NBER WORKING PAPER SERIES

EDUCATIONAL PRODUCTION

Edward P. Lazear

Working Paper 7349
http://www.nber.org/papers/w7349

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 1999

This research was supported in part by the National Science Foundation. I am grateful to Gary Becker, Janet Currie, Jeffrey Grogger, James Heckman, Sherwin Rosen, and participants of workshops at Stanford's GSB and Hoover, UCLA and Rand, and the University of Chicago for comments and discussions. I especially thank Michael Schwarz for outstanding research assistance. The views expressed herein are those of the authors and not necessarily those of the National Bureau of Economic Research.

Educational Production
Edward P. Lazear
NBER Working Paper No. 7349
September 1999

ABSTRACT

The literature on class size yields a number of findings. First, class size effects are difficult to find except when using data where class size variations are truly exogenous. Second, Catholic schools have large classes and better performance. Third, to the extent that class size matters, it is more important for disadvantaged children. Special education classes are smaller than advanced placement classes. Fourth, when many students have joined a class recently, the joiners and their classmates do worse.

The theory presented below reconciles all of these facts by recognizing that classroom teaching is a public good where congestion effects are potentially important. Because the optimal class size is larger for well-behaved students, the observed relation of educational output to class size is small or even positive. However, increasing class size to ranges away from equilibrium levels will adversely affect educational output. The theory argues for a particular non-linear relation of educational output to class size and is consistent with observed variations in class size by grade level, student and teacher characteristics. Class size effects are more significant in small classes than large ones. There is a special function that maps the substitution of discipline for class size, which may explain why Catholic schools, with large classes, out-perform public schools. The same technology also implies that class size effects are larger for problem children than for well-behaved children.

Private schools, which charge a positive price and compete with free public schools, attract better students. This selection may help explain why Catholic schools out-perform public schools even though expulsion rates are lower in Catholic schools than in public ones.

Teachers may prefer smaller class size than student or parents either because wages do not reflect working conditions fully or because teachers as a group can raise the demand for their services by lowering class size.

The theory provides a measurable and operational way to define school quality that can be tested empirically. Finally, because public schools that operate in a centralized environment do not capture the returns to their successes, public school incentives differ from those of private schools.

Edward P. Lazear
Graduate School of Business
Stanford University
Stanford, CA 94305-5015
and NBER
lazear_edward@gsb.stanford.edu

Results from the enormous literature that studies educational production are varied and equivocal. A reading of that literature leaves a number of impressions. First, class size effects are difficult to find and are generally found only in the natural experiments literature. Other analyses are frequently inconsistent with those findings. Second, studies that examine public versus private education find that the performance of students from Catholic schools is better than that from public schools, despite the larger class size found in the Catholic schools. Furthermore, Catholic schools have lower expulsion rates than public schools so the obvious selection argument does not seem to be operative. Third, class size seems to matter more for children from disadvantaged homes and for special needs children than it does for average children.

How can these findings be reconciled? The approach used here is to note that classroom education is a public good, and as such, subject to extreme complementarity. The complementarity gives rise to particular a production technology that can make sense of the diverse findings a tie together a number of other facts.

The class size effect puzzle is most perplexing. It is almost impossible to find strong, consistent effects of class size on student performance. At some basic level, this fact makes no sense because observed class size is generally smaller than the entire number of students at any particular grade level. Why bear the expense of having four kindergarten classes of 30 rather than one class of 120 if class size truly does not matter? Furthermore, observed class size varies with age of the student. Pre-school classes are smaller than large lecture classes for college students. How is this to be explained if class size is irrelevant? Furthermore, there is pressure by teachers and administrators, if not parents, to lower class size. California recently cut class size by about 10%, imposing significant costs to the taxpayers.

The answer to the class size puzzle is that class size is a choice variable and the optimal class size varies inversely with the attention span of the students. It is efficient to use fewer teachers and a higher student-teacher ratio when the students are better behaved. But actual educational output varies directly with the behavior of the student, despite the fact that fewer teacher inputs are used. This means that changing class size matters in the expected direction, but the observed relation of educational output to class size may be small or even positive.

The confusion over facts results from a failure to embed them in a theoretical framework. The main purpose of the model presented here is to tie together a wide variety of facts and to integrate the literature on class size and student performance. A further goal is to provide a new empirical strategy for understanding student performance and the determinants of it. This is an ambitious goal, but it is hoped that some progress can be made by emphasizing the role that behavior plays in the determination of class size.[1]

The theory set out below is consistent with a number of facts. The main fact to be explained is that class size does not seem to affect educational output.[2] To the extent that it does matter,

[1]Other models of education with externalities have been offered by Caucutt (1996) and Fernandez and Rogerson (1998). The focus of these studies is somewhat different from the one here, but there is overlap in deriving sorting equilibria. Positive assertive mating, of the kind derived in these studies, goes back at least as far as Becker (1991).

[2]Hanushek (1998b) reports that expenditures per student more than doubled between 1960 and 1990 at a time when there was no steady trend in test scores or other measures of

changing class size has greater impact at small class sizes than at large ones.[3] Some additional facts are relevant. There is evidence that private schools do a better job in training students than public schools. Catholic schools do well, notwithstanding the large class sizes. Additionally, private schools segregate students by academic achievement. This is obviously true at the college level, but is increasingly true for k-12 education, where test scores, grades, interviews and letters of recommendation are required for admission. Public schools may also be segregated on the basis of academic achievement, but tuition is not the instrument used to accomplish the sorting. There is evidence that competition both lowers costs and improves school performance.[4]

The basic structure begins with the recognition that education in a classroom environment is a public good that is full of social interaction effects. But as with most public goods, classroom learning has congestion effects, which are negative externalities created when one student impedes the learning of all other classmates. There is a great deal of empirical support for this proposition.

---

performance. Other recent papers to examine the relation between expenditures and class size are Card and Krueger (1992) and Betts (1996).

[3]See Hanushek (1998a,b) who finds little evidence that anything matters including class size reductions. Coleman and Hoffer (1987) and Coleman, Kilgore, and Hoffer (1981) point out that Catholic schools with large class sizes produce better students than public schools with smaller class sizes. Class size effects are documented by a number of authors. The literature goes very far back. For example, Blake (1954) summarized a literature where 35 studies found smaller class size was better, 18 found larger class size was better and 32 were inconclusive. More recent are Hanushek (1998), Hoxby (1998) and Krueger (1997, 1998). Angrist and Lavy (1997) find that class size matters when for elementary school children in 4th and 5th grade in Israel.

[4]See Hoxby (QJE paper and 1998a) ,McMillan (1999), and Urquiola (1999).

Peer effects have long been recognized as crucial in education.[5] While hardly novel, to understand peer interaction effects it is necessary to embed the spillovers in a framework where changing the size of a class or its composition has a cost. The primary cost takes the form of teacher salary and infrastructure. The tradeoff between costs and benefits of changing size and structure are modeled explicitly and the framework is used to answer the following questions:

1. What is the optimal class size and what are its determinants? Are public and/or private schools led to choose the right structures?

2. What accounts for the small class size effects that are found empirically?

3. How do students choose between public and private schools and what empirical selection biases does this create?

4. Should more resources be devoted to certain students while others are neglected? rational behavior?

5. How does educational output depend on discipline or classroom etiquette?

6. Do teachers' and administrators' incentives deviate from those of the consumers of educational services? If so, why and in which direction?

7. Which variable is appropriate for measuring school quality? How does class size vary with student behavior?

Model

[5]There is a large literature here. An early empirical paper is Henderson, Mieskowski and Sauvageau (1977).

The driving force is the idea is that peer effects are important in classroom education. At some level the point is obvious. A classroom almost defines what is meant by a public good.[6] In the context of the public goods discussion, the cost of adding additional students can be thought of as congestion effects. In this setting, however, it is better to model the congestion effect more explicitly by thinking in terms of negative externalities that students may convey on one another. 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. Thus, let p be the probability that any given student is not impeding his own or other's learning at any moment in time. Thus, 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.[7]

The impediment may take a variety of forms. Student disruption provides a concrete example. Neither the student nor his classmates can learn much when the student is misbehaving, causing the teacher to allocate her time to him. Less nefarious, but equally costly is time taken by a student who asks a question the answer of which is known to all other students. One can think of

---

[6] Heckman (1998,1999) points out that learning is a lifetime affair and that the emphasis on formal schooling is misplaced, particularly when it is recognized that early success breeds later success. There is one major difference between formal schooling and the learning that occurs from infancy and continues after the cessation of school. It is that formal classroom schooling has public good attributes and externalities are key, whereas training given by a parent to his child or by an employer to his employee is essentially a private good.

[7] Actually, 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. Furthermore, it could also be assumed that the individual who asks the question benefits from that time even if others do not. This will change the functional form only slightly as educational output per person rises from p^n to p^{n-1}. All results remain qualitatively identical.

p as the proportion of time that a given student does not halt the public aspects of the classroom education process. Thus, the assumption made is that one child's disruption destroys the ability of all students (including himself) to learn at that moment.[8] It is expected that p would be relatively high because even having p=.98 in a class of 25 students results in disruption 40% of the time ($1-.98^{25} = .40$). Disruptive behavior may be viewed as deviant behavior, but most students are capable of disrupting for at least some fraction, in this case, 1-p, of the time, especially when disruption is interpreted to mean asking a question to which others know the answer.

Public goods are also provided by other students. Although true, it is uninteresting to look to the range of class size values where adding students produces positive rather than negative externalities. Because increasing class size reduces cost per student, the profit maximizing school will always increase class size to at least the point where additional students have negative effects on others. The optimum must be in the range where externalities are negative and so the focus is on this part of the story throughout the paper.

When there is peace in the classroom, learning occurs for each student with probability q, which may vary with the age of the student and the nature of the material being taught. Then the amount of learning that occurs at a given moment in a class of size n is

$$n\, q\, p^n \quad .$$

This technology has two obvious properties. First,

[8]This technology is that of perfect complementary. Kremer (1993) also presents a model using a perfect-complements-style technology.

$$\frac{\partial p^n}{\partial n} = p^n \ln(p) \quad < \quad 0 \quad .$$

and

$$\frac{\partial^2 p^n}{\partial n^2} = p^n (\ln(p))^2 \quad > \quad 0$$

so that learning per student declines with class size, but at a decreasing rate. The change in learning that occurs in going from a one student tutoring situation to a two student one is greater than the decline that occurs in going from a class of 30 to a class of 31. At p=.95, the effect of reducing class size from 6 to 5 students is over twelve times as great as the effect of changing class size from 30 to 29. This implication is consistent with the facts.[9]

This is a technological relation and says nothing about optimality. To determine what schools do when faced with different conditions, let us begin by asking how much would a student pay to be in a class of size n. Suppose that the value of a unit of learning is given by v. Then the value of learning that a student expects to derive from a class of size n is v is vqp^n. For most purposes, it is only the product vq that is relevant. Therefore, define

---

[9]Rosen (1987) argues that teaching is a labor intensive service that does not lend itself to mass production. This can be interpreted in this context as saying that what one student needs to know another does not, which can be interpreted as $p<1$. When the student is addressing the specific needs of one student, the rest of the class is not benefitting, or at least not benefitting by as much as they would were there personal needs addressed.

$V \equiv vq$.

As a modeling strategy, little emphasis is placed on variations in V. The goal is to attempt to reconcile the findings in the literature by focusing on variations in p. The production function chosen here is somewhat specific, which is intentional. It is intended to be a specific description of the production process and a parsimonious specification. Variations in teacher and student quality as well as other differences in the production function can always be thought to enter through V, but they are not the focus of the model. Many of the puzzles in the empirical findings can be explained by considering only congestion, p, and class size, n.

The parameters in this model, namely p and V, and not mere abstractions. They can be observed and estimated, at least in theory. It would be possible to go into a classroom and measure p, the proportion of time that a given student initiates a disruption. Further, once p is known, then estimating V is straightforward human capital analysis because V is the amount by which a moment spent learning raises wages. The effect of spending a moment investing in education is known (or at least there are many estimates of it). This equals Vp^n so V can be inferred, given that p and n are known.

Consider a school of Z students with m teachers and m classes. Let the cost of a teacher and the rental value of the associated capital for her classroom be given by W. For now, W is assumed to be independent of P and other working conditions. Then a private school that wanted to maximize profits could sell each moment at the school for $Z\ V\ p^n$, but would face a total cost of Wm. Maximization of profit would meaning choosing m so as to maximize

(1) Profit = Z V $p^{Z/m}$ - W m

because each class has n = Z/m students in it.

The first-order condition is then

$$(2) \qquad -V\frac{Z^2}{m^2}p^{\frac{Z}{m}}\ln(p) - W = 0 \quad .$$

Comparative statics can be derived from (2). Using the implicit function theorem, note first that

$$(3) \qquad \frac{\partial^2 profit}{\partial m^2} = VZ^2 p^{\frac{Z}{m}}\ln(p)\frac{2m+Z\ln(p)}{m^4}$$

which is negative for the solution to be an interior one. This implies that

2m+Z ln(p) > 0 .

Next, because ∂/∂W = -1,

$$(3a) \qquad \frac{\partial m}{\partial W}|_{f.o.c.} = \frac{1}{\partial^2 profit/\partial m^2} < 0 \quad .$$

Also,

(3b) $$\frac{\partial m}{\partial p}\Big|_{f.o.c.} = \frac{-VZ^2 p^{\frac{Z-m}{m}} (m + Z\ln(p)) / m^3}{\partial^2 profit / \partial m^2} < 0$$

for interior solutions, which are guaranteed for p near 1. Further,

(3c) $$\frac{\partial m}{\partial Z}\Big|_{f.o.c.} = \frac{VZp^{\frac{Z}{m}} \ln(p)(2m + z\ln(p)) / m^3}{\partial^2 profit / \partial m^2}$$

$$= m / Z \qquad > 0$$

Also note that

$$\frac{\partial m}{\partial Z}\frac{Z}{m} = 1$$

which is as expected. A ten percent increase in the size of the school is matched by a ten percent increase in the optimal number of teachers. The size of the school does not matter because all effects are assumed (unrealistically) to take place in the classroom only.[8]7

Finally,

[8]See Bedard, Brown and Helland (1999) for a discussion of school size effects.

(3d) $$\frac{\partial m}{\partial v}|_{f.o.c.} = \frac{\frac{Z^2}{m^2} q p^{\frac{z}{m}} \ln(p)}{\partial^2 profit / \partial m^2} > 0$$

and

(3e) $$\frac{\partial m}{\partial q}|_{f.o.c.} = \frac{\frac{Z^2}{m^2} v p^{\frac{z}{m}} \ln(p)}{\partial^2 profit / \partial m^2} > 0$$

Market equilibrium is achieved in two ways, given any existence of positive profits. Competitive entry of firms into the education industry drives up W through demand pressure on wages in the teachers market. At the same time, the supply of educated graduates to the labor market drives down v. Equilibrium occurs when maximization of profit results in zero overall profit to the competitive supplier of education.

This is the basic model and a number of implications can be derived from it.

Class Size:

The main purpose of the model is to provide a framework for discussing class size and how it varies with a number of factors.[9] To get a feel for this, consider an example. Suppose that V is

[9]The model could, in principle, be generalized to consider other kinds of expenditures as well. Focusing on class size is probably the place to start, particularly since Flyer and Rosen (1997) show that almost all of the rise in the cost of schooling over time is a result of reductions

normalized to 1. Then W, the price of a teacher's time must be priced relative to V. In equilibrium, the price of teacher time relative to the productivity of student time must be sufficiently low to make the activity profitable. If it were not, private schools could not exist. That is, from (1), a necessary condition for schooling to be worthwhile is that W is less than Z /m. The cost of a teacher must be less than the maximum educational output in her class.[10]

Suppose that the ratio of W to Vp^n is 5. The teacher's time is five times as valuable as what any one student gets out of the class. Then, if z=100 and p=.99 so that 99 percent of the time, any given student is not causing enough disruption to interrupt learning in the classroom, the first order condition (2) yields an optimum m of 3.94, which gives a class size of 25 students.

This example makes clear why it is so difficult to find significant class size effects.[11] Increasing class size from 25 to 27 would reduce educational output per student by only about two percent.

The more important point is that class size is a choice variable. This fact will cause researchers to observe small, or possibly even positive class size effects in cross-sectional data. Note from (3b) that $\partial m/\partial p$ is negative. The optimal number of teachers declines with p, which means that better behaved students are in larger classes. More disruptive students, who are themselves poorer learners, are found in smaller classes, but the effect of reducing class size is not sufficient to overcome their deficiencies. (This is proved in the appendix.) There is substitution, but it is

---

in class size.

[10]This is the maximum because $p^n \leq 1$.

[11]Krueger (1997) is one of the few exceptions.

incomplete. Thus, educational output in high p classes is higher. If class size varies primarily because schools are adjusting class size in response to the behavior of the students, then the larger classes will have the better students and higher educational output, providing a positive observed relation between class size and educational output[12]. In the example above, when p=.99, optimal class size is 25 and educational output per student is .78. When p falls to .98, the optimal class size is 19 and educational output per student is .68 because the effect of the lower value of p swamps the effect of the reduced class size. This impairs the ability of the researcher to find improved educational output when class size is reduced.[13]

It is also clear why class size effects are potentially quite important despite the inability to observe them in the data. If a given group of students with a given value of p were to be placed in a larger class, educational output would fall. This is why we observe four kindergartens of 30 students rather than one kindergarten of 120 students. A natural experiment that leaves p constant and changes class size should induce the expected class size effect (e.g., Angrist and Lavy (1997) and Krueger (1997)), whereas class size variation that reflects optimal adjustment to different values of p does not produce the expected result. Even if large classes have more educational output, reducing the size of a given class would increase educational output further.

---

[12]Becker and Murphy (1992) show that optimal teacher-student ratio varies with the distance from the final product produced. This, too, provides a relation of equilibrium class size to technological parameters, although their focus is somewhat different from the one of this paper.

[13]Olson and Ackerman (1999) find that wages are positively related to the pupil/teacher ratio of the schools that they attended. They attribute this to differences in teacher quality across districts.

The point is that even if class size effects are potentially important, in equilibrium, marginal changes in class size may have small effects on observed educational output. If large gains were available from lowering class size, then those changes would have been made.

Since class size varies inversely with W, large class size effects are most likely to be observed when the cost of teachers is low. Low teacher salaries imply low optimal class sizes. Reducing class size has a larger effect on educational output in small classes than in large ones. The empirical implication is that class size effects are most likely to be observed when teachers are relatively inexpensive. Pre-school teachers are less expensive than professors, which generates smaller class size for pre-schoolers. Class size effects should be more important in pre-school classes than they are at the college level.

Behavior and Class Size:

What happens as p, the probability that a student is a non-disruptive learner, changes? There are two obvious applications, one relating to age and the other to underlying social behavior of a student body. Age is the most straightforward. Younger children have shorter attention spans than does the typical college student. In a class of kindergarten children, the probability that a child is behaving is lower than that in a college class.

From (2), it is clear that when $p=1$, no interior solution for p can be found. The logic is straightforward. If $p=1$, then the class should be of infinite size. Any reduction in m reduces cost, but does not reduce output. If m must be a positive integer, then the solution is to set $m=1$. In general, it is clear from (3b) that deviations of p from 1 imply higher numbers of teachers and smaller class sizes. In the numerical example given above, the optimal class size was 25 when p=.99. If p falls to .95, then the optimal number of teachers rises from four to seven and class size is reduced from 25 to 14.

Even within grade level, class size varies with topic. Some topics require much discussion and tend to have lower class size. If other student time were as valuable as teacher time, there would be no need to have small classes. It would not matter whether another student or the teacher was speaking. Because these effects are negative, at least on the margin, "air time" devoted to other students have negative effects on learning. These lower p classes have higher optimal m values, other things the same.[14]

The implication is that if V, the value of a minute of first grade schooling is the same as the value of a minute of college schooling, then the first-grade class should be smaller than the college class. Although not particularly surprising, it is a direct implication of a model that takes into account the cost of teacher time and trades this off against the gains from having a less frequently disrupted class.

The same point can be made in another context. Suppose that there are two types of students and p differs between the groups. Let the *A* group have a higher value of p than the *B* group so that $p_a > p_b$. While *A* and *B* need not refer to grades earned by such students, the interpretation is not an

[14]An early economics paper on class size by Summers and Wolfe (1977) found different effects for low achievers than for high achievers. Low achievers benefit from smaller class size, but the reverse is true for high achievers. Their results are somewhat consistent with this model because low achievers probably have higher values of p, where the reduction in size effects are greater.

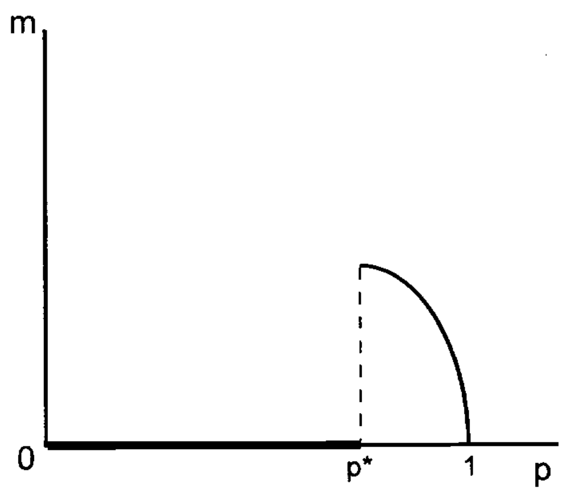
unnatural one.[15]

A direct implication of (3b) is that class size for *A* students is larger than class size for *B* students. The marginal value of reducing class size is greater for low p students in the relevant range than for high p students.[16]

If one thinks of the *B* group as being comprised of children with special needs, then this gives the implication that special education classes should be smaller than regular education classes. Sufficient for this implication is that the value of a moment of education is as high for *B*s as it is for *A*s.

For very low p, profitable learning is impossible. In order for the solution in (2) to be maximum, it is necessary that the second order condition, (3), be negative. It is possible for any given p and Z to choose an m sufficiently high so as to make the r.h.s. of (3) negative, but for sufficiently low values of p, the m that solves the first order condition yields negative profits. This is

Figure 1



[15]In the context of two different types, a question by a *B* student might be viewed as disruption to *A* students if all the *A* students already know the answer and *B* students do not.

[16]The relevant range is when an interior solution yields positive profit.

obvious from (1) since for p=0, profits are negative for any positive value of m. Thus, at sufficiently low values of p, it is better to simply give up. Figure 1 depicts the relation between the optimal m and p. As p declines, m rises until p reaches p*, at which point it does not pay to supply any education. At sufficiently low p, large classes where there is little pretense of education are created. This might be thought of as a babysitting role of schools.

A private school that accommodates low p students cannot compete. Either society decides that the value of educating these individuals is higher than the private value of V or society allows them to go without schooling. For young children the choice is the former. Society places a sufficiently high value on the education of young children that it puts them in schools, often in small classes, despite that fact that these children have a high tendency to disrupt and a low ability to learn. Implicitly, society is imputing a high value V to early education. But society is generally reluctant to do this for older individuals. College subsidies do not go to very low p individuals. This is consistent with a view that holds that there is little deviation between the private value of schooling and the social value of schooling for college students.

Special Programs:

Schools sometimes set up special programs and allocate slots in these programs on some kind of lottery or first-come-first-served basis. The programs often involve smaller class sizes and sometimes feature a different or extended curriculum. In this section, it is shown that this cannot be efficient.

When lotteries or other non-attribute based selection process is used, the pool of selected students will have the same characteristics as those of the non-selected attributes. Specifically, p does not differ across groups. Suppose that certain students are selected to be in smaller elite classes. Let m* denote the optimal number of teachers (and classes) to be used if all classes were symmetric, i.e., it is the solution to (2). Given Z, m* implies that the optimal number of students per class is Z/m* which is defined as n*. Now, if some students are selected for smaller classes, then some classes have $n' < n^*$ students and some have $n'' > n^*$ students. In order for it to be optimal to have asymmetric classes, the profit per class must be higher in the classes that have n' or n" students than in the classes that have n* students. That is,

$$n' V p^{n'} - W \quad > \quad n^* V p^{n^*} - W$$

or

$$n'' V p^{n''} - W \quad > \quad n^* V p^{n^*} - W$$

or both. But if this is true, then it would be possible to choose every class such that n= n' (or n") and do even better. This implies that n*, and therefore m*, are not optimal for the symmetric classes, which results in a contradiction. Therefore, there are no asymmetrically efficient solutions when all students have the same p.

Note that nothing in this proof requires that the material taught in the different classes be the same. The same logic applies to new superior curricula. If it is optimal to provide the new curriculum to some group of the students, then it is optimal to provide it to all students because students are ex ante identical. Special classes allocated randomly fail the test of ex post fairness and are also inefficient. Probably the best argument for asymmetric classes is that of experimentation. There may be a justification for the small classes as pilot programs. To the extent that educators do not know which method works best, it may be preferable to set aside a few experimental classes before implementing the program on a massive scale.

The Cost of Teachers and the Value of Education:

It is unnecessary to say much about increasing the cost of factors of production. The expression in (3a) is always negative (for interior solutions) which implies that an increase in the price of teacher time (or other educational inputs that are tied to her) means a reduction in the number of teachers and an increase in class size. As above, it is possible for a given p that W is so high to push the solution to a corner, i.e., m=0. As teachers' wages rise, it does not pay to allocate their time to students.

As already mentioned, an increase in V implies an increase in the optimal number of teachers. This follows directly from (3d). Because elementary education is the foundation for all that follows, the value of knowledge acquired during a year of schooling is highest at the earlier grades. ("Everything I need to know, I learned in kindergarten.") The fact that p is lower for young children than for older ones coupled with the observation that wages for elementary school teachers are significantly lower than those for college professors, reinforces the condition that pushes for smaller class size.

Sorting:

The schools may create problems by mixing non-disruptive students with disruptive ones. Given the technology postulated, a student with a high p gets virtually nothing out of an educational experience that is shared with a number of low p individuals. A classroom filled with disruptive individuals does not produce much learning, even for those who are themselves anxious to learn.

Whether schools should have mixed or matched classes depends on the value to the students. The result is that segregation by academic type is optimal. This is shown in two steps: First, the social optimum is computed. Then it is demonstrated that private schools, which operate in a market environment, attract the better students. Throughout, segregation and integration refer only to mixing or matching on the basis of an individual's type in terms of probability of disruption. Demographic characteristics are not considered here.

First, consider a school with Z students and m classes so that each class has n=z/m students. Let α of the students be *A*s and 1-α of the students be *B*s. Then a school with matched classes[17] has output

(4a) Output of matched school = $\alpha\, m\, n\, p_A^n + (1-\alpha)\, m\, n\, p_B^n$

because each class has n students and αm of them are *A* classes, whereas (1-α)m of them are *B* classes.[18] A school with mixed (in proportion) classes has output

[17]It is assumed that α and m are such that proportions work out to guarantee an integer number of classes, each of which has n students.

[18]It is possible to allow different class sizes in the segregated schools. It has already been shown that the optimal class sizes depend on p. This just adds unneeded notation. If segregation dominates when class sizes are assumed to be the same, then segregation surely dominates when class size in segregated schools are allowed to be different.

(4b) Output of mixed school = $m\ n\ p_A^{\alpha n}\ p_B^{(1-\alpha)n}$.

To show that it is always better to match then mix, it is merely necessary to show that the difference between the r.h.s. of (4a) and (4b) is positive.

Dividing through by n and m, the difference is given by

(5) $$\text{diff} = \alpha\ p_A^n + (1-\alpha)\ p_B^n - p_A^{\alpha n}\, p_B^{(1-\alpha)n}.$$

When $p_A = p_B$, diff = 0, as it must because then there is only one type so mixing and matching is irrelevant. Next, differentiate (5) with respect to p_A to obtain

$$\partial\,\text{diff} / \partial\, p_A = \alpha\, n\ p_A^{n-1}\,(1 - p_B^{(1-\alpha)n} / p_A^{(1-\alpha)n}) \quad > 0$$

for $p_A > p_B$. Thus, the difference is zero for $p_A = p_B$ and becomes positive for $p_A > p_B$. School output is maximized by matching rather than mixing student types in classes.[19]

Coleman and Hoffer (1981) and Coleman, Kilgore and Hoffer (1987) report that performance is higher in private schools and Catholic schools in particular than in public schools.[20] One possibility is that Catholic schools expel the troublemakers, leaving a population of students who are easier to teach than those in the public schools. However, the facts are that expulsion rates are

---

[19]Throughout this proof, it was assumed that *B*s prefer to be with *A*s because *A*s are less disruptive. Under the interpretation that a disruption consists of asking a question the answer to which is know by others, then *B*s might actually prefer to be with other *B*s because *A*s ask questions that are too advanced for *B*s. If true, this only reinforces the result that segregation is efficient. If segregation is efficient even when *B*s prefer to be with *A*s, it is surely efficient if *B*s prefer to be with *B*s.

[20]See also Evans and Schwab (1995), Neal (1997, 1998) and Sander (1997) for more recent papers.

lower in the Catholic schools than they are in the private schools. It is possible that sorting occurs at admission. Given that public schools are free and that private schools cost, there is a positive difference in price associated with going to a private school. (If all schools were private and competitive, *A* schools would be less costly than *B* schools since the latter optimally use more teachers per student.)

It is straightforward to show that As are willing to pay a higher price for admission to an all A school than are *B*s although both are willing to pay a positive price. An *A* receives value p_A^n from an all *A* school and $p_A^{\alpha n} p_B^{(1-\alpha)n}$ from an integrated school. (V is normalized to 1.) A *B* receives value $p_A^{n-1} p_B$ from an all *A* school and $p_A^{\alpha n} p_B^{(1-\alpha)n}$ from an integrated school. The difference between what an *A* will pay and what a *B* will pay to move from the integrated school to an all *A* school is then $p_A^n - p_A^{\alpha n} p_B^{(1-\alpha)n}$ which is positive. Thus, *A*s are more likely to pay private school tuition to get into an all *A* school than are *B*s, given that the alternative is free public school.[21] The direct implication is that Catholic schools selectively attract *A*s because *B*s are less willing to pay tuition for private education, given free public schools. The students who will pay the most to go to a private school

[21]In equilibrium, if all *A*s go to private schools, then the public schools consist only of *B*s. It remains true, however, that *A*s will pay more than *B*s to be in an all *A* school than in an all *B* school. The difference between what an *A* will pay and what a *B* will pay is

$$[Vp_A^n - Vp_B^{n-1}p_A] - [Vp_A^{n-1}p_B - V\, p_B^n]$$

which is positive for $p_A > p_B$.

Further, all *B* private schools charge more than all *A* private schools because optimal class size is smaller in the all *B* school, raising costs. If public schools are already exclusively *B*, private *B* schools could not compete unless they differed on other dimensions of quality.

are students with higher values of p.[22]

Private schools with *A*s should have smaller classes and should be cheaper than those with *B*s. This is surely true, at least at the extremes. Private schools for special needs children have small classes and are very expensive (although they are sometimes subsidized by the state). Schools for higher achieving students have larger classes than the special needs schools and are generally cheaper. But there are exceptions. Some private schools for highly talented students are expensive and have small class size. This is most likely a result of differences in q or V, which stands for learning ability. From (3e), class size declines with q. Less disruptive students are also better learners, which creates some empirical ambiguity since the effects on class size work in opposite directions.

Public k-12 schools use neither type-specific prices nor admissions criteria to sort students. It is true, however, that the implicit price of k-12 public schools does vary through housing prices and local taxes. Furthermore, these indirect price variations appear to sort students, albeit imperfectly.[23] The within-school variation in educational attainment, even in public schools, is small relative to the total variation. Some of this may be a result of the schooling itself and some is a result of the characteristics of the underlying student bodies.

---

[22] Rothschild and White (1995) show that if type-specific prices can be charged, then a competitive equilibrium results in optimality. Becker and Murphy (2000) discuss market induced sorting in the presence of externalities. Without a sufficient number of prices, there is inefficient allocation. In this context, scholarships allow for enough prices to induce social efficiency. Epple and Romano (1998) derive similar results and simulate some voucher experiments. The results of these studies apply here directly.

[23] See Hoxby (1998).

Setting A Good Example:

Low p students might benefit from being with high p students. One argument in favor of integrating schools along a variety of dimensions is that disadvantaged students can learn from more advantaged ones, without any detrimental effect on the better behaved group. The argument is not merely that disadvantaged students benefit from being around less disruptive students. The previous section showed that despite this factor, segregation is efficient. A stronger claim must be made to justify segregation on efficiency grounds. Necessary is that *B*s can be *transformed* into *A*s by being around them. If true, then *B*s benefit from *A*s in two ways. Classes that consist in part of *A*s are more peaceful than those with all *B*s. In addition, when *B*s are integrated with *A*s, *B*s may change their own behavior and become like *A*s. If this effect is strong enough, integrated classes are efficient.[24]

A generalization of (4b) that allows for *B*s to be transformed into *A*s is given by

(6) Output of mixed school with transformation $= m\, n\, p_A^{(\alpha+\theta(1-\alpha))n}\, p_B^{(1-\alpha)(1-\theta)n}$

where there are α *A*s, $(1-\alpha)$ *B*s, and where θ of the *B*s are transformed into *A*s by virtue of contact.

[24]Interestingly, there is an increasing trend toward using other factors, one of which is community service, as a criterion for admission to college. Perhaps colleges believe that those *As* who engage in community service are also likely to encourage *B*s to behave like *A*s.

In order for integration to be efficient it must be true that there exist values of the parameters such that net output of the integrated school exceeds net output of segregated schools, or

$$\text{(7)} \qquad m\,n\,(p_A^{(\alpha+\theta^*(1-\alpha))n}\,p_B^{(1-\alpha)(1-\theta^*)n} - W/n) > \alpha m_A n_A (p_A^{n_A} - \frac{W}{n_A}) + (1-\alpha) m_B n_B (p_B^{n_B} - \frac{W}{n_B})$$

where the *A* and *B* subscripts to the m and n variables refer to optimal values of those variables for segregated *A* and *B* classes, respectively. To see that integration may be efficient, consider an extreme case where $\theta=1$. All *B*s who have contact with *A*s become *A*s, but not the reverse: No *A*s are susceptible to negative influence of *B*s. Then, the l.h.s. of (7) is simply

$$m\,n\,(p_A^{\,n} - W/n)$$

or,

$$m_A n_A (p_A^{n_A} - \frac{W}{n_A})$$

because the integrated class consists of all *A*s (either by birth or transformation). This expression must exceed the r.h.s. of (7) since $p_A > p_B$ and a convex combination of a higher output and lower output schools cannot be as high as the higher output school alone.

Integrated schools may be efficient if integration turns *B*s into *A*s. The likelihood of this occurring must depend on the ratio of *A*s to *B*s. For example, if a school of 100 had 99 *B*s and 1 *A*, it would be difficult to believe that the one *A* student could change the behavior of all of the other *B* students. The reverse would be more likely to hold. The argument for integrated schools is therefore more persuasive when integration occurs on a limited basis. Flooding an *A* school with

*B*s is unlikely to be useful. Mixing a population of, say, 10% *B*s with 90% *A*s has a greater chance of success.[25]

Single-Sex Education:

Among the more obvious differences in k-12 education is that girls are less disruptive than are boys. Following the logic above, boys would want to be in all girl schools, but girls would not want them there. . The competitive equilibrium would drive price down to marginal cost, which is lower in an all girl school than in an all boy school or mixed school. Girls' schools should be cheaper than boys' schools and less disrupted. As such, single sex education should be a female issue rather than a male issue. If boys could be charged a higher price than girls for the same education, efficient segregation could be obtained without formal exclusion.[26]

At the college level, there is clear evidence of this pattern. There are currently 83 colleges and universities that list themselves as women only.[27] There are no corresponding cites for men only colleges. There are some famous historical counter-examples coming from exclusive or military schools that prohibited women from attending. Were there an equivalent female school, the

---

[25]Lazear (1999) presents a model of cultural acquisition. There, it is argued that incentives to become assimilated into the majority culture depend on the size of the relevant groups. The smaller is the minority relative to the majority, the greater is the incentive of a minority member to acquire the culture of the majority.

[26]Hoxby (1998b) finds that female classes perform better in the lower grades and that gender composition alters classroom conduct. There is a recent push to create girls-only classes in the public schools.

[27]See Converse College, "Facts about Today's Women's Colleges," Internet cite http://www.converse.edu/benefit7.htm, September, 1999.

argument that females would prefer exclusivity would hold. The push by females to enter male schools occurred because there were not equivalent female institutions.

Endogenous Discipline:

Catholic schools are known for strict discipline and some attribute the success of their educational programs to this discipline. Catholic schools also have large class sizes. Strict discipline is a substitute for small class size, given the production technology postulated. For any given level of educational output per head, X, there is always a tradeoff between class size and p:

$$p^n = X$$

or

$$p = X^{1/n}$$

so

$$dp/dn = -X^{1/n} \ln(X) / n^2 > 0$$

since X<1. To keep educational output constant, there is an increase in p that can always offset an increase in n. Larger classes produce the same output as smaller ones if they have sufficiently better behaved students.

Discipline is one way to produce higher p in the classroom. Another may be to promote a particular classroom etiquette. Although students are generally allowed to ask questions in class in college and graduate courses, questions are generally discouraged in large lecture classes having a few hundred students. Etiquette varies directly with class size. As class size increases to numbers like 500, p needs to approach 1 for there to be any educational output at all.

The functional relationship between class size and discipline is very simple. In order to increase class size by a factor of *k* and keep educational output per student constant, it is necessary to improve discipline such that p rises to $p^{1/k}$.[28] For example, in a class of 25, where p=.98, educational activity occurs 60% of the time. To double class size and obtain the same level of educational output, it is necessary to raise p from .98 to $\sqrt{.98}$, which equals .99. Thus, each student must behave one percent more of the time. Although this may seem like a relatively minor improvement in behavior, the statement can be turned around: The amount of misbehavior must be cut from 2% of the time per student to 1% of the time per student. This implies that a halving of misbehavior is necessary to effect a constant educational output while doubling class size. Whether this is large or small depends on student responsiveness to disciplinary incentives and this cannot be stated a priori. Nevertheless, it does make sense of the finding that Catholic schools have higher educational output per student even though class sizes are very large.

Of course, neither discipline nor etiquette comes without cost. Were it free to produce high levels of p, then all classes would consist of an extremely large group of passive and silent students. At some point, q, the learning component suffers when questions are prohibited.

An additional implication that finds support in data comes directly from this approach. The

---

[28]To see this, let p_1 be initial p, p_2 be the new p. To keep educational output constant while changing class size by a factor of k, it is necessary that

$$p_1^n = p_2^{nk}$$

or

$$p_2 = p_1^{1/k} .$$

proportionate increase in educational output associated with a decrease in class size from nk to n students is

$$(p^n - p^{nk}) / p^{nk} .$$

Differentiate with respect to p to obtain

$$d/dp = -n\, p^{(n-nk-1)} (k-1) < 0$$

for k>1. The proportionate effect of a class size reduction is greater for low p students than for high p students. Angrist and Lavy (1997) report that the class size effects are more important for disadvantaged students than for others. If disadvantaged students are also lower p students, then their results follow directly from this specification.

Movers

Using data from Texas, Hanushek, Kain and Rivkin (1999) find that children who switch schools perform poorer than those who do not and that moving imposes costs on other students. Both can be interpreted as implications of the model presented above. Since movers are unaccustomed to the new classroom, their questions are more likely to be disruptive in that they relate to material that their new classmates have already covered. Their p is low relative to the non-movers. This has two implications. First, movers own learning is slower than it would have been had they not moved. Second, because they lower the average p for the class, learning by others is reduced as well. This exactly what is found in the Texas study.

Teacher Preferences:

As many of us who are teachers know, teachers may have preferences about classroom composition. A primary consideration of teachers is class size. Despite the paucity of evidence supporting the claim, smaller classes are justified by the presumed effect on the academic achievement of the students. Another more cynical view is that smaller class size is a teacher benefit, not a student benefit. It is useful, therefore, to ask whether teacher interests are aligned with the interests of their students.

There are two issues that relate to class size. First, how does class size affect the teacher's workload and/or her utility from teaching? Second, are differences in class size reflected in the wage that she is paid? The first question relates to technology. The second relates to markets.

There is no need for teacher tastes to mimic the preferences of students or their parents. Much of the "taste" for small classes may have to do with wanting a reduced workload. On the other hand, a teacher who would like to get her message out to as much of the world as possible might prefer larger classes for a given amount of work. How do teacher and student desires differ?

Students care about class size but they also care about costs. In competition, the school must cover its costs which equal

$$W m / Z$$

per student. Initially, suppose that teacher salary were independent of class size. Then, students would want to maximize

$$p^{z/m} - W m / Z$$

which has f.o.c.

$$\frac{-Z}{m^2} p^{\frac{Z}{m}} \ln(p) - \frac{W}{Z} = 0$$

This is the same condition as above in (2), but it differs from the tradeoff that teachers make. If teacher salary were is independent of class size, then a teacher would prefer a class size that conforms directly to her utility function. That is, if teacher utility is given by

U(n, W) ,

teachers choose the n (equivalent to Z/m) that maximizes U for a given W. If size always enters negatively, then teachers prefer a class size that goes to zero. If size enters positively up to a point, then the teacher preferred class size equals that point. But the costs that the school faces are irrelevant to the teachers, even though they are not irrelevant to the students. It is not unreasonable that teachers would push for smaller classes than students or their parents, under these circumstances.

Of course, in an efficient labor market, teacher salary would not be independent of class size. This is the point of Antos and Rosen (1975) who estimate teacher hedonic wage functions. In a competitive labor market, teacher salaries adjust to incorporate the marginal teacher's willingness to pay for smaller class size. Teachers who teach in larger classes receive higher salaries to compensate. As long as the market adjusts salary efficiently, there is no tension between student preferences and teacher preferences. If teachers prefer smaller classes than students, then they will give up enough earnings to compensate students for the smaller class size.[29] This will always result

[29] Because the majority of private schools are Catholic, the reality is actually the opposite of what would be predicted. Teachers in public school receive higher salaries than those in

in efficiency. The maximization problem (both for society and the private school) now changes from (1) to

(1') $$\text{Profit} = Z\,V\,p^{Z/m} - m\,W(z/m)$$

which has the first-order condition

(2') $$-V\frac{Z^2}{m^2}p^{\frac{Z}{m}}\ln(p) - \left[\ W(\frac{Z}{m}) - \frac{Z}{m}W'(\frac{Z}{m})\ \right] = 0$$

The term enclosed in brackets in (2') is merely the effect of raising the number of teachers on costs. It has two components. Increasing the number of teachers implies an increase in costs because teacher salary is positive. Increasing the number of teachers also implies a decline in cost per teacher because class size is reduced and teachers (at least on the margin) are willing to give up wages for smaller classes. The private school does exactly the right thing from a social point of view in a competitive labor market because teacher salary adjusts to class size.

There may be a distinction here between public and private schools. Because of the wage setting process within public schools (especially in a union environment), wages are unlikely to adjust appropriately to class size.[30] If wages do not adjust, then there is likely to be pressure from teachers for smaller than optimal class size. Competition from teachers who are employed at private

---

private schools, despite the larger class sizes associated with private schools.

[30]See Hoxby (1996).

schools may place constraints on this, but given tenure and other ex post wedges, senior teachers especially may have incentives to push for deviations from the optimum.

A recent study[31] found that reducing class size improved student performance, but adding a teacher's aide to the classroom had no effect on performance. This is consistent with the model under a couple of interpretations. If disruption occurs for an entire classroom when one child is disruptive, then the fact that there is a teacher's aide in the class may have little effect. Two teachers in the room may not alter the fact that learning occurs only p^n of the time.[32] Another interpretation is that teachers substitute teacher aide time for their own. The teacher's aide may be a benefit to the teacher, rather than to the student. Total teaching time might increase some, but not enough to be discernable if teachers withdraw almost the same amount of time as aides add.[33]

Even if markets fully internalize the non-monetary tradeoffs in class size, there is another reason why teachers as a group might prefer smaller class sizes. A reduction in class size, coupled with compulsory schooling, implies an increase in the demand for teachers. The recently mandated class size reduction in California meant a significant increase in demand for teachers. Some was probably reflected in increased salaries, some in ease of obtaining a teaching job and some in a lowering of standards for teachers. It is not surprising that an organized group would push for policies that increase the demand for its services.

---

[31]See Krueger (1997, 1998).

[32]To the extent that crowding effects reflect the need for individual attention, rather than disruption, then a teacher's aide would be expected to have positive effects on performance.

[33]University professors may substitute research time, not additional teaching time, when they are given a teaching assistant who assists in grading and other chores.

Homework:

The difference between homework and classwork is that neither positive nor negative externalities are present when homework is done. Homework can be viewed as a substitute for classwork, but in equilibrium, it may also be a complement.

Because homework involves no interaction, one might predict that the higher is p, the higher would be the ratio of class time to home time. But the ratio of class time to home time spent on schooling is higher for younger children than it is for older ones despite the fact that p is lower for younger children. The explanation must result from an interaction between age and the efficiency of learning under different technologies. The probability that an undisturbed six-year-old learns on his own is smaller than the probability that an undisturbed sixteen-year-old learns on his own. Additionally, children of different ages acquire different skills at school. Most of what college students learn is cognitive; a large part of what elementary school students learn is social. The cognitive component is better learned at home than is the social component. These complications are ruled out by assumption in the model, which expresses output in efficiency units.[34]

Teacher Performance, School Quality and Incentives:

Private schools charge a price that depends on the quality of the education that it produces. While other factors may affect the price (such as charity to needy students), a school that competes in a private market is constrained to charge a price that is commensurate with the quality of the

[34]This is no more nor no less abstract than standard human capital analysis. There, human capital is treated as the generic output of schooling. An extension of this model could allow for different types of learning, with different p and q values.

education it produces. Even though schools are non-profit organizations, the revenue that is generated by an improvement in quality of the product goes directly to the school in question. This, in turn, makes the jobs of administrators easier. Also, workers in non-profit organizations may capture at least some of the returns to higher net revenue. The wages of administrators and teachers are likely to be related to the revenue of the school.[35]

Public schools are different. If a district offers higher quality schooling, the value of land in that district rises, but the school is unlikely to capture much of the revenue from improved quality. Even if tax revenue goes up as a result of increasing land prices, only a fraction of the tax revenue may be passed to the schools in general and the school that actually generates the additional revenue is unlikely to capture much of the return. The more centralized is the school system, the less direct impact on individual schools.[36] As a result of the centralized revenue pool and decentralized work decisions, there may be more incentive slippage in the public schools than in the private ones.[37]

Wages may be somewhat responsive to firm profitability and to individual performance in private schools, but are almost totally divorced from productivity considerations in the public

---

[35]In a completely competitive labor market with free mobility, there is no necessary relation between profits and wages. To the extent that there is an attachment that is firm specific, there can be some ex post rent sharing.

[36] States differ in this regard. In California, property taxes are collected and distributed at the state level subject to rigid egalitarian formulas. In Illinois, taxes are county based and revenues that flow to schools are more directly tied to the specific county's tax revenue.

[37]See Hoxby (1998a). Also, it is conceivable that the same problem could plague private schools that were part of a chain. Just as a single McDonald's fails to take into account the effect of its actions on McDonald's brand name, so too might one branch of a large system of private schools ignore the impact of its actions on the value of the system.

schools. Wage setting is done centrally and the group of government workers, to which teachers and school administrators belong, have wages that are based on rigid grade and step scales. Finally, private schools can go out of business more easily than can public schools, although this should not be overemphasized because public schools can and frequently do close their doors.

The consequence of these differences is to create incentives in public schools that differ from those in the private ones. It can be argued that limitations on the ability to take compensation directly in the public sector induce officials to take their compensation in other forms.[38] Just as contractors shade quality when they are paid a fixed price for a construction job, so too might school administrators shade effort when they are paid a fixed wage for a job that is independent of quality. If the wage in the public sector is less responsive than the wage in the private sector, then incentives to perform are likely to be weaker in public schools.

The institution of tenure, granted to most public and many private school teachers, is often criticized as an extreme manifestation of the unresponsiveness of compensation to performance. The reality is that teachers are only somewhat more protected than other workers from layoff. Even in the non-education sector, once an individual has been with the firm for a few years, the probably of a layoff is very low.[39]

It is possible to overstate the differences in incentives between public and private schools, which share some common attributes. First, there is a market for successful administrators and

---

[38]See Alchian and Kessel (1960).

[39]An early study documenting the decline in the turnover hazard with tenure is Mincer and Jovanovic (1981).

teachers that operates in both public and private schools. Second, the incentives to keep a constituency happy are present in both sectors. Parents and students can make their displeasure known as readily in the public sector as in the private one. Parental unhappiness may have less impact on school revenue in the public sector than in the private sector, but this does not negate the fact that teachers and administrators must deal with complaints in both sectors. Third, just as trustees of private schools can reward administrators who do well, so too can public officials reward administrators of public schools through promotions and other non-monetary benefits. Finally, pride in a job well done may motivates educators and there is no obvious reason why this would differ between sectors.

An Empirical Strategy:

As already mentioned, p, the fraction of the time that a student is not an initiator of disruption, is not a mere abstraction, but is observable. One could imagine obtaining information on p by surveying teachers or by actually observing a classroom. Given p, quality of education should vary with p^n. Since n is also observable, p^n can be thought of as a measure of quality that is different from educational expenditures, used by others.[40] A year of adjusted schooling could be defined as a year, multiplied by p^n.

The model has very specific predictions about the relation of n to p. The first order condition in (2) implies that class size should be smaller, the larger is p. This is testable using data such as TIMS, where the classroom experience is taped, reviewed and graded.

[40] See, for example, Card and Krueger (1992).

Operationally, it is probably easier to observe p^n than p. An alternative to direct viewing of classrooms is surveying the teachers on the proportion of their class time spent in actual teaching versus discipline or disruption. This provides an estimate of p^n ,which when coupled with information on n, provides an estimate of p. For the purposes of quality adjustment, p^n by itself is all that is of interest. For normative purposes, for example, determining the optimum class size by grade level, p is useful. It might also be of interest to compare student characteristics with p. How does p vary with age, socio-educational background and parent's income? Understanding variations in p may provide some implications for school reform.

The Division of Labor is Limited by the Extent of the Market:

Class size may be limited by another factor, namely the extent of the market. Consider, for example, a small college that has students in both literature and economics. Suppose that the college has 100 students, half of whom are in each field. On the basis of congestion effects, the condition in (2) might imply an optimal class size of 100. This is infeasible because only 50 want to study each field. Heterogeneous study preference limit class sizes beyond congestion considerations. It is for this reason that schools do not put high school seniors in the same class with first-graders. Absent other constraints, one would conclude that a school of 100 is too small. Because class sizes must be below the efficient number of students, a merger of two small schools into one larger one could achieve both division of labor and class size efficiency. To the extent that a desire for neighborhood schools or other preferences limit the size of a school, heterogeneous learning preferences limit the size of classes.

That having been said, a limited extent of the market cannot be used to explain a number of facts that are consistent with the congestion hypothesis. First, congestion implies that schools may have 4 identical classes of 30 students at a particular grade level, whereas heterogeneous preferences would imply one class of 120. Second, heterogeneity would argue for smaller classes among older students, not the reverse. The curriculum for pre-school students is more homogeneous across students than is the curriculum for high school students who are following different paths. Pre-school classes are smaller than high school classes, not because different courses are being offered, but because things get out of control when there are more than ten pre-schoolers in the same place. Similarly, special education classes are small not because the course material is so varied, but because p is low for special students.

Conclusion:

Classroom teaching is a public good. As such, congestion effects can be important. A student who is disruptive or who takes up teacher time in ways that are not useful to other students affects not only his own learning, but that of other in the class. It is for this reason that class size may have important effects on educational output. Much of the empirical evidence, however, suggests otherwise. Class size effects are small or non-existent in most studies.

A theory of educational production, with particular emphasis on classroom dynamics has been presented. Because schools have choice over class size, and because optimal class size varies directly with the quality of the student, observed class size effects are small or sometimes positive: Larger classes produce higher educational output per student. It is for this reason that natural experiments tend to find benefits to class size reductions whereas other analyses do not. The natural experiments do not have the problem that optimal class size varies directly with student behavior.

The theory provides implications for class size and its effect on total output. The analysis predicts variations in class size by grade level and by other student and teacher characteristics. Among the findings are that in equilibrium, class size matters very little. To the extent that class size matters, it is more likely to matter at lower grade levels than upper grade levels where class size is smaller.

The technology implies that class size effects are more pronounced in smaller classes and for lower values of p. This result is consistent with the finding that class size reductions provide better results for disadvantaged and special needs children.

Discipline is a substitute for class size. Educational output per student remains constant when class size is increased by a factor of k so long as the proportion of the time that students behave rises from p to $p^{1/k}$.

Under most circumstances, segregating students by academic ability maximizes total educational output. Self-selection induces the more attentive students to attend private schools. This point and the one on the effect of discipline may explain why Catholic schools out-perform public schools, despite both larger class sizes and lower expulsion rates in Catholic schools.

Teachers may prefer smaller class size than students or parents either because wages do not reflect working conditions fully or because teachers as a group can raise the demand for their services by lowering class size.

Finally, because public schools that operate in a centralized environment do not capture the returns to their successes, public school incentives differ from those of private schools. These differences do not imply that administrators and teachers in public schools are without incentives; many forces operate similarly across sectors. There are, however, some incentives which operate in the private sector that are absent in the public sector.

References

Angrist, Joshua and Victor Lavy. "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement," NBER Working Paper #5888, January, 1997.

Antos, Joseph R., and Rosen, Sherwin. ``Discrimination in the Market for Public School Teachers," *Journal of Econometrics* **3** (May 1975): 123-50.

Alchian, Armen, and Kessel, Reuben. ``Competition, Monopoly, and the Pursuit of Pecuniary Gain," *Aspects of Labor Economics* (1960): 156-183. YA 2070.

Becker, Gary S. *A Treatise on the Family.* Enlarged Edition. Cambridge: Harvard University Press, 1991.

Becker, Gary S. and Kevin M. Murphy. "The Division of Labor, Coordination Costs, and Knowledge," *Quarterly Journal of Economics,* CVII, 4, (November, 1992), 1137-1160.

Becker, Gary S. and Kevin M. Murphy, *Social Markets and Social Economics,* forthcoming, Cambridge: Harvard University Press, 2000.

Bedard, Kelly, William O. Brown Jr., and Eric Helland. "School Size and the Distribution of Test Scores," unpub. manuscript, Claremont McKenna College, 1999.

Betts, Julian. "Is there a link between school inputs and earnings? Fresh scrutiny of an old literature," in *Does money matter? The effect of school resources on student achievement and adult success*, ed. by Gary Burtless. Washington: Brookings Institution, 1996, pp 141-91.

Blake, Howard. "Class Size: A Summary of Selected Studeies in Elementary and Secondary Public Schools," doctoral dissertation at Columbia University, 1954.

Caucutt, Elizabeth M. "Peer Group Effects in Applied General Equilibrium," University of Minnesota Center for Economic Research Discussion Paper: 293, November 1996.

Card, David and Alan Krueger. "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States," *J. Polit. Econ.* 107, February, 1992, 1-40.

Coleman, James et al., *Equality of Educational Opportunity*, Washington 1966.

Coleman, James, and Hoffer, Thomas. *Public and Private Schools: the Impact on Communities*. New York: Basic Books 1987.

Coleman, James, Kilgore, Sally, and Hoffer, Thomas. *Public and Private High Schools*. Washington, D.C.: National Center for Educational Statistics 1981.

Epple, Dennis and Romano, Richard E. "Competition between Private and Public Schools, Vouchers, and Peer-Group Effects," *American Economic Review*, v88 n1 March 1998.

Evans, William N. and Schwab, Robert M. "Finishing High School and Starting College: Do Catholic Schools Make a Difference?" *Quarterly Journal of Economics*, v110 n4 November 1995.

Fernandez, Raquel and Rogerson, Richard. "Public Education and Income Distribution: A Dynamic Quantitative Evaluation of Education-Finance Reform," *American Economic Review*, 88,4, September 1998.

Flyer, Frederic and Sherwin Rosen. "The New Economics of Teachers and Education," *Journal of Labor Economics*, vol. 15, number 1, part 2, (1997), S104-S139.

Hanushek, Eric A. "The Evidence on Class Size," Occasional Paper Number 98-1, W. Allen Wallis Institute of Political Economy University of Rochester, February 1998a

Hanushek, Eric A. "Conclusions and Controversies about the Effectiveness of School Resources," Federal Reserve Bank of New York Economic Policy Review, v4 n1 March 1998b.

Hanushek, Eric, John F. Kain and Steven G. Rivkin, "The Cost of Switching Schools," unpub. manuscript, April, 1999. Presented at 1999 Meetings of the Society of Labor Economists, Boston, May, 1999.

Heckman, James J. "Doing it right: job training and education," *The Public Interest*, 135, (Spring, 1999), 86-107.

Heckman, James J. "Rethinking Myths About Education and Training; Understanding the Sources of Skill Formation in a Modern Economy," David Kinley Lecture and Presidential Address to the Midwest Economics Association, March, 1998.

Henderson, Vernon, Mieszkowski, Peter, and Sauvageau, Yvon. "Peer Group Effects and Educational Production Functions," *Journal of Public Economics*, v10 n1 Aug. 1978.

Hoxby, Caroline M. "What Do America's "Traditional" Forms of School Choice Teach Us about School Choice Reforms?" Federal Reserve Bank of New York Economic Policy Review, v4 n1 March, 1998. (1998a)

Hoxby, Caroline M. "The Effects of Class Size and Composition on Student Achievement: New Evidence from Natural Population Variation," mimeo, Harvard University, December 1998b.

Hoxby, Caroline M. "How Teachers' Unions Affect Education Production," *Quarterly Journal of Economics*, 111,3, August, 1996, 671-718.

Kremer, Michael. "The O-Ring Theory of Economic Development," *Quarterly Journal of Economics*, v108 n3 August 1993.

Krueger, Alan B. "Experimental Estimates of Educational Production Function," NBER Working Paper no. 6051, June 1997.

Krueger, Alan B. "Reassessing the View That American School Are Broken," Federal Reserve Bank of New York Economic Policy Review, v4 n1 March 1998.

Lazear, Edward P. "Culture and Language," *Journal of Political Economy*, December, 1999.

McMillan, Robert. "Parental Pressure and Competition: An Empirical Analysis of the Determinants of Public School Quality," mimeo Stanford University, January 1999.

Mincer, Jacob, and Jovanovic, Boyan. ``Labor Mobility and Wages," in *Studies in Labor Markets*, 21-64, ed. Sherwin Rosen. Chicago: University of Chicago Press for NBER, 1981.

Neal, Derek. "The Effects of Catholic Secondary Schooling on Educational Achievement," *Journal of Labor Economics*, v15 n1 Part 1 January 1997.

Neal, Derek. "What Have We Learned about the Benefits of Private Schooling?" Federal Reserve Bank of New York Economic Policy Review, v4 n1 March 1998.

Necheba, Thomas J. "Mobility, Targeting and Private School Vouchers," unpub. manuscript, Stanford University, 1999.

Olson, Craig A. and Deena Ackerman, "High School Inputs and Labor Market Outcomes for Male Workers in Their Mid-Thirties: New Data and Estimates from Wisconsin," unpub. manuscript, Univ of Wisconsin, 1999. Presented at 1999 Meetings of the Society of Labor Economists, Boston, May, 1999.

Rees, Daniel I. and Argys, Laura M.; Brewer, Dominic J. "Tracking in the United States: Descriptive Statistics from NELS," *Economics of Education Review*, v15 n1 February 1996.

Robertson, Donald and Symons, James. "Self-Selection in the State School System," London School of Economics Center for Economic, Performance Discussion Paper: 312 November 1996.

Rosen, Sherwin. "Some Economics of Teaching," *Journal of Labor Economics*, vol 5, number 4, part 1, (1987), pp. 561-75.

Rothschild, Michael and Lawrence White. "The Analytics of the Pricing of Higher Education and Other Services in Which the Customers are Inputs," *Journal of Political Economy*, 3, 103 (June 1995), pp. 573-586.

Sander, William. "Catholic High Schools and Rural Academic Achievement," *American Journal of Agricultural Economics*, v79 n1 February 1997.

Summers, Anita A. and Wolfe, Barbara L. "Do Schools Make a Difference?" American Economic Review, v67 n4 Sept. 1977.

Urquiola, Miguel "Demand Matters: School District Concentration, Composition, and Educational Expenditure," UC Berkeley: Center for Labor Economics, Working paper #14, April, 1999.

Appendix:

To show: $\dfrac{\partial p^{1/m(p)}}{\partial P} > 0$

where m(p) is the function derived from the first-order condition that gives the optimal m as function of p. (Note: V is normalized to 1).

Proof:
Rearranging terms of the (2), one obtains

$$p^{1/m(p)} = \frac{-Wm^2(p)}{1n\, p} \qquad (A1)$$

The second-order condition is:

$$2p^{1/m} 1n\, p \frac{1}{m^3} + p^{1/m} 1n^2 p \frac{1}{m^4} < 0 \qquad (A2)$$

which implies

$$2m(p) + 1n\, p > 0 \qquad (A3)$$

for an interior solution to be optimal.

Define $D_0 \equiv \dfrac{\partial p^{1/m(p)}}{\partial p}$.

Then, D_0 can be written as

$$D_0 \equiv \frac{\partial p^{1/m(p)}}{\partial p} = \frac{p^{(1/m(p)-1)}}{m^2(p)}(m(p) - pm'(p)\ln p) \qquad (A4)$$

Combine (A1) and (A4) to obtain

$$D_0 = \frac{-W}{p \ln p}(m(p) - pm'(p)\ln p) \qquad (A5)$$

Differentiate (2A) w.r.t. p to obtain

$$-\frac{\partial}{\partial p}(p^{1/m(p)})\ln p - \frac{1}{p}\, p^{1/m(p)} - 2Wm(p)m'(p) = 0 \qquad (A6)$$

Using (A2), (A8) and the definition of D_0 , it follows that

$$-D_0 \ln p + \frac{Wm^2(p)}{p \ln p} - 2Wm(p)m'(p) = 0 \qquad (A7)$$

Substitute $m'(p)$ from Equation (A6) it into (A5) to obtain

$$pm'(p)\ln p = \frac{-D_0\, p \ln^2 p + Wm^2(p)}{2Wm(p)} \qquad (A8)$$

Next, substitute (A8) into (A5) to obtain:

$$D_0 = \frac{-W}{p\ \ln p}(m(p) + \frac{D_0 p\ \ln^2 p\ -\ Wm^2(p)}{2Wm(p)}) \qquad (A9)$$

Rearranging the terms in (A9) yields

$$D_0\ p\ \ln p = -Wm(p) - \frac{D_0 p\ \ln^2 p - Wm^2(p)}{2m(p)}$$

so that

$$D_0(p\ \ln p + \frac{p\ \ln^2 p}{2m(p)}) = -Wm(p) + \frac{Wm(p)}{2}$$

or

$$D_0 \frac{2p\ \ln p\ m(p) + p\ \ln^2 p}{2m(p)} = -\frac{Wm(p)}{2}$$

and finally, that

$$D_0 = -\frac{Wm^2(p)}{p\ \ln p(2m(p) + \ln p)}$$

The fact that p<1 combined with (A3) imply that $D_0 > 0$. |||