Discussion

IMPROVING EDUCATIONAL QUALITY: HOW BEST TO EVALUATE OUR SCHOOLS?

Peter J. Dolton*

Hanushek and Raymond have provided us with an elegant and coherent set of arguments for accountability in the public provision of education. Their main empirical contribution is the characterization of the different types of accountability systems in the United States. They go on to show how the operation of incentives in the public sector is not easy to administer and evaluate. Specifically they draw attention to the distinction between inputs, process variables, and outcomes, and they show how incentive systems reliant on input and process measurements may be ineffective. The authors find empirically that the presence of an accountability system leads to modest growth in achievement but caution that evidence needs to be treated carefully to recognize the possibility of gaming and the consequent interpretation problems. In short, they suggest that accountability incentives matter.

This discussion of their paper recaps the problems with incentive structures in public provision of education and raises the issue of exactly what is meant by accountability. I also consider how the effect of accountability on achievement measurement and on performance should be assessed. Then I will provide examples of the problematic working of incentives from the U.K. education system.

WHAT ARE THE PROBLEMS WITH INCENTIVE STRUCTURES?

There is now widespread evidence that incentives work in the public sector. The issue is designing incentive structures that are not subject to distortion or "gaming." The education production process is very reliant

*Professor of Economics, University of Newcastle and University of London.

on teacher labor as the most important factor of production. In practice, it is very difficult to write complete labor contracts in education to generate the appropriate incentives for teachers. To a large extent this is a principal/agent problem. However there are several extra dimensions to this problem in education.

The literature (Dixit 2002 and Burgess and Metcalfe 1999) suggests that using incentive structures in the public sector could induce dysfunctional behavior in the sense that employees could direct their effort on some aspects of their work, to the detriment of other aspects, or in a counterproductive way, when teamwork or the cooperation of colleagues is involved.

The essential problem of public sector educational provision is that education is not a single output, and any education system must have multiple goals. Dixit (2002) lists the multiple goals of public education as the following:

1. Imparting basic skills of literacy, mathematics, and science for communication, reasoning, and calculation;
2. Fostering the emotional and physical growth of children;
3. Preparing students for work, by teaching them vocational skills and attitudes suitable for employment;
4. Preparing them for life, by teaching them skills of health and financial management;
5. Preparing them for society, by instilling ideals of citizenship and responsibility;
6. Helping them to overcome disadvantageous circumstances at home, including in many cases poor nutrition and poor study environments; and
7. Providing an environment free from drugs and violence.

Dixit suggests that although these goals are not mutually contradictory, they do compete for resources. To this degree they are alternative outputs in the educational production process, and teacher effort put into one of these objectives may detract wholly, or in part, from one or more of the other goals. Hanushek and Raymond view educational production in much narrower terms, as very few of these goals appear on their list of accountability variables.

The essential problem of education is that with multiple goals, it is unclear how to direct effort. Holmstrom and Milgrom (1991) develop a model that explains that the way incentives work may not be appropriate, even when accurate performance measures are available. They extend the standard principal/agent model to one in which there are several dimensions to effort. The general result is that the agent will have an incentive to divert effort away from the less accurately measured task. Hence, it is shown that if the principal wishes the agent to allocate effort

towards a task that is not easily measured, then incentives on the measurable tasks must be weakened.

The second essential feature of any education system is that it has multiple principals. As a consequence, the actions of any individual teacher (agent) could be affected by many other people (principals) who are in a position of influence. Most specifically the wishes of parents, headteachers, teacher unions, local or federal authorities, taxpayers, employers, religious and ethnic pressure groups, governors, and even pupils may influence the actions and decisions of individual teachers. However, Dixit (1997) shows (under regularity conditions) that the existence of several principals makes the overall incentives for the agent much weaker. This weakening of incentives occurs because each principal will seek to divert the agent's effort to his most preferred dimension. Obviously the more principals that are involved with competing interests the more diluted will be the incentive structure for the agent. Hanushek and Raymond do not discuss the multiple principal incentive problem or its implications for accountability.

WHAT DO WE MEAN BY ACCOUNTABILITY?

The concept of accountability is a difficult one. Fearon (1999) suggests that "one person, A (the agent), is accountable to another B (the principal), if two conditions are met. First, there is an understanding that A is obliged to act in some way on behalf of B. Second, B is empowered by some formal institution or perhaps informal rules to sanction or reward A for her activities or performance in this capacity" (p. 55).

Laver and Shepsle (1999) provide a different definition. They suggest that a political agent is accountable to a principal when the principal, having the means to do so, has no inclination to replace the agent with a feasible alternative. Hence Laver and Shepsle view accountability as both an equilibrium state and a mechanism for change.

Ferejohn (1999) suggests there are three serious limits to accountability. First, the nature of the accountability mechanism (voting rule) may mean that minorities are ignored or indeed that electoral heterogeneity makes it possible for officials to play off some voters against others. Second, the institutions of accountability operate in real time—and this provides the officials with the opportunity to avoid responsibility. Third, officials typically enjoy an immense informational advantage over consumers.

Hanushek and Raymond do not have a clear definition of what they mean by accountability. They suggest that accountability creates incentives—although they recognize that such incentives may not always have desirable consequences. What is unclear in their exposition is whether they believe the mere publication of information on standards in schools

will provide an adequate incentive for efficient resource allocation. Surely a necessary (but not sufficient condition) for such efficiency is that this accountability be directly linked to a quasi market, or the power of consumers to choose alternative providers in a competitive market.

Hence what I am suggesting is that effective accountability in education necessitates, first, that the education system provide consumers with full information to make decisions; second, that consumers have the power to influence the balance of priorities across the multiple goals of educational provision; third, that consumers have the means to choose alternative providers in a competitive or quasi-competitive environment; and, fourth, that any incentives that operate on education providers do not act to distort their incentives regarding their provision in ways that are counter to the wishes of consumers.

A definition of accountability that includes customers' wishes relies on being able to identify who these customers are; identifying correctly what their views are; aggregating their views into a consensus to establish what the ranking of priorities is; and then implementing these views effectively. Any decisions concerning public expenditure and investment in education constitute a social-choice problem (see Majumdar 1983). This issue is rarely examined.

We also have to assume that the views of parents are responsible and representative of the whole customer base. Such an assumption may be unrealistically ideal. Aoki and Feiner (1996) discuss how the parents whose views are more effectively heard are disproportionately those who live in affluent areas, are more highly educated, and have higher-status occupations. Such evidence means that establishing precisely who an educational system is accountable to, and what the mechanism is for the transmission of the influence, is important. Most concretely, are the customers of education the parents or the pupils? Undoubtedly the priorities of the pupils, if consulted, may be different from their parents'.

At the heart of effective public service provision is the possibility of competition among providers. Unless there are alternative schools for parents to send their children to, there is no incentive mechanism for each school to compete in the quasi market. Another problem with this model arises if there are private schools outside the public sector. Friedman (1962) advocates a voucher system in which essentially all schools would be private. In the context of the present system where private and state schools operate in the same area, there is the "exit and voice" issue, which states that there will not be an effective mechanism for change if the most influential parents choose to "exit" from the state schools to the private schools rather than "voice" their views in an attempt to change the state schools. More research is necessary into how the public sector in education can efficiently co-exist alongside a private sector (see Hirschman 1970).

How Can Educational Output and Efficiency be Measured? What Is the Effect of Accountability?

Hanushek and Raymond show that most states rely on student performance as the main outcome measure. They come to the conclusion that the "individual-gain score" (or value-added) measure of student performance between years is the most valid. It is worth reviewing exactly what such measures can and cannot say. Todd and Wolpin (2003) provide the most general discussion of the assessment of educational evaluations and rigorously discuss modeling using the value-added method. To fix ideas, we consider that pupil attainment is determined by a production function relation, and we may use the following notation:

A_{ijkt} attainment of pupil i, in class j, in school k, at the end of time period t[1]

X_{it} characteristics of pupil i at time t that may affect attainment

S_{ijkt} resources of pupil i's class j, in school k, at time t

F_{it} family resources devoted to pupil i at time t

μ_i innate ability endowment of pupil i

Assuming that educational attainment of the pupil is a function of individual attributes and ability, school inputs, and family inputs, we may write the general production function type model of what determines pupil attainment as:

$$A_{ijkt} = g(X_{it},\ S_{ijkt},\ \Sigma\ F_{it},\ \mu_i). \tag{1}$$

Simplifying this production relation to consider the influence on initial attainment, prior to school, we suggest that

$$A_0 = g_0(X_0, F_0, \mu), \tag{2}$$

where we are dropping the i subscript for the individual. In each subsequent period the family adds more input based on its decision process, and the school contributes resources, S, in the manner suggested by the production function. The schooling input decision for any pupil will be determined as a result of the pupil's ability and prior attainment, that is,

$$S_{ijkt} = \varphi(A_{ijkt-1}, \mu_i). \tag{3}$$

We can write the production function (1) as an econometric model for period 1 as:

---

[1] For notational convenience we will think of pupils' attainment being tested at the end of each school year t, so that A_i represents the attainment acquired in that year.

$$A_{ijk1} = \alpha_1 X_{i1} + \beta_1 S_{ijk1} + \delta_1 F_{i1} + \gamma_1 \mu_i + \epsilon_{i1} + u_{j1}, \tag{4}$$

where α, β, δ, and γ are parameters and ϵ_i and u_j represent unobserved heterogeneity at the individual and school level. Likewise for period 2 we can write,

$$A_{ijk2} = \alpha_2 X_{i2} + \beta_2 S_{ijk2} + \delta_2 F_{i2} + \gamma_2 \mu_i + \epsilon_{i2} + u_{j2}. \tag{5}$$

If we take the difference between (5) and (4), we get an expression for the fixed effects estimator that is so common in econometric applications. Writing this as:

$$\Delta A \equiv A_{ijk2} - A_{ijk1}, \tag{6}$$

we can see that this is equivalent to

$$\Delta A = \alpha_2 X_{i2} - \alpha_1 X_{i1} + \beta_2 S_{ijk2} - \beta_1 S_{ijk1} + \delta_2 F_{i2} - \delta_1 F_{i1}$$
$$+ \gamma_2 \mu_i - \gamma_1 \mu_i + \epsilon_{i2} - \epsilon_{i1} + u_{j2} - u_{j1} \tag{7}$$

The question is, under what circumstances is the individual student-gain estimator a valid estimate of pupil progress? We place some restrictions on this model to make explicit the necessary assumptions:[2]

- **A1. The pupil attributes, X_i, remain constant across time.** This means we can write: $\alpha_2 X_{i2} - \alpha_1 X_{i1} = \alpha X_i$, where $(\alpha_2 - \alpha_1) = \alpha$ and $X_{i1} = X_{i2}$. This is a restrictive assumption since it means that variables that represent motivation and effort, like propensity to complete homework, remain fixed. This is clearly wrong, as such attributes are often age-related for the pupil.
- **A2. There exists a sufficient statistic for the changing value of school inputs that is observable and that school effects are time invariant.** This means we can write $S_2 - S_1 = S$. Then we can write the school effects term as βS.
- **A3. The change in parental input can be proxied by some observable family characteristic F.** This is equivalent to assuming that $F_2 = F_1$ and $\delta_2 - \delta_1 = \delta$ and hence the family effects can be represented by δF. Although naive and restrictive, it is unlikely that values of family inputs will be observed at different points in time.

---

[2] It should be appreciated that more than one set of assumptions can be made in order to make this model useful with the data.

A4. **The impact of ability endowment on achievement is indepen-
dent of time.** Hence $\gamma_2 = \gamma_1$ and so the unobservable ability term
can be netted out. This would suggest that the importance of the
application of innate ability is not age-specific. Again, this is
naïve, as maturity may well affect the pupil's potential for
attainment.

A5. **Input choices made by schools and parents are invariant to
prior achievement outcomes.** Thus S_1 and F_1 are uncorrelated
with A_0. Taking A1 through A5 together we can now rewrite (7)
to give:

$$\Delta A = \alpha X_i + \beta S_{ijk} + \delta F_i + \epsilon_i + u_j. \qquad (8)$$

As restrictive as the above assumptions are, the possible estimation of
equation (8) still has very demanding data requirements.

The problem of Hanushek and Raymond is more complex. They
wish to establish the effect of introducing accountability on student
achievement. One approach would be to estimate equation (8)—possibly
aggregated at the level of the school or the state, splitting the sample
according to whether the state operated with accountability. This would
require that the decision to invoke accountability was exogenous to
achievement. Their adopted approach, due to inadequacies in their data,
is less sophisticated. They use state-level data with no information on
school or family resource decisions. They use their own measure of
accountability as a regressor into equation (8). The question is, under
what circumstances is it valid to assume that such a regressor is
exogenous with an additively separable effect? Clearly one would expect
the level of school resources and family factors to be affected by the level
of accountability.

There are other subtle ways in which the inputs and outputs of the
education production process are difficult to observe. The raw material,
or input, a teacher works with is highly variable. It is well known that
teaching the same material to children from poor homes in deprived
areas is more difficult than teaching to motivated children from middle-
class homes. Even if one tries to measure value-added in terms of
improvement of exam scores, these can be a distortion of the improve-
ment in attainment as such a calculation assumes that other factors and
their influence are fixed over time. There is often a huge variation in the
resources at the school's disposal—many of which are not easily mea-
sured.

A second often-overlooked issue is the measurement of peer effects
in schools (see Lazear 2001). It is possible that some of the results relating
to the absence of a pupil-teacher ratio resource effect (see Hanushek 1997
and Burtless 1996) may be due to ignoring peer effects. Indeed, one study
in the United Kingdom shows this dramatically with a complete change

in the resource coefficient when peer-group effects are proxied. (See Dolton 2002a.)

A third important limitation of achievement gains models is that a pupil's learning may not be apparent until years after his schooling. Often the value of what is learned by the pupil is not used or tested until several years after. As Hanushek and Raymond point out, one further important limitation to the value-added model is that achievement gains cannot be identified for those who move geographical location.

Finally, it is not impossible that teachers (or principals other than government) may view educational output differently from the government. Teachers may want to promote curiosity, induce creative thinking, provide pastoral care, and develop a wider curriculum. The government may prefer to structure the curriculum, standardize teaching methods, meet minimum standards on basic skills, and maximize performance on SAT test scores. The wider benefits of learning are rarely added into achievement-gain calculations.

Incentives and Quasi Markets in U.K. Schools

There has been a major shift in the way in which public sector education has been provided in the United Kingdom over the last 20 years. The educational system has changed to one dominated by incentive structures and quasi markets. These changes have produced a revolution in state educational provision. The results and consequences so far have been mixed. I will highlight how several of these quasi markets have been working, including some of their unintended consequences.

In the United Kingdom the Education Reform Act was passed in 1988. The general aim of the reform was to introduce a more competitive quasi-market approach to the allocation of resources in the education system. It introduced financial delegation to schools, and this involved the introduction of "formula funding" in which school income is based directly on pupil numbers. The Act insisted on the publication of school league tables and introduced the principle that parents had the right to send their children to any school they wished. The idea was that popular schools were allowed to expand without limit and conversely unpopular schools, mostly in inner cities, to contract or even close. The principles of parental choice and devolved school funding linked directly to pupil numbers establishes the conditions under which—theoretically—a quasi market can operate. This approach was designed to provide teachers and schools with appropriate incentives for efficiency and effectiveness. Although requiring schools to live within their budgets, this approach does not provide the same incentives for employees as knowing that their efforts contribute to the profit "bottom line" of a firm.

One clear feature of the state education system in the United

Kingdom is that there is a lack of competition. State schools in the United Kingdom, in many areas, operate essentially as monopoly providers. Only around 7 percent of school children in the United Kingdom attend independent schools. Because of the scale of their fees, these independent schools do not present a realistic alternative to state schools for most parents. It was this lack of competition that was part of the rationale for the 1988 Education Act providing parents with choice. The central idea behind the creation of a quasi market in state education is the theory that the introduction of competition would provide the appropriate incentives to schools to become more efficient. Theoretically this, in turn, may provide incentives for teachers to improve their performance. However, this naive faith in the power of market forces must be tempered by the reality that multiple tasks and multiple agents will weaken the power of such incentive structures.

Empirical evidence from the United States (Chubb and Moe 1990) supports the view that decentralized schooling systems produce better results, measured in terms of educational outcomes. The 1988 Act also devolved the administrative and financial control of schools to each headteacher and the school's governing body. The governing body was also to have representation from parents.

Bartlett (1993) reports that the effect of the reform has been a large shift in the distribution of resources between schools. Schools in the poorest inner-city areas have received reduced funding while funding has increased for schools in the more prosperous areas of the country. Likewise, the appointment of proactive parent governors in middle-class areas is straightforward but finding any parents willing to do the job in deprived areas is difficult. Overall, the effect of the quasi-market reforms on educational outcomes and efficiency in the United Kingdom is hard to judge, not least because there are several initiatives acting on the market at the same time. Nevertheless, there are some microeconometric studies that suggest that efficiency improvements can be directly attributed to the quasi market (Bradley, Johnes, and Millington 2001).

In reality, access to oversubscribed schools remains rationed with some selectivity and "cream-skimming" operations. This has been reflected in the market-clearing mechanism of rising house prices in localities with the best performing schools. (See Gibbons and Machin 2002).

Since 1995, the government has published school league tables of the results of all schools in the United Kingdom based on national examinations for pupils aged 7, 11, 14, 16, and 18. Some commentators, for example, Glennerster (2002), have suggested that these results show how educational standards have improved in the United Kingdom over the last six years. Table 1 shows a remarkable rise in the performance of 14-year-olds in the United Kingdom on reading, math, and science. The proportion reaching the expected standard in reading has risen from 49

Table 1
United Kingdom National Achievement Tests at Level 3, Aged 14
Percent Reaching Expected Levels

	1995	1996	1997	1998	1999	2000	2001
Reading	49	57	67	71	78	83	81
Math	45	54	62	58	69	72	70
Science	70	62	69	69	79	85	87

Source: Glennerster (2002), Table 6.

percent in 1995 to 81 percent in 2001. In math, the proportion has risen from 45 percent to 70 percent; and science has jumped from 70 percent to 87 percent over the same period. Such statistics raise the following questions:

- To what extent are these tests based on absolute standards that have not been manipulated by a government that has declared, as if by decree, that educational standards will rise over the next five years? Alternatively, have the exams become easier or have the pupils improved their performance over time because of the predictable nature of the exams and rote learning?
- To what extent has there been misallocation of resources towards median and marginal pupils at the threshold of achievement levels in order to maximize the number of pupils passing the thresholds?
- Has the introduction of these tests diverted resources away from the least able and most able, towards the average child?
- If the improvement has been real—is it really a treatment effect that results directly from the operation of the quasi market rather than a redirection of effort on literacy and numeracy in the curriculum?
- Are the long-term consequences of increasing marginal standards on narrowly focused tests in math and English valuable for long-term educational objectives like citizenship and transferable skills?
- Is it possible to reconcile these data with results from Gundlach, Wöessman, and Gmelin (2001), who suggest that the United Kingdom along with other OECD countries has experienced a dramatic fall in school productivity over the last 25 years?

POSTSCRIPT: A LESSON FROM HISTORY?

The attempt to introduce incentives and monitoring into schools in order to secure their efficiency and make the best use of public money is not new. Neither is the possibility that such attempts may lead to counterproductive consequences of these incentives.

In 1857, the Newcastle Commission surveyed schools in Great Britain and recommended "a searching examination by competent authority of every child in every school to which grants are to be paid with a view to ascertaining whether these indispensable elements of knowledge are thoroughly acquired, and to make the prospects and position of the teacher dependent to a considerable extent on the results of this examination" (Armytage 1964, p. 124).

Armytage (1964) reports on the unforeseen by-products that led to many unfortunate practices: "The cult of the 'register,' acquiescence in large classes, the deliberate cultivation of rote-memory to defeat the inspectors; even, we are told, the presentment of sick children for attendance grant." The possibility of gaming the system and its consequences were recognized even then. Matthew Arnold reported the process "as a game of mechanical contrivance in which the teachers will and must more and more learn how to beat us" (Report of the Committee of Council for 1865, p. 291, quoted in Armytage 1964, p. 125).

After 30 years, the system of public funds based on performance was abolished largely because of the problem of designing the appropriate incentives. Perhaps we can learn a lesson from history.

References

Armytage, W. H. G. 1964. *Four Hundred Years of English Education*. Cambridge: Cambridge University Press.

Aoki, Masoto and Susan Feiner. 1996. "The Economics of Market Choice and At-Risk Students," in *Assessing Educational Practices: The Contribution of Economics*, edited by W. Becker and W. Baumol. New York: Russell Sage Foundation.

Bartlett, Will. 1993. "Quasi-Markets and Educational Reform," in *Quasi-Markets and Social Policy*, edited by J. LeGrand and W. Barlett. New York: Macmillan.

Bradley, Steve, Geraint Johnes, and Jim Millington. 2001. "School Choice, Competition, and the Efficiency of Secondary Schools in England." *European Journal of Operational Research* 135: 527–44.

Burgess, Simon and Paul Metcalfe. 1999. "Incentives in the Public Sector: A Survey of the Evidence." CMPO Discussion Paper 99/016, University of Bristol.

Burtless, Gary, ed. 1996. *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*. Washington, DC: Brookings Institution.

Chubb, John E. and Terry M. Moe. 1990. *Politics, Markets and America's Schools*. Washington, DC: Brookings Institution.

Dixit, Avinash. 1997. "Power of Incentives in Public versus Private Organizations." *American Economic Review* 87 (2): 378–82.

———. 2002. "Incentives and Organizations in the Public Sector: An Interpretative Review." *Journal of Human Resources* 37 (4): 696–727.

Dolton, Peter. 2002a. "Evaluating Education Inclusion." University of Newcastle-upon-Tyne, unpublished paper.

———. 2002b. "Is Performance Related Pay for Teachers Possible?" Institute of Education, unpublished paper.

Dolton, Peter, Stephen McIntosh, and Arnaud Chevalier. Forthcoming. "Teacher Pay and Performance: A Review of the Literature." *Bedford Way Papers*. London: Institute of Education Publications.

Fearon, James. 1999. "Electoral Accountability and the Control of Politicians: Selecting Good Types Versus Sanctioning Poor Performance" in *Democracy, Accountability and Repre-*

sentation, edited by A. Przeworski, S. Stokes, and B. Manin. Cambridge: Cambridge University Press.

Ferejohn, John. 1999. "Accountability and Authority: Toward a Theory of Political Account-ability," in *Democracy, Accountability and Representation*, edited by A. Przeworski, S. Stokes, and B. Manin. Cambridge: Cambridge University Press.

Friedman, Milton. 1962. "The Role of Government in Education," in *Capitalism and Freedom*. Chicago: University of Chicago Press.

Gibbons, Steve and Stephen Machin. 2002. "Valuing Primary Schools." London School of Economics, unpublished paper.

Glennerster, Howard. 1991. "Quasi-Markets for Education?" *Economic Journal* 101 (408): 1268–76.

———. 2002. "United Kingdom Education 1997–2000." *Oxford Review of Economic Policy* 18 (2): 120–36.

Gundlach, Erich, Ludger Wöessman, and Jens Gmelin. 2001. "The Decline of Schooling Productivity in OECD Countries." *Economic Journal* 111 (471): C135–47.

Hanushek, Eric. 1997. "Assessing the Effects of School Resources on Student Performance: An Update." *Educational Evaluation and Policy Analysis* 19 (2): 141–64.

Hirschman, Albert O. 1970. *Exit, Voice, and Loyalty*. Cambridge, MA: Harvard University Press.

Holmstrom, Bengt and Paul Milgrom. 1991. "Multi-Tasking Principal-Agent Analyses: Linear Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics, and Organization* 1 (7): 24–52.

Ladd, Helen F. 1996. *Holding Schools Accountable: Performance-Based Reform in Education*. Washington, DC: Brookings Institution.

Lazear, Edward. 2001. "Educational Production." *Quarterly Journal of Economics* 116 (3): 777–804.

Laver, Michael and Kenneth Shepsle. 1999. "Government Accountability in Parliamentary Democracy," in *Democracy, Accountability and Representation*, edited by A. Przeworski, S. Stokes, and B. Manin. Cambridge: Cambridge University Press.

Majumdar, Tapas. 1983. *Investment in Education and Social Choice*. Cambridge: Cambridge University Press.

Przeworski, Adam, Susan Stokes, and Bernard Manin, eds. 1999. *Democracy, Accountability and Representation*. Cambridge: Cambridge University Press.

Todd, Petra and Kenneth Wolpin. 2003. "On the Specification and Estimation of the Production Function for Cognitive Achievement." *Economic Journal* 113 (485): F3–33.