SWEDISH ECONOMIC POLICY REVIEW 10 (2003) 179-214

Teacher incentives

Edward P. Lazear*

Summary

■ Like all workers, teachers may be expected to respond to incentives inherent in compensation structures. As such, general theories of compensation should apply to teaching. Those theories suggest that output-based pay is best used when output is well defined and easily measured. Input-based pay is best when jobs are inherently risky and when output is not easily observed. The main difficulty with output-based pay is that even if teachers can affect their students' earnings, the evidence does not show up until many years after the student leaves the teacher's class. The National Education Longitudinal Study of 1988 provides evidence that earnings later in life are related to test scores when children are as young as twelve years old.

Actual compensation structures in both the US and Sweden are examined. The main features for both countries include relatively low pay, the reluctance to tie compensation to output and a pattern pay compression, both across fields and by teacher performance. Low pay makes it difficult to attract a large enough quantity of high quality teachers. Compression results in some adverse selection, where the highest quality teachers may be induced to leave the profession.

A primary reason to increase teacher pay and to tie it to performance is that teacher quality would be improved by such policies. Finally, teacher and student preferences may deviate from the optimum. This occurs primarily because of the failure to price working conditions and school inputs appropriately. As a result, teacher, student and parent decisions may be distorted. ■

JEL classification: M5, I2.
Keywords: Incentives, education, teacher pay.

* *Edward P. Lazear is Professor at the Hoover Institution and the Graduate School of Business, Stanford University*

Teacher incentives

Edward P. Lazear*

Education has been shown to be key to economic growth[1] and to the determination of earnings of individuals within a society.[2] Because it affects society so directly and dramatically, the education industry is often singled out for scrutiny at all levels. Coupled with the fact that most education is public, it is frequently claimed that performance in education is hampered by a failure to provide appropriate incentives.

Some incentive problems involve product market competition. Because education is state run, the "firms," in this case schools or school districts, do not face as much market pressure as firms in the private sector. It has been argued that the failure to allow competition to work hinders productivity and raises costs.[3] More directly, it has also been argued that teacher (and school administrator) incentives have been distorted, perhaps as a result of the same failure to face competition in the market. The more specific focus of teacher incentives is the subject of this essay.

In particular, three questions are addressed. First, what are the principles behind creating optimal teacher incentives, and how close do the actual structures in Sweden and the US conform to the ideal ones? Second, how much is performance affected by creating incentives for current teachers, and how much by changing the pool of teacher applicants? Third, do teacher preferences align with those of their students and of society in general, and if not, why not? Associated with each of these questions are policy implications that may remedy existing distortions.

* *This research was supported in part by the National Science Foundation. Comments by Peter Fredriksson and Bertil Holmlund are gratefully acknowledged.*

[1] See Barro (2001), Hall and Jones (1999).

[2] There is an enormous literature on this that stems from the work of Schultz (1961), Mincer (1962), and Becker (1962).

[3] See Hoxby (1996, 2003).

1. Theory of compensation

The guidelines for thinking about the theory of compensation come from a comparison of payment on input versus payment output.[4] One common criticism of the teacher incentive problem is that compensation is not sufficiently "results oriented." Most detractors contend that a reward structure that ties pay to student performance would remedy many of the problems. Would this happen, and is it feasible were it not for the opposition of teacher organizations?

In the context of the teaching profession, payment on input means payment on the basis of skills and time worked, whereas payment on output usually refers to some metric of the performance of the students whom they teach. When feasible, payment on output has two advantages over payment on input. The first, most frequently emphasized, is incentives. The second, equally important, but receiving less attention, is sorting or selection.

The incentive argument is relatively straightforward. If evaluators can agree on the appropriate metric of performance, then rewarding teachers on the basis of the metric aligns teacher incentives with those of their students and potentially of society as a whole.[5] It is important to align incentives for a variety of reasons, most of which have nothing to do with selfish behavior by teachers. Part of the argument is informational; if wages are contingent on student performance, then payment on the basis of student performance provides forceful signals to teachers about what is valued and what is not. Part depends on alignment of agendas. Teachers may be very hard-working in general, but individual teachers may believe that some things are more important than others in the curriculum. A teacher with idiosyncratic views might provide the wrong kind of education to his or her students. Tying compensation to the appropriate metric provides incentives to move in the direction that has been agreed on. Finally, there is always some tension between a teacher's own preferences and those of his students. For example, as a teacher myself, I might fail to assign term papers because the grading of them is onerous, even though I recog-

[4] See Lazear (1986, 1995, 1998).

[5] The argument for social efficiency is slightly more complex. If there are positive externalities to education that accrue to those in society other than the student, then the social return to education is higher than the private return. Rewards could be associated with the social return, inducing teachers to take into account the full value of their effort in making the work choice decisions.

nize that the exercise would be of value to my students. Compensation based on performance provides some incentive for me to do the right thing.

There are two major problems with the incentive argument. First, it presumes that we can agree on what is desirable in the absence of a strong competitive sector. Second, it assumes that we can measure it accurately.

The problem of assigning value is particularly acute in education because we do not have much competition for state-run schools at the primary and secondary levels. In most of business, the market disciplines firms into providing products that consumers value. But schooling is both compulsory and to a major degree provided by the state. As such, students who do not like the offerings of the publicly run schools find it difficult to vote with their feet. For the most part, they must buy the product from the state, like it or not. Of course, parents and students have voice and are able to express any discontent by talking to teachers and school administrators. But the most effective form of discipline, namely boycotting the product, is almost absent.

1.1. Output-based pay

Because there is no strong market competition, schools must set their standards and hope that they get them right. What is the goal of education? Although there may be many goals, the public's most immediate concern in educating its children is providing the skills necessary to ensure a productive populace. Without job market skills, individuals cannot generate earnings necessary to sustain themselves and must rely on the general support of society. The initial goal of compulsory education is basic literacy, but most developed societies now demand far more than that from their schools. They want populations that can produce output sufficiently high to create economic well-being and growth. The bottom line, then, is that generating earnings or at least earnings capacity is an, if not the most important goal of education.

The problem is that the relevant earnings do not show up until many years after the individual has received the education. Therefore, it is impossible to tie teacher compensation to the earnings of her students, even if that is the relevant metric. Instead, proxies are used. The proxies are usually achievement test scores. Many educators believe that higher test scores are themselves a goal of education, but most think of them as proxies for other things. Literacy, both verbal and mathematical, educational attainment (itself an intermediate factor), and subsequent economic outcome are the most obvious. Indeed, there has been some research[6] that shows that test scores are eventually related to earnings. Below, additional original evidence on the connection between test scores and earnings is provided.

Continuous or discrete standard

Even if it is agreed that pay should be based on measures of output, it remains to be determined whether continuous or discrete standards should be used and where should the target be set. This depends on a number of factors, many of which have both efficiency and equity consequences. For example, the current "No Child Left Behind" program at the US federal level rewards schools on the basis of the number of students who attain a certain level of literacy as determined by an annual examination. This is discontinuous in two respects. First, the test is treated discontinously at the level of the student who either passes or fails the test. Second, a school is singled out for disciplinary attention if it fails to make "adequate yearly progress." Making adequate yearly progress depends on meeting some criterion, which either happens or does not. An alternative structure, closer to piece rates paid in manufacturing, would be simply to reward or punish on the basis of average test scores or changes in those averages. Probably the largest difference between this method and the current one is that it gives credit to gains at all parts of the achievement distribution rather than to those close to the cutoff. For example, those in the bottom 5 percent of the achievement distribution might be too far away from the literacy standard to pass the test after one or two year's time, but raising their test scores might be worthwhile nonetheless. The discontinuities are less important than the fact that paying on the basis of averages weight more of the distribution than paying on the basis of exceeding some cutoff. A hybrid is also possible. One could create a metric that was a formula and gave different weights to improving test scores at different parts of the achievement distribution. It is possible to weight the bottom more than the top, the middle more than either end, and so forth. Furthermore, there is a conceptual basis on which to anchor the weighting function. In theory, point increases at different parts of the achievement distribution correspond to the effect of the point increase on the relevant dependent variable.

[6] See Neal (1997) and Neal and Johnson (1996), for example.

If that were determined to be wage increases, it might be that a point gain at the bottom of the ability distribution affected earnings by twice as much as an increase at the top of the distribution. If so, one could weight gains at the bottom of distribution twice as heavily as the top. The formal definition of the index is found in Appendix A.

The pay structure takes on a very simple form when wealth rises linearly with scores on the achievement test. If scores can be re-scaled so that the mapping of the transformed scores into wealth can be made linear, then the optimal payment function is

$$\text{Teachers salary} = \mathrm{N}\left(\Delta Score\right)\beta - K \tag{1}$$

where $\Delta Score$ is the change in the class average test score as that results from teacher intervention, β is a number that comes from the data that relate earnings to test scores, and K is the constant that sets salaries at their appropriate level. This formula is very easy because it says that the school need only take the increase in average test scores that results from the teacher's activity and multiply it by some scalar to get the appropriate increase in salary.

For example, suppose that the form that fits the data best (a common specification in the labor earnings literature) is

$$\log\left(\text{Wealth}\right) = \lambda + \gamma A. \tag{2}$$

Then, define a transformed test score measure, $A*$, such that

$$A* = \lambda e^{\gamma A} / \theta \tag{3}$$

where θ is an arbitrary scalar. The transformed test scores can then be compensated according to the increase of the mean of $A*$.

The complications come in when we seek to determine the size of β and when we try to determine how much of the increase in test scores was due to teacher effort versus the activity of others. While these issues present technical challenges, the derivation here shows that it is possible to condition teacher salary on performance to produce the correct incentives.

There are distributional consequences in addition to efficiency effects. Suppose, for example, that increases at the top of the achievement distribution were more (rather than less) productive in raising wages. Society might still prefer to weight the bottom more heavily than the top to encourage teachers to focus on the bottom where earnings gains, small though they may be, were deemed to be more important.

Sorting or incentives?

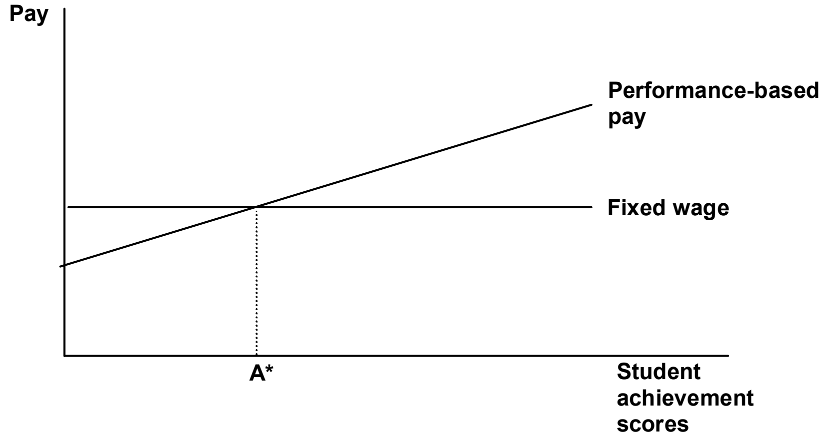
Most of the discussion on performance or output-based pay centers on incentives. Relatively little attention is given to the nature of people that various compensation schemes attract, but selection is equally important and in the context of education, perhaps the most important aspect of compensation effects.

The distinction between sorting and incentives is this: Incentive effects are those that are associated with a given individual. If person A is confronted with a different compensation scheme, say, one that pays on the basis of test scores, the incentive effect is that A will attempt to raise test scores to enhance compensation. Sorting works differently. Even if teachers are unable to alter their behavior to enhance test scores, some people are inherently better at affecting test scores than others. When a school shifts from paying a straight annual salary to paying on the basis of test scores, those who are best at getting high test scores out of their students benefit relative to those who are not able to affect test scores. As a result, payment based on output tends to attract those who can produce output and to discourage those who cannot. In sorting or selection, individual B then replaces individual A. When incentives are the key force, A is not replaced, but merely changes her behavior.

The point is easily seen in Figure 1. Consider two schedules. The first pays a fixed wage and is flat, i.e., independent of student achievement scores. The second pays a base wage, but increases pay based on student performance. Those teachers who can raise student achievement scores above A^* earn more on the performance-pay schedule. Those who cannot raise achievement scores above A^* earn more on the fixed wage schedule. The point is that pay based on student achievement scores favors higher ability teachers relative to lower ability ones, where ability is defined in terms of the teacher's ability to raise achievement scores.[7]

[7] The point does not require that those most able to elicit high test scores necessarily prefer performance-based to fixed pay. Rather, those who find it most difficult

Figure 1.



The distortions that arise in the incentive context are similarly relevant when sorting is the mechanism of interest. Suppose that what we really care about is subsequent earnings and that test scores are correlated with, but do not cause higher earnings. Performance pay based on test scores causes a distortion when incentives are involved because teachers are motivated to raise test scores, but not to raise subsequent earnings. Similarly, a distortion in hiring can result under the same circumstances. If there are teachers who are good in raising test scores, but not in raising earnings, then pay based on achievement test scores attracts those who are capable at the former without getting those who excel on the latter.[8]

In some recent work[9] that examined a completely different (and perhaps irrelevant) industry, the sorting effect was found to be as large as the incentive effect. That is, when the firm in question switched its compensation scheme, there was a 22 percent increase in productivity as a result of the incentive effect and one of almost exactly the same size as a result of the change in personnel. Those who replaced the former workers who left the job (mostly through quits) were on average 24 percent more productive than the workers they

to improve the achievement scores of their students are at a comparative disadvantage when performance-based pay is used.

[8] The situation is somewhat less problematic, however, if those who are good at raising test scores are also good at raising wages. Then, paying on test scores attracts the right people to the job.

[9] Lazear (2000).

replaced. Again, there is no reason to expect that the same results would prevail in education, but the size and relative importance of the two effects are noteworthy.

In the context of education, there is some new evidence that incentive pay can produce desirable outcomes. Lavy (2002) reports that in an experiment run in Israel, where teachers and schools are rewarded according to a tournament-like structure, the reward structure induced better performance among teachers. This was particularly true of the rewards that were teacher- rather than school-based.

Team incentives

Some observers believe that education, and in particular education in Sweden, is moving more toward team production, where teachers share classes, engage in communication, and run educational projects together.

The most standard way to motivate a team is to compensate the team for group output. This works reasonably well when the team size is small, say, fewer than ten individuals. When teams become large, free rider effects of the type analyzed in Kandel and Lazear (1992) become pronounced. The problem in large groups is that each individual captures only 1/N of the return to his or her effort, and as N gets large, the dilution effects become too pronounced. Indeed, the evidence discussed above provided by Lavy bears on the point. Rewards that were school- based were less effective in raising student performance than those that were individual-teacher based. If the team cooperation effect induced by school compensation outweighed any free rider effects, then school-based award should have been more effective than individual awards. Lavy finds the reverse. In another context (that of medical practices), Encinosa, et. al. (1998) find that larger medical groups use more individual incentive pay, presumably to offset free rider effects associated with group pay.

Teams incentives should be used only when production is truly joint. There are almost no work situations in which production is individual, so it is a question of degree. As educational technology becomes more accommodating of team work, team compensation should grow in importance. But as of right now, there is little evidence that team compensation dominated individual compensation for motivating teachers.

The link between earnings and test scores: new estimates.

Most of the emphasis on performance pay revolves around payment on the basis of test scores. But how closely are test scores linked to the relevant dependent variable? There is little doubt that programs like "No Child Left Behind" in the US have as the target increasing the earning power of currently disadvantaged individuals. Are test scores early in a child's learning a good proxy for earning power later in life? This question is difficult to answer because it is necessary to obtain data both on early test scores and then on earnings many years later. Fortunately, a data set has recently become available that allows such an analysis to be done.

The National Education Longitudinal Study (NELS) of 1988 provides a great deal of information on tests taken as early as eighth grade. There is also follow-up in subsequent years with earnings reported. The survey was done in 1988 so the oldest individuals in the sample are only about 27 years old in 1999 (which was the last follow-up wave) with the mean being 25. Still, income in the mid-20s provides some information on lifetime income.[10] Using these data, it is possible to estimate the effect of test scores on lifetime wealth or at least on income at early ages. It is also possible to examine the relation of educational attainment to early test scores.

Table 1 reports the means of the relevant variables and regression results are contained in Table 2. The basic specification and most fundamental results are contained in column 1. There, "SCORE," which is defined as the sum of the standardized scores on reading, math, history, and science tests given at the outset of the survey, is entered to control for aptitude going into the period. Then ΔSCORE is the change in the score between eighth grade and the follow-up four years later, for most, at the end of high school. Since the scores are standardized, there is no reason to expect a positive change in scores. As seen in Table 1, the mean of ΔSCORE is about 4 points on an average SCORE level of 202.8.

[10] The major problem is that those who will have very high earnings later on may show up as having relatively low earnings during their 20s because they are investing in additional skills.

Table 1.

Variable	Mean	Standard deviation
SCORE	202.80	49.60
ΔSCORE	4.36	39.40
INCOME	$23,609	$20,209
LN(INCOME)	9.97	0.77
AGE at fourth follow up	25.30	0.54
Dummy Parent College	0.29	
WHITE	0.69	
Black	0.08	
Hispanic	0.11	
Education level	13.10	2.60

The basic evidence is contained in regression column 1 of Table 2. Note that both the coefficients on SCORE and ΔSCORE are both statistically significant and virtually identical. One interpretation of the ΔSCORE variable is that of learning. Higher increases in ΔSCORE imply that more learning has occurred. When SCORE is increased between eighth and twelfth grades, there is a direct positive effect on earnings. For every one-tenth of a standard deviation increase in SCORE during the first and third wave of the test, there is an increase in earnings of about 1.2 percent per year.

One major problem is this: It is very possible that those who have high test scores when young are more able than those with low test scores. If subsequent income is related to underlying ability, then there will be a correlation between test scores and income, even if there is no causal relation. That is, the correlation might be there, but raising test scores might not necessarily raise subsequent income and it is the causal relation that is of interest.[11]

Although there is the worry that the independent variables are picking up only unobserved ability differences across individuals, this seems unlikely for a variety of reasons. First, there is some evidence that the coefficient on SCORE itself reflects the effect of prior learning rather than unobserved ability. Recall that the coefficient on SCORE is identical to that on ΔSCORE. Thus, a point of SCORE acquired before the first test has the same effect on earnings as a

[11] Still, because the direction of the bias is clear, we can think of the obtained estimates as being an upper bound to the causal effect.

point acquired after the first test.[12] Were the variables picking up unobserved ability, there would be no reason to expect that the coefficient on the unobserved ability component and the ΔSCORE variable would be the same. The fact that they are the same suggests that they are measuring precisely the same thing. Were that same thing unobserved ability, one variable would likely capture most of the effect. A reasonable candidate is that they both measure learning and whether it comes before the first test or between the first and third test is inconsequential.

Table 2. Regression results: Earnings, education and test scores

	1	2	3	4	5	6
Variable	**ln(income)**	**ln(income)**	**ln(income)**	**Education level**	**Education level**	**Education level**
SCORE	.00318 (.00003)	.00145 (.00002)	.00296 (.00032)	.031 (.001)	.0166 (.0006)	.013 (.001)
ΔSCORE	.00317 (.00004)		.00287 (.00053)	.025 (.001)		.007 (.001)
Dummy (parent went to college)[a]			.044 (.023)			3.43 (0.05)
Age at first survey			-.002 (.018)			-.171 (.042)
$(\Delta SCORE)^2$			.00000007 (.00000027)			.000022 (.000006)
Intercept	9.31 (0.06)	9.68 (0.04)	9.39 (0.48)	6.90 (.18)	9.85 (.13)	13.8 (1.1)
R-square	.02	.01	.02	.17	.10	.52

Notes: Standard errors in parentheses. [a] "Parent" refers to parent who answered the parent questionnaire. Number of observations = 6278.

Perhaps SCORE and ΔSCORE both measure unobserved ability in exactly the same way and unobserved ability loads on each factor identically. Although possible, it seems less likely conceptually. Regression column 2 provides evidence. If ΔSCORE were picking up unobserved ability, then presumably that unobserved component would be positively correlated with the unobserved component

[12] Another way to put this is that SCORE3, defined as the sum of test scores on the 3rd wave of testing (done for most at the end of high school), captures all of the effect. Rather than putting in SCORE and ΔSCORE, the regression could have had only SCORE3 on the r.h.s., since SCORE3=SCORE+ΔSCORE. If the coefficients on SCORE and ΔSCORE are the same, then the one variable specification is identical to entering SCORE and ΔSCORE.

picked up by SCORE. Omitting ΔSCORE from the regression should result in omitted variable bias in the coefficient on SCORE. Under the interpretation that ΔSCORE is unobserved ability, its effect on ln(Income) should be positive. Furthermore, since ΔSCORE and SCORE are presumed to measure the same thing, the regression coefficient of ΔSCORE on SCORE should be positive. If so, the coefficient on SCORE should be higher in column 2 than it is in column 1. In fact, it is lower. Indeed, a supplementary regression reveals that

$$\Delta\text{SCORE} = 115 - \underset{(.01)}{.54}\ \text{SCORE}. \tag{4}$$

An alternative interpretation and one that favors a learning interpretation over unobserved ability is that those who have learned the material before the first wave are less likely to learn it between the first and third wave. The differences between SCORE and ΔSCORE would then relate to the timing of the learning rather than to unobserved ability.

Additional evidence is provided in column 3 of Table 2. A dummy for the parent's (who answers the parent questionnaire) having attended college is entered, as is the respondent's age in 1999. Both can be thought to be proxies for underlying ability since age is the age at the fourth follow-up, which is equal to age in eighth grade plus 11. The higher ages then refer to children who were older in eighth grade and are therefore likely to have been slower to reach that grade. Age does not enter significantly. Parent's education does affect subsequent earnings. Having parents who completed college raises earnings by about 4.5 percent per year. This could reflect ability differences or human capital that educated parents impart to their children. There is some decline in the coefficients on SCORE and ΔSCORE, but they are not statistically different from those obtained in column 1.

The same regression also allows for a non-linear ΔSCORE term. The variable does not enter significantly. Finally, it is not possible within the limited ages of the participants to determine the effect of SCORE and ΔSCORE on earnings at different ages. There is not enough age variation in the data, either cross-sectionally or across waves, to estimate any significant differences in effects at different ages in the labor force.

It is interesting to examine the effect of test scores on another dependent variable, namely the level of attained education. The dependent variable in columns 4 through 6 of Table 2 is education measured in years of schooling. As seen in column 6, virtually all variables considered affect attained level of education. In particular, SCORE and ΔSCORE have large (although different size) effects on schooling attained. Furthermore, having a parent who attended college increases the average level of education by about 3.4 years, a very large effect.

The same effects are estimated in Table 3 for racial subgroups. All coefficients have the same positive (and significant) signs found in Table 2. The only notable difference is that the effect of raising test scores is more than twice as large for blacks as for whites, with Hispanics in the middle.

Table 3. Regression results by race and schooling completion

	Whites: ln(income)	Blacks: ln(income)	Hispanic: ln(income)	No post-HS: ln(income)	Some post-HS: ln(income)
SCORE	.00257 (.00034)	.00560 (.00094)	.00412 (.00093)	.00387 (.00046)	.00255 (.00040)
ΔSCORE	.00272 (.00043)	.00610 (.00121)	.00378 (.00110)	.00446 (.00054)	.00196 (.00052)
Intercept	9.46 (0.07)	8.82 (0.18)	9.08 (0.18)	9.16 (0.09)	9.46 (0.09)
R-square	.01	.07	.03	.03	.01
Number of observations	4461	493	657	2510	3767

Finally, the last two columns of Table 3 split the data by levels of education completed. Interestingly, the relation of earnings to test scores is stronger for less educated individuals. This is relevant because many special government programs are targeted at disadvantaged students who never attend college. To the extent that early test scores do well at predicting this group's earnings, the case for tying compensation and rewards to test scores is strengthened.

The bottom line is that test scores appear to be a potential proxy for subsequent earnings. This result is helpful irrespective of sorting or incentive view. For sorting, teachers who are more effective at raising test scores will be most attracted by payment based on ΔSCORE. For incentives, to the extent that a teacher can affect ΔSCORE by altering her behavior, there is evidence that such changes will result in increased earnings of her students.

1.2. Input-based pay

Although there is some relation between test scores and subsequent earnings, the link is a weak one.[13] As a result, it is useful to consider the implications of paying on the basis of input, rather than or in addition to paying on output.

Paying on input has the advantage that it removes risk from the worker and prevents workers from focusing too much on the metric that is easily measured. The general notion is that workers are risk averse. In order to induce them to accept a risky compensation package, the overall average level of pay must be higher. In the case of teachers, the risk argument is probably not too important. Teachers are paid on the basis of student performance, and if teachers have anywhere from 30 to 150 students per year, the year-to-year variation in the average test scores for their classes is likely to be small. Furthermore, only a fraction of their wages would be tied to performance measures even under the most ambitious schemes. It is difficult to imagine a scenario where earnings would fluctuate more than a couple of percent from year to year around some basic trend.

The second issue, that teachers focus on measured performance, is more fundamental and is the one that has been the subject of the last few paragraphs. Often, this is referred to as "teaching to the test." In the case of manufacturing, the concern is that paying a piece rate will induce workers to push for high quantity and shade on quality.[14] Stated differently, when one dimension of the product is easily measured (in manufacturing, quantity), but another dimension is not (quality), workers will put their effort into maximizing the measurable one at the expense of the unmeasured one. In the context of teaching, paying on the basis of reading and math test scores will induce teachers to favor those subjects and pay less attention to say, creative activity. The weightings chosen by the teachers will not necessarily reflect the values to the students or society as a whole.

Payment on the basis of perfectly observed input creates appropriate incentives without distortion. In the extreme case, workers would be paid on the basis of some perfect measure of disutility associated

[13] Coefficients on ΔSCORE are not huge and the amount of earnings variation accounted for by SCORE and ΔSCORE are small with R-square values being very low.

[14] Again, see Lazear (1986), Baker (2000) and Holmstrom and Milgrom (1991) for formal statements of the problem.

with the work. Then, there would be no incentive to favor one part of the curriculum over another. Teachers would be fully compensated for any disutility associated with their efforts and would be willing, therefore, to simply do what was best. Even if teachers had preferences for teaching one subject over another, compensating on the basis of disutility would keep them indifferent because they would receive just enough more for teaching the disagreeable subjects.

This scenario creates measurement problems of its own. To the extent that the total hours of work are a good proxy for the disutility of teaching, then the problem is manageable. But when teachers care greatly about how the hours are spent and not just on the total amount of time spent—the more realistic case—input-based pay will not be a solution either.

Fortunately, there are some ways to make the problem tractable. The market provides signals on the disutility associated with teaching various subjects. For example, business schools compensate those in accounting better than they compensate those in organizational behavior. In part, the difference is compensation for perhaps less pleasant or more difficult work. The wage differences do not reflect caprice; they are the result of market forces that prevent business schools from hiring professors of accounting at the same wage that they can attract professors of organizational behavior.

As a practical matter, American schools base pay on skill measures, paying teachers with master's degrees more than they pay teachers with bachelor's degrees. The evidence suggests that there is little basis on which to create these salary differentials. Hanushek (2002) finds no effect of the teacher's having a master's degree on performance of the students they teach. They point out the importance of teacher quality and show that different teachers have very different effects on the performance of their students, but unfortunately, the effects are attributable to unobservables rather than to observable characteristics on which compensation can be based.

1.3. Subject-based pay

One significant problem in the US is that in some subject areas, there is an ample supply of teachers while in others, there is a dearth of qualified teachers. The shortages are most pronounced in math and sciences. Shortages exist in markets when one and only one condition holds: Prices are not allowed to adjust freely to equate supply with demand.

In the context of teachers, the problem is too much pay compression, or "equity" based pay structures. Because math teachers and history teachers work approximately the same number of hours and put in approximately the same amount of effort, some believe that they should receive the same compensation for the job. A policy of this sort results in disequilibrium since math and science teachers have better outside options than do history, language, music and physical education teachers. As a result, math and science courses are often staffed by those who never studied in college the subjects that they teach.

The discrepancy can be remedied only by allowing the salary to be different (either in monetary or non-monetary components) or by increasing the supply of those in short supply by direct subsidies to education in the scarce fields. Either way, those who enter math and sciences must earn more. They do so by earning on the job or by receiving preferential treatment when they are studying in college, but their lifetime compensation, net of education cost, must be higher, irrespective of the method used. Failing to differentiate on the basis of subject area will invariably result in shortages in some fields and surpluses in others.

2. Context specific pay structures: Sweden and the US

It is useful in the context of Sweden and the US to examine first the relative earnings of teachers. Figure 2 shows the ratio of primary school teacher salaries to per-capita GDP. Both the US and Sweden are low relative to other OECD countries. This does not imply that in absolute terms teachers are paid less in Sweden and the US than in other OECD countries. It implies only that, relative to the country's per capital GDP, teacher salaries are low.[15]

The figure is informative. It points out that attracting talent into teaching in both Sweden and the US is likely to be difficult, given low relative salaries of teachers as compared with other OECD countries.

[15] Part of this could reflect differences in per-capita GDP, but it does not seem to. For example, Switzerland (a very wealthy country) and Korea (a less wealthy country) are among the highest in terms of teacher pay relative to GDP.

Figure 2. Ratio of teacher salary to GDP per capita

Absolute levels of teacher salaries are low in both Sweden and the US. In 2002, the average salary in education in Sweden was USD 28,500 overall for teachers in the municipal sector, which accounts for the bulk of employees. This is 99 percent the salary earned by secretaries and 75 percent that of police working for the central government. In fact, teachers in Swedish primary education only earn 13 percent more than housekeepers employed by the municipal sector.[16] The average public school teacher in Sweden receives about 82 percent the salary of the average Swedish worker who has at least two years of college but less than a doctorate. Thus, teacher salaries are well below that of individuals with comparable training and skill. Of course, teachers do not work a full year, and part of the pay differentiation reflects shorter hours of work. But, in the US, teachers work only 6.5 percent fewer hours than the typical college graduate, and their pay is 36 percent less than the average college graduate, based on data from 2001.[17]

Still, levels of pay are not everything. Working conditions vary, and sometimes allowances are made for differences in working conditions

[16] Statistiska centralbyrån (2002a,b) and World Almanac and Book of Facts (1999).
[17] Current Population Survey (2002).

and qualifications. This seems to be more the case in Sweden than in the US. In Sweden, adjustments to base salary are made at least in some districts for managerial responsibilities, teaching children with special needs, performing special tasks, teaching more classes or hours than required by full time contracts, some locational allowances, outstanding performance in teaching, completion of professional development activity, teaching courses in a particular field and being qualified in multiple subjects. In the US, adjustments to base salary are made primarily for managerial responsibilities, higher educational attainment of the teacher, and specific activities. The primary reason for paying teachers more in the US is that they have "better" credentials, e.g., a master's degree in addition to the four-year college bachelor of arts or sciences.[18] But as mentioned earlier, there is evidence that the teachers' possessing higher degrees has no observable effect on student performance.

The determination of teacher salaries is purely a local matter in the US. For the most part, the same is true in Sweden; however, the teachers' union and the Swedish Association of Local Authorities have agreed to a set of minimum salaries for newly employed teachers through March 2005. The same agreement also outlines a goal to create a process by which pay and performance are linked. There is also language that emphasizes that wages should be set so that individual performance is prioritized. Specifically, actual pay is to be individual and differentiated and is to reflect goals and results. "The procedure for pay reviews should be as follows: after presenting a pay review proposal to an individual employee, the employer passes it on to the local trade union of which the employee is a member; "If the trade union does not call for local negotiations, the employer's proposal is accepted."[19] A timetable was established for results at the national level. If improvements are not obtained, then a process of special negotiations between the parties is initiated in order to create improvements.

The American program initiated by the Bush Administration during the past two years and referred to above, entitled "No Child Left Behind (NCLB)," has a flavor similar to that of the Swedish structure. The NCLB plan focuses primarily on failing schools. Low-performing schools are targeted for some federal funds if they can improve at a

[18] OECD (2001).
[19] Berg (2001).

sufficiently rapid rate. Those that do not are subject to control by outside authorities, competition from private schools, and eventual receivership. The primary goal of the test is a level of literacy in both reading and mathematics. There is a presumption that performance on the tests (which can be determined by the individual states) is correlated with some specific goal like providing labor market skills, reducing the risk of crime, drug use and incarceration, or even a more intermediate variable like high-school completion.

Given the principles discussed in the earlier section, how do current pay structures in the US and Sweden stack up?

The most important fact is that teachers are not well paid in either Sweden or the US. Additionally, in the US, and perhaps to a somewhat lesser extent in Sweden, there is too much pay compression. Teachers pay is not sufficiently differentiated by field to attract talent in scarce fields without creating large surpluses in other fields.

To say that teachers are not well-paid enough requires some underlying understanding of the educational production function and the social efficiency of using more highly skilled labor in teaching. There are many jobs where it is simply not cost effective to hire more expensive, more talented labor. Although there is a benefit to the higher-skilled individual, the value of the additional output is insufficient to cover the higher cost of the more talented labor. The presumption, however, is that teacher productivity is probably too low, in part because of the comparison with the past. We know (Hoxby, 2003; Lakdawalla, 2001), for example, that in the US, teacher salaries have fallen in relative and real terms, teacher quality has fallen, and productivity in education has declined by about 30 percent over the past few decades. Both relative levels of education and value of human capital have declined during this period. Lakdawalla estimates that per-student cost of education has risen dramatically as schools have substituted more teachers for better educated ones. It is possible, of course, that teacher quality was too high in the past, especially given suppressed wages of female teachers that resulted because females were precluded from entering other fields. But it is at least questionable that such large changes reflect a new optimality. In the next section, I will argue that a specific technology that is consistent with the evidences suggests that teacher quality may be a key variable in determining educational output. Although this does not prove that actual quality is below optimal quality, it does suggest that much attention should be paid to this key variable.

In addition to the level of pay, the form is important. With the exception of recent modifications, pay is based primarily on input rather than output in both Sweden and the US. Pay is based on skills, on difficulty of performing the task, and sometimes hours of work, all measures of input. With rare exception in either country is it the case that pay is based on any measure of output. Again, because it is difficult to measure the actual desired output (such as labor market success) until many years after a student has left the school, it is virtually impossible to tie pay to the actual output goal. But it is rare that pay is tied even to output proxies, such as test scores.

The conclusion is that teacher pay structures are probably not efficient in that they are too low, too compressed, and perhaps too unrelated to proxies for output.

3. An educational production function.

In "Educational Production,"[20] I suggested a technology that provided implications consistent with a large number of empirical findings, especially those that relate to class size. The idea relies on the notion that peer effects, which are modeled as negative externalities that students may convey on one another, are important in classroom education. In classroom education, the ability of one student to get something out of a moment of class time depends on the behavior of others in the class. This is a clear application of the bad apple principle. If one child is misbehaving, the entire class suffers, in part because the teacher must devote her time to dealing with the disruptive student. Thus, let p be the probability that any given student is not impeding his own or others' learning at any moment in time. Then, the probability that all students in a class of size n are behaving is p^n so that disruption occurs $1-p^n$ of the time. [21]

The primary implication of this technology is that p is a very important force in determining educational output. To get a sense of this, note that when p=.99 so that each student is behaving 99 percent of the time, in a class of 25 at least one student will be misbehaving 22 percent of the time. Now let p fall to .98 so that each student is behaving 98 percent of the time rather than 99 percent. At least one student will be misbehaving 40 percent, disrupting educational pro-

[20] Lazear (2001).

[21] Actually, $1-p$ is the probability that a given student initiates a disruption. It does not matter, given the technology postulated, that others may or may not follow.

duction for all 24 other students as well as himself. It is straightforward to show that the optimal response to decreases in p is to lower the class size, but the effects of such class size reduction will be weak in equilibrium.

More important for the purposes here is that p can be thought of as a teacher-determined parameter as well as a student-determined one. Whether students behave or not depends on the teacher. Boring teachers are likely to produce less attentive students than fascinating ones. It is here that the role that teacher quality plays can be so important. Lakdawalla's evidence on the substitution of quantity for quality may indicate that American schools have exacerbated the problem of falling productivity.

In Lazear (2001), a simulation was performed to estimate the effects of improving teacher quality. Under not unreasonable assumptions, it was shown that a huge increase in salary could be cost effective because its effects on student productivity were so profound. Of course, whether the simulation is valid or not is an empirical question, and no mere example can determine the relevance of the numbers cited. But the mechanism emphasized in the simulation was sorting, not incentives. There was no assumption that higher pay would result in more effort by the teachers. Pay was not tied to output in any formal way. Instead, the idea was that more high-quality labor would be drawn into teaching from other occupations. The increase in pay allowed for a better applicant pool and better selection. In order to make the mechanism effective, it is necessary that the schools hire and retain the best from the applicant pool and allow those who are less well suited to teaching move out of the profession. In this sense, pay and performance are clearly linked. But there is no explicit tying of teacher pay to test scores or any other specific metric. This approach means that the incentive mechanism is not on hours worked or on effort per hour. Rather, the incentives work to modify occupational choice by raising the relative attractiveness of teaching to other occupations.

3.1. Tenure

In most districts in the US, teachers are awarded tenure, which guarantees long term employment early in their careers (usually after three years and sometimes as early as after one year.) In the US, tenure is specific to the school or school district; it rarely carries across districts. Districts are geographically small, usually limited to a city or county, so that if a teacher wanted to move from San Francisco to Los Angeles, both in California, the already tenured-in-San-Francisco teacher would have to earn tenure again in the Los Angeles school. Of course, a school could award tenure to an already-tenured teacher in order to attract her, but there is no requirement that this be done.

In Sweden, employment guarantees, explicit or implicit, are also the rule. How does the institution of tenure affect teacher incentives?

The main mechanism is through adverse selection. If turnover rates were in the range of 10 percent per year, and were independent of years of service, then after seven years, more than half of the school's teachers will have left. It is easily possible to affect the quality of teachers with such a high turnover rate. But the independence assumption is invalid. Departures are not independent of service, nor of teacher quality. Instead, given the low pay of teachers relative to other college graduates, it is likely that disproportionately the most dedicated and those who cannot obtain alternative high offers remain. The former group is desirable, but the latter is not. Most of the adverse consequences result directly from pay compression. Because pay is compressed, the least productive teachers earn more relative to their alternatives than the most productive ones. Even if pay is related to performance, the weaker link inside the teacher profession than in the rest of the labor market means that poor teachers are relatively overpaid and good teachers are relatively underpaid. Departures of many of the better teachers is the result.

The institution of tenure formalizes employment guarantees in teaching, but long-term employment is the rule for most senior workers in other occupations as well. Tenure per se, coupled with pay compression, results in difficulties primarily because it does not encourage the best to stay and the worst to leave.

4. Reshuffling

Some incentives serve to reshuffle rather than change the level of effort or average quality of individuals in teaching. One example is instructive. A few years ago, California instituted a policy requiring that class size in the first three years of school be reduced from their current levels to no more than 20 students per class. The effect was to increase the demand for teachers, which meant that new teachers were brought into the profession from the outside. For the most part, the new teachers were young, inexperienced and lacking in teacher training. This had two effects. First, it lowered the average quality of teachers. Second, it moved teachers around in a way that was surely unintended when the initial policy was enacted. Specifically, because of the increase in positions, teachers, especially the best among them who had previously taught in schools in areas with disadvantaged students, were now able to obtain jobs in the wealthier suburbs where teaching was less difficult. This teacher flight to the suburbs meant that schools in disadvantaged areas had to be staffed by the newer, probably lower quality teachers.

An increase in demand for teachers that is unaccompanied by increases in compensation results in a lower quality pool of teachers. More important for the purposes here, the policy resulted in significant reshuffling.

Reshuffling can occur from any number of policies. When is reshuffling optimal? This question is subtle and depends on the nature of the educational production function. It is reasonable to expect that educational output depends on the ability of the student and the teacher, but the question is one of interaction effects. Formally, write

$$\text{Educational output for student } i = f\left(q_i, Q_j\right) \tag{5}$$

where q_i is a measure of the ability of student i and Q_j is a measure of the ability of teacher j, $f_1 > 0$, and $f_2 > 0$. The issue is the sign of f_{12}. If f_{12} is positive, then from a pure efficiency point of view, it is optimal to put the best teachers with the best students. If, on the other hand, $f_{12} < 0$, it is optimal to do the reverse so that the better teachers are in the schools with the poorest students. Although the sign of f_{12} is an empirical question, there is some reason to believe that f_{12} is positive. If we examine sectors of education that are most competitive, for example, the college level, it seems clear that the most distinguished faculty tend to be at the universities that have the best students. This suggests that although everyone benefits from having the better teachers, the best students benefit most from a given increment in teacher quality.

Equity is a major issue here, however, and equity and efficiency are likely to be at odds. Even if it would be efficient to put the best teachers at schools with the highest achieving students, doing so is likely to fly in the face of fairness. To the extent that education is almost a necessary ticket to high incomes, and to the extent that teacher quality can affect the eventual incomes that their students earn, placing the better teachers in the schools with the best students is likely to exacerbate inequality. If $f_{12} < 0$, then there is no tension: The better teachers should be placed in the schools with the poorer students.

In any event, the goal is not so much to shift teachers around from school to school, but rather to change the quality of the overall pool. The only way to accomplish this is to attract higher-quality people to teaching from other occupations. Higher salaries seem to be the answer, but there remains the question as to how elastic is the supply of labor to the teaching profession. Again, this is an empirical issue, but the direction of the effect is clear. Irrespective of the elasticity, it seems quite likely that reshuffling teachers cannot be the only solution to the problem.

5. Working conditions and teacher preferences

At a given wage, workers prefer more pleasant working conditions to less pleasant ones. Teachers are no different. For example, it is no surprise that teachers prefer reductions in class size. Even if reductions in class size had no beneficial effects on students, many teachers, myself among them, would prefer smaller class sizes simply because teaching a smaller class is easier and more pleasant. The key is the relation of wages to working conditions. Because teacher salaries may not fully reflect working conditions, teachers do not have proper incentives.

It would be quite easy to induce efficient provision of job amenities. The hedonic price literature[22] provides extensive analysis of the way that competitive markets price amenities to bring about the appropriate combination of wages and working conditions. The following simple example illustrates the point. Suppose that teachers prefer that the school provide parking that is next to the school building. Let the value that teachers place on parking be x per day. Suppose further that the rental cost of the land for the parking space is equal to y. In the absence of parking, teachers earn k dollars. Then in a competitive labor market, firms can offer a job to teachers where parking is provided that pays a wage of $k - y$ at the same net cost as the job without parking that pays k. The value of the job with parking is the wage, $k - y$, plus the value of the parking x, or

[22] The seminal paper is Rosen (1974). Antos and Rosen (1975) applied this analysis to examine salaries and attributes in the teaching profession.

$$k - y + x. \tag{6}$$

The value of the job without parking is k so the teacher takes the job with parking when

$$k - y + x > k \tag{7}$$

or when

$$x > y. \tag{8}$$

This is, of course, the efficiency condition because it says that parking is provided when the cost of parking, y, is less than the value of parking x.

Distortions arise when wages are not adjusted appropriately. For example, suppose that jobs that provide parking paid $k - z$ where $z < x < y$. Teachers would prefer the job that provided parking because

$$x > z \tag{9}$$

implies that

$$k - z + x > k. \tag{10}$$

However, providing parking is inefficient because $y > x$ means that the cost of the parking exceeds its value.

Under these circumstances, there is a distortion between the preferences of the school districts and the teachers. The district takes into account the cost; the teacher takes into account the benefit. The cost to the district of a job with parking is $k - z + y$. The district prefers to withhold parking when

$$k - z + y > k \tag{11}$$

or when $z < y$. The teacher prefers parking when $x > z$. But when $z < x < y$, the teachers push for parking and the district resists. The district behaves appropriately, but teachers are induced to push for better working conditions not because the working conditions are valuable to them, but because they are not priced appropriately.

Competition does not prevail in education because there is a very large subsidy to attending public schools over private ones. At the extreme, one can think of the situation as being one of monopsony. But a profit maximizing monopsonist should not err in the provision of working conditions relative to wages. The monopsonist who behaves efficiently with respect to this tradeoff can extract the maximum amount of surplus from its exploited workers. To see this, consider the previous example. In the case of monopsony where the monopsonist can extract all rent from the employment relationship, the monopsonist can push the worker to his reservation level of utility. Let that (measured in kronor) be denoted u. The worker is indifferent to being paid u without parking or $u - x$ with parking; both allocations yield the reservation level of utility. The firm can provide the former at cost u and the latter at cost $u - x + y$. The firm then prefers to offer parking whenever

$$u - x + y > u \tag{12}$$

or whenever $x < y$. Once again, the firm chooses to do what is efficient. The fact that the firm has monopsony power in no way distorts the allocation between wages and working conditions.

If job amenities are not priced appropriately, it is likely to be a result of political rather than economic considerations. It is well known that unions compress the wage structure. There is also evidence[23] that public utilities and governments also have more compressed wages than other firms. The logic behind the empirical phenomenon is less than clearly spelled out, but we can probably take as given that job attributes are not priced appropriately in the public sector in general and in teaching in particular.

[23] Pergamit (1983, 1985).

6. Divergence between teacher and student preferences

The discussion in the last section implies that there is a difference between teacher desires and administrator desires. Is there an analogous difference between teacher desires and student (or parent) desires?

Since public schools are provided without explicit charges to the student, students and their parents would like the school to provide any extra service that has positive benefit. For example, it is often the case that parents clamor for smaller class size. However, if smaller class size were explicitly tied to tax increases or to a reduction in other school inputs like fewer textbooks, the decision would be less clear.

Thus, whether parents or students internalize decisions appropriately depends on their view of the costs. If they believe that every change in school inputs is passed on to them directly in the form of increased taxes, then there is no reason to expect that students and parents would demand too many school inputs.[24] But this connection is likely to be loose. However, if students and parents understand that the school district faces a budget constraint, then decisions should be aligned. For example, if the school can spend only 15,000 per student, a reduction in class size has a direct impact on other resources. If textbooks are a more important input into the production process than is a reduction in class size, students should not favor a class size reduction. Of course, this requires two pieces of information, which may require a high degree of sophistication. First, students must know the educational production function. Second, they must recognize the absolute nature of the budget constraint. Relaxing either one of these considerations will induce students to have preferences that deviate from efficiency.

Under such circumstances, student preferences may also deviate from teacher preferences. Suppose that neither teachers nor their students regard the addition of school resources as being costly to themselves. Will they prefer the same resources? The answer is a clear no. Take the example of teacher parking above. If the value of teacher parking is x to the teacher, it is likely to be far less than x to the student who does not benefit directly from the parking space that the teacher is given. Class size reduction is usually the policy that receives the most attention. There is no reason to expect that teacher and student values on class size reductions will be the same. Neither side may

[24] There might be heterogeneity in preferences, but even this could be handled by sorting across neighborhoods and school districts.

get it right in the sense of selecting the right allocation of resources, but it is far from clear whether teachers or their students will push harder for class size reduction.

7. Conclusions

There are many lessons from the compensation literature that apply to teacher incentives. Still, the logic of the preceding essay suggests a number of different implications.

The primary issue relates to the level of pay. Certainly in the US, and perhaps in Sweden, teachers are not paid as much as the typical college graduate, even when shorter hours are taken into account. As a result, the applicant pool is restricted and probably lower quality than would be optimal were education run through a private market.

If pay is increased, the size of the applicant pool will be enlarged. The strategy has the most value when those who are likely to be the best teachers can be identified during the hiring process. Unfortunately, there is no strong evidence that observable characteristics of the teacher are good predictors of the ability to affect student performance. This implies that turnover is an important part of the selection process.

Separating the lowest quality teachers requires that teacher quality can be identified. Again, we are back to the problem that much of the output of the teacher is not observed until many years after the student has had her course. But this problem is not specific to teaching. Even in private industry, there are many jobs for which it is difficult to measure the output and for which the results do not come in until long after the actions have been taken. (Research and development is an obvious case in point, but the same would be true of a vice-president of finance and many other managers.) Yet, firms deal with this problem by using a variety of evaluation techniques. Many times, the evaluations are subjective. One well-known company executive imposed a 10 percent rule on his managers: All workers are ranked and each year the lowest 10 percent are terminated. This somewhat draconian rule may not be ideal, but it emphasizes that ranking employees and terminating those who are at the bottom of the distribution is common in all areas.

The institution of tenure tends to reduce the flexibility that schools have in encouraging departures of their least effective teachers. Tenure, de facto or de jure, is unlikely to be eliminated, but if the time period before which tenure is granted were lengthened, the process would be more likely to weed out the least effective teachers, especially since teacher fixed effects seem to be very important in determining output.

Second, pay uniformity across fields results in disequilibrium where certain subjects have an ample supply of applicants and other subjects go wanting.

Third, a general reduction in pay compression would likely reduce the tendency for the best teachers to be attracted to other opportunities and away from teaching. Because better teachers are not paid much more than poorer teachers and because better teachers are also likely to be better at other jobs as well, the top of the profession finds it easier to move than the bottom.

Fourth, the limited evidence suggests some positive effects of output-based pay. The pay-performance relation can work through the incentive or through the sorting channel. However, it is inefficient to tie compensation to the wrong metric. Given that the current state of knowledge on the relation between output and pay is limited, administrators are unlikely to place too heavy a weight on objectively measured output-based pay. Whether the primary effect of tying pay to test scores is on incentives for incumbent teachers or has its primary effect by selecting those teachers who are best at improving test scores remains to be determined.

References

Antos, J.R. and Rosen, S. (1975), Discrimination in the market for public school teachers, Journal of Econometrics 3, 123-150.

Baker, G.P. (2000), The use of distortionary performance measures in incentive contracting, American Economic Review 90, 415-420.

Barro, R.J. (2001), Human capital and growth, American Economic Review 91, 12-17.

Becker, G.S. (1962), Investment in human capital: A theoretical analysis, Journal of Political Economy 70, 9-49.

Berg, A. (2001), Five-year pay agreement for teachers finally concluded, Eironline (http://www.eiro.eurofound.ie/2001/01/inbrief/SE0101179N.html).

Current Population Survey (2002), March (http://www.bls.gov/cps/).

Encinosa, W.E. III, Gaynor, M. and Rebitzer, J.B. (1998), The sociology of groups and the economics of incentives: theory and evidence on compensation systems, Working Paper, Case Western Reserve University, Cleveland.

Hall, R.E. and Jones, C.I. (1999), Why do some countries produce so much more output per worker than others?, Quarterly Journal of Economics 114, 83-116.

Holmstrom, B. and Milgrom, P. (1991), Multi-task principal-agent analyses: Incentive contracts, asset ownership and job design, Journal of Law, Economics, and Organization 7, 24-52.

Hanushek, E.A. (2002), The long-run importance of school quality, NBER Working Paper 9071, National Bureau of Economic Research.

Hoxby, C.M. (1996), How teachers' unions affect education production, Quarterly Journal of Economics 111, 671-718.

Hoxby, C.M. (2003), School choice and school productivity: (Or could school choice be a tide that lifts all boats?), in C.M. Hoxby (ed.), The Economics of School Choice, University of Chicago Press, Chicago.

Kandel, E. and Lazear, E. P. (1992), Peer pressure and partnerships, Journal of Political Economy 100, 801-817.

Lakdawalla, D. (2001), The declining quality of teachers, NBER Working Paper 8263, National Bureau of Economic Research.

Lavy, V. (2002), Evaluating the effects of teachers' group performance incentives on pupil achievement, Journal of Political Economy 110, 1286-1317.

Lazear, E.P. (1986), Salaries and piece rates, Journal of Business 59, 405-431.

Lazear, E.P. (1995), Personnel Economics, MIT Press, Cambridge, MA.

Lazear, E.P. (1998), Personnel Economics for Managers, John Wiley & Sons, Inc., New York.

Lazear, E.P. (2000), Performance pay and productivity, American Economic Review 90, 1346-61.

Lazear, E. P. (2001), Educational production, Quarterly Journal of Economics 116, 777-803.

Mincer, J. (1962), On-the-job training: Costs, returns, and some implications, Journal of Political Economy 70, S50.

Neal, D. (1997), The effects of Catholic secondary schooling on educational achievement, Journal of Labor Economics 15, 98-123.

Neal, D. and Johnson, W.R.. (1996), The role of premarket factors in black-white wage differences, Journal of Political Economy 104, 869-895.

OECD (2001), Education at a Glance, Table D1.3 (http://www1.oecd.org/els/education/ei/eag/list.htm).

Pergamit, M.R. (1983), Wages and employment in regulated industries, Ph.D. Dissertation, University of Chicago, Chicago.

Pergamit, M.R. (1985), Wages in regulated industries, Working Papers 152, Bureau of Labor Statistics.

Rosen, S. (1974), Hedonic prices and implicit markets: product differentiation in pure competition, Journal of Political Economy 82, 34-55.

Schultz, T.W. (1961), Investment in human capital, American Economic Review LI, 1-17.

Statistiska centralbyrån (Statistics Sweden) (2002a) (http://www.scb.se/statistik/am0106/am0106_tab2eng.asp).

Statistiska centralbyrån (Statistics Sweden) (2002b) (http://www.scb.se/statistik/am0104/am0104_tab2eng.asp).

World Almanac and Book of Facts (1999), Press Pub. Co., New York.

Appendix A

Formally, an index can be derived as follows: Suppose that the ex ante density of achievement test scores, A, a teacher faces is given by $f(A)$. Let the (causal) transformation of achievement test scores into lifetime wealth or whatever other metric society deems to be the relevant dependent variable be given by $h(A)$. If it were wealth, then

$$\text{Wealth} = h(A) \tag{A1}$$

so improving a student's achievement score from A_0 to A_1 would result in an increase in lifetime wealth of

$$\Delta\ \text{Wealth} = h(A_1) - h(A_o). \tag{A2}$$

The value that a teacher creates when she shifts the distribution of test scores from $f(A)$ to $g(A)$ is then

$$\int_0^\infty g(A)h(A)da - \int_0^\infty f(A)h(A)da. \tag{A3}$$

Think of the $g(A)$ distribution as the achievement distribution that is produced only as a direct result of teacher effort. At the end of a year's time, achievement scores may shift because of effort by students, parents, or other actions having nothing to do with the teacher. Let the shift in A that is reflected in the $g(A)$ density be that part that is directly attributable to teacher activity.

For a school to provide perfect incentives, the teacher must receive the full benefit of her effect on productivity. Suppose that her reservation wage is W per year. Then a compensation scheme that has the following form provides perfect incentives and induces her to accept the job as teacher:

$$N\int_0^\infty \left[g(A) - f(A)\right]h(A)da - K \tag{A4}$$

where N is the Number of students she encounters in a year and

$$K = N\int_0^\infty \left[g(A) - f(A)\right]h(A)da - W. \tag{A5}$$

Appendix B

Let

$$\text{Wealth} = \alpha + \beta A. \tag{B1}$$

Then substitution of (A1) into (2) yields

$$\int_0^\infty \left[g(A) - f(A)\right](\alpha + \beta A)dA - K \tag{B2}$$

which can be rewritten as

$$\beta\int_0^\infty \left[Ag(A) - Af(A)\right]dA + \alpha\left[\int_0^\infty g(A)dA - \int_0^\infty f(A)dA\right] - K \tag{B3}$$

or

$$\beta\int_0^\infty \left[Ag(A) - Af(A)\right]dA - K \tag{B4}$$

because the last two integrals equal 1.

The term is the difference between the ex post average score and the ex ante average score times *β*.