.012 (-1.94)
Test score in 1996/97		-0.0010 (-16.28)	-0.0010 (-15.76)	-0.0011 (-16.84)
75th percentile test score minus 25th percentile test score in 1996/97			0.0004 (4.45)	0.0003 (3.88)
Pupil/teacher ratio 1996/97				-0.017 (-2.43)
Building enrollment in 1996/97				0.0038 (1.40)
Free or reduced price lunch eligibility fraction in 1996/97				-0.02 (-4.05)
R-square	0.011	0.33	0.35	0.38

Table 6. (Continued)

75th Percentile

Specification:	A	B	C	D
PSA=1	0.012 (1.87)	-0.0017 (-0.322)	-0.0005 (-0.098)	-0.008 (-1.27)
Test score in 1996/97		-0.0009 (-14.60)	-0.0009 (-13.44)	-0.0010 (-14.20)
75th percentile test score minus 25th percentile test score in 1996/97			-0.0002 (-2.68)	-0.0002 (-2.51)
Pupil/teacher ratio 1996/97				-0.012 (-1.77)
Building enrollment in 1996/97				0.005 (1.77)
Free or reduced price lunch eligibility fraction in 1996/97				-0.02 (-3.55)
R-square	0.004	0.28	0.29	0.31

Note: Difference is equal to 1997/98 level minus 1996/97 level.

Source: MEAP scores, student characteristics, financial and management data obtained from the Michigan Department of Education. MEAP scores, enrollment, pupil/teacher ratio, average salary, and expenditure/pupil are expressed in logs. Sample size = 537.

Table 7. Coefficient Estimates from a Model Explaining Differences in MEAP Science Test Score Levels at Distributional Quartiles (*t*-statistics in parentheses)

25th Percentile

Specification:	A	B	C	D
PSA=1	0.068 (0.666)	-0.0004 (-0.039)	-0.0000 (0.000)	-0.004 (-0.402)
Test score in 1996/97		-0.0004 (-10.83)	-0.0011 (-9.20)	-0.0012 (-9.77)
75th percentile test score minus 25th percentile test score in 1996/97			0.0005 (2.13)	0.0005 (1.98)
Pupil/teacher ratio in 1996/97				-0.010 (-0.758)
Building enrollment in 1996/97				0.012 (2.24)
Free or reduced price lunch eligibility fraction in 1996/97				-0.03 (-3.14)
R-square	-0.001	0.17	0.17	0.19

Table 7. (Continued)

50th Percentile

Specification:	A	B	C	D
PSA=1	-0.005 (-0.634)	-0.009 (-1.43)	-0.009 (-1.43)	-0.020 (-2.56)
Test score in 1996/97		-0.0008 (-10.73)	-0.0009 (-10.81)	-0.001 (-11.93)
75th percentile test score minus 25th percentile test score in 1996/97			-0.0002 (-1.24)	-0.0002 (-0.946)
Pupil/teacher ratio in 1996/97				-0.018 (-2.05)
Building enrollment in 1996/97				0.007 (1.88)
Free or reduced price lunch eligibility fraction in 1996/97				-0.03 (-3.89)
R-square	-0.001	0.17	0.17	0.19

Table 7. (Continued)

75th Percentile

Specification:	A	B	C	D
PSA=1	-0.0029 (-0.421)	-0.008 (-1.20)	-0.007 (-1.15)	-0.019 (-2.67)
Test score in 1996/97		-0.0008 (-10.40)	-0.0008 (-10.11)	-0.001 (-11.79)
75th percentile test score minus 25th percentile test score in 1996/97			-0.0005 (-3.47)	-0.0004 (-2.66)
Pupil/teacher ratio in 1996/97				-0.018 (-2.14)
Building enrollment in 1996/97				0.0073 (2.12)
Free or reduced price lunch eligibility fraction in 1996/97				-0.03 (-5.00)
R-square	-0.001	0.16	0.17	0.22

Note: Difference is equal to 1997/98 level minus 1996/97 level.

Source: MEAP scores, student characteristics, financial and management data obtained from the Michigan Department of Education. MEAP scores, enrollment, pupil/teacher ratio, average salary, and expenditure/pupil are expressed in logs. Sample size = 572.

References

Argys, Laura M. and Daniel I. Rees (1995). "Unionization and School Productivity: A Reexamination." in *Research in Labor Economics* 14: 49-68.

Ballou, Dale and Michael Podgursky. (1997). *Teacher Pay and Teacher Quality*. Kalamazoo, MI: W. E. Upjohn Institute for Employment Research.

Clotfelter, Charles T. and Helen F. Ladd. (1996) "Recognizing and Rewarding Success in Public Schools." *Holding Schools Accountable* . Washington, DC: The Brookings Institute.

Eberts, Randall W. and Joe A. Stone. (1991) "Unionization and Cost of Production: Compensation, Productivity, and Factor-Use Effects." *Journal of Labor Economics* 9(2): 171-185.

—. (1987) "Teachers' Unions and the Productivity of Public Schools." *Industrial and Labor Relations Review* 40:355-63.

—. (1986) "Teacher Unions and the Cost of Public Education." *Economic Inquiry* 24:631-44.

—. (1984) *Unions and Public Schools: The Effect of Collective Bargaining on American Education*. Lexington, MA: Lexington Books.

Figlio, David N. and Joe A. Stone. (1999) "Are Private Schools Really Better?" Forthcoming in *Handbook of Labor Economics*, edited by Solomon Polachek.

Firestone, William A. (1994) "Redesigning Teacher Salary Systems for Educational Reform." *American Educational Research Journal* 31(3): 549-574.

Greene, Jay P., Paul E. Peterson, and Jiangtao Du. (1997) "The Effectiveness of School Choice: The Milwaukee Experiment." Harvard University Education Policy and Governance Occasional Paper 97-1 (March).

Grimes, Paul W. and Charles A. Register. (1990) "Teachers' Unions and Student Achievement in High School Economics." *The Journal of Economic Education* 21:297-308.

Hanushek, Eric A. with others. (1994) *Making Schools Work: Improving Performance and Controlling Costs*. Washington, D.C.: The Brookings Institution.

Hanushek, Eric A. (1986) "The Impact of Differential Expenditures on School Performance." *Educational Researcher* 18(4): 45-62.

Hassel, Bryan C. (1999) *The Charter School Challenge: Avoiding the Pitfalls, Fulfilling the Promise*. Washington, DC: Brookings Institution Press.

Hatry, Harry P., John M. Greiner, and Brenda G. Ashford. (1994) *Issues and Case Studies in Teacher Incentive Plans*, 2nd edition. Washington, DC: The Urban Institute Press.

Heckman, James, Carolyn Heinrich, and Jeffrey Smith. (1997) "Assessing the Performance of performance Standards in Public Bureaucracies." *AEA Papers and Proceedings* 87(2): 389-395.

Hedges, Larry V., Richard D. Laine and Rob Greenwald (1994) "Does Money Matter? A Meta-Analysis of Studies of the Effects of Differential School Inputs on Student Outcomes." *Educational Researcher* 23(3): 5-14.

Horn, Jerry and Gary Miron. (1999) "Evaluation of the Michigan Public School Academy Initiative." Final report. The Evaluation Center, Western Michigan University (January).

Hoxby, Caroline Minter. (1996) "How Teachers' Unions Affect Education Production." *Quarterly Journal of Economics* 111:671-718.

—. (1994) "Does Competition Among Public Schools Benefit Students and Taxpayers." NBER Working Paper No. 4979. Cambridge, MA: National Bureau of Economic Research.

Johnson, Susan Moore. (1986). "Incentives for Teachers: What Motivates, What Matters." *Educational Administration Quarterly* 22(3): 54-79.

Kelley, Carolyn. (1999) "The Motivational Impact of School-Based Performance Awards." *Journal of Personnel Evaluation in Education* 12(4): 309-326.

—. (1997, March) "The Kentucky School-Based Performance Award Program: School-Level Effects." Paper presented at the meeting of the American Educational Research Association Annual Conference, Chicago, IL.

Kleiner, Morris M. and Daniel L. Petree. (1988) "Unionism and Licensing of Public School Teachers: Impact on Wages and Educational Output." in *When Public Sector Workers Unionize*. Richard B. Freeman and Casey Ichniowski, ed. Chicago: University of Chicago Press.

Kruse, Douglas L. (1993) *Profit Sharing: Does It Make A Difference?* Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

Ladd, Helen F., ed. (1996) *Holding Schools Accountable: Performance-Based Reform in Education*. Washington, D.C.: The Brookings Institution.

Lawler, E. E., III . (1983). *Pay and Organization Development*. Reading, MA: Addison-Wesley.

Milgrom, Paul and John Roberts. (1988) "An Economic Approach to Influence in Organizations." *American Journal of Sociology* 94: S154-79.

Milkman, Martin I. (1997) "Teachers' Unions, Productivity, and Minority Student Achievement." *Journal of Labor Research* 18:137-50.

—. (1989) *Teacher Unions and High School Productivity*. Ph.D. Dissertation. University of Oregon.

Murnane, Richard J. and David K. Cohen. (1986) "Merit Pay and the Evaluation Problem: Why Most Merit Pay Plans Fail and a Few Survive." *Harvard Educational Review* 56(1): 1-17.

Nathan, Joe. (1999) *Charter Schools: Creating Hope and Opportunity for American Education*. San Francisco: Jossey-Bass Publishers.

Neal, Derek. (1997) "The Effects of Catholic Secondary Schooling on Educational Achievement." *Journal of Labor Economics* 15: 98-123.

Nelson, F. Howard and Michael Rosen. (1996) "Are Teachers' Unions Hurting American Education? A State-by-State Analysis of the Impact of Collective Bargaining among Teachers on Student Performance." Technical Report. Institute for Wisconsin's Future. Milwaukee, WI.

Prendergast, Candice. (1999) "The Provision of Incentives in Firms." *Journal of Economic Literature* 37: 7-63.

Rothstein, Richard, Martin Carnoy and Luis Benveniste. (1999) *Can Public Schools Learn From Private Schools: Case Studies in the Public and Private Nonprofit Sectors*. Washington, DC: Economic Policy Institute.

Rouse, Cecilia Elena. (1998) "Private School Vouchers and Student Achievement: An Evaluation of the Milwaukee Parental Choice Program." *The Quarterly Journal of Economics* (May): 553-602.

Stone, Joe A. (1998) "Collective Bargaining and Public Schools: A Critical Survey of the Evidence," Presented at the Conference on Teacher Unions and Educational Change, September 23-25, 1998, sponsored by the Program on Educational Policy and Governance at the Taubman Center for State and Local Government, Kennedy School of Government.

Tirole, Jean. (1992) "Collusion and the Theory of Organizations." In *Advances in Economic Theory: Sixth World Congress*, Vol. 2, edited by J. Laffont. Cambridge: Cambridge University Press.

Tulli, Dennis J. (1991) *An Assessment of Student Achievement Before and During a Merit Pay Program for Teachers of the Penn Manor School District*. Ed.D. dissertation, Temple University.

Witte, John F. (1997) "Achievement Effects of the Milwaukee Voucher Program." University of Wisconsin at Madison mimeo, January.