The Economic Journal, **113** (*February*), F3–F33. © Royal Economic Society 2003. Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA.

ON THE SPECIFICATION AND ESTIMATION OF THE PRODUCTION FUNCTION FOR COGNITIVE ACHIEVEMENT*

Petra E. Todd and Kenneth I. Wolpin

This paper considers methods for modelling the production function for cognitive achievement in a way that captures theoretical notions that child development is a cumulative process depending on the history of family and school inputs and on innate ability. It develops a general modelling framework that accommodates many of the estimating equations used in the literatures. It considers different ways of addressing data limitations, and it makes precise the identifying assumptions needed to justify alternative approaches. Commonly used specifications are shown to place restrictive assumptions on the production technology. Ways of testing modelling assumptions and of relaxing them are discussed.

An extensive, multidisciplinary empirical literature studies the determinants of cognitive achievement in children. The early childhood development (ECD) branch of this literature seeks to understand the role of parental characteristics and the early home environment in producing cognitive skills. The education production function (EPF) branch of the literature examines the productivity relationship between schooling inputs and test score outcomes for school-age children.

In the EFP literature, researchers draw an analogy between the knowledge acquisition process of a human being and the production process of a firm. A primary goal of empirical research is to understand the technology for combining school inputs to create cognitive achievement outcomes. The production function analogy provides a conceptual framework that guides the choice of variables and enables a coherent interpretation of their effects. In ECD studies this analogy is less transparent, although the goal is highly similar and both literatures could be placed under the same umbrella.

Ideally, in analysing cognitive achievement of children, it would be useful to have access to data on all past and present family and school inputs as well as information about children's heritable endowments. Because existing data sets are deficient in one or more of these domains, an important issue confronted in both the ECD and EPF literatures is the problem of missing data. Data sets used in ECD studies often have rich longitudinal information on early childhood environments but lack data on schools. Data sets used in EPF studies have data on schooling inputs, at least at one point in time, but contain limited and mostly contemporaneous family background information and may lack data on historical schooling inputs.[1] For this reason, the EPF literature often treats early childhood

* We thank Christian Dustmann, James Heckman, Lance Lochner, Steve Machin, and three anonymous referees for helpful comments.

[1] For an example of a richer dataset that includes information both about school and family inputs, see Dustmann *et al.* (2003).

inputs as unobservables and invokes assumptions under which the unobservables can be eliminated or ignored.[2] Common estimating equations include so-called 'value-added' specifications, which assume that a lagged test score provides a sufficient statistic for all historical inputs and heritable endowments, or 'fixed-effect' specifications, which attempt to difference out unobservables over time or over multiple children in the same family.

In both the ECD and EPF literatures, there is a remarkable lack of consensus over which inputs increase children's achievement and to what extent; see e.g. Parcel and Menaghan (1994), Hedges *et al.* (1994) and Hanushek (1986, 1996). For example, many child development researchers examine the question of whether maternal employment in the early years of a child's life is detrimental to children's cognitive and social development. Even when studies are based on the same data source, estimates range from maternal employment being detrimental (Baydar and Brooks-Gunn, 1991; Desai *et al.*, 1989; Belsky and Eggebeen, 1991) to its having no effect (Blau and Grossberg, 1992) to its being beneficial (Vandell and Ramanan, 1992). As most of these studies employ conventional regression models, differences in conclusions are undoubtedly due to variations in sample inclusion criteria and to the choice of conditioning variables.

There are similar disagreements in the schooling quality literature over whether schooling inputs, such as class-size, teacher experience, teacher education, and term length, matter in producing cognitive skills in children. In one of the earliest investigations of the link between school inputs and achievements outcomes, Coleman (1966) found surprisingly small effects of school resources on student achievement. His influential report was the impetus for hundreds of empirical studies of the school-quality–achievement relationship that, thus far, do not appear to be converging towards a consensus. Recent exchanges between Hanushek and Krueger provide examples of the debate that has characterised this literature, a debate that is continued in their contributions to this issue. For example, Hanushek (1998) and Krueger (1998, 2000) analyse US aggregate time series data on expenditure and NAEP (National Assessment of Educational Progress) test scores[3], with Krueger concluding that increases in expenditure have led to modest gains in test scores and Hanushek finding 'no strong or consistent relationship between school resources and student performance'.

The aggregate evidence of NAEP test scores suggests that test scores did not improve much between 1960 through 1990, despite substantial improvements in the quality of schooling as indicated by lower class sizes, rising teacher education levels, and increases in overall school expenditure. (Hanushek, 1998). However, simple correlations between test score outcomes and contemporaneous quality input measures are difficult to interpret, because other factors such as family inputs are entirely left out of the analysis. As Hedges and Greenwald (1996) argue, a possible explanation for the lack of a strong correlation between quality inputs and test scores is that parental inputs into the achievement process have shown net

[2] See the literature review in Hanushek (1986), and more recently, Goldhaber and Brewer (1997), and Hanushek *et al.* (1998).

[3] See Hanushek (2003) for discussion of the NAEP exam.

declines. The decline is often associated with rising female labour force participation rates and the rising prevalence of single-parent households, although these factors could be offset by other factors potentially beneficial to children's learning, such as rising parental education levels (Burtless, 1996). Hanushek (2003) argues that changes in family inputs are not sufficient to explain the lack of improvement in test scores.

Empirical studies employ a wide variety of estimating equations. Hanushek (1996, 2003) summarises in several meta-analyses the findings of EPF studies and concludes that although some papers find statistically significant effects, the overall pattern suggests that none of the measured schooling quality inputs are reliably associated with achievement. Krueger (2003) takes issue with the conclusions drawn by Hanushek, arguing that Hanushek's sample of estimates are biased towards his conclusion. A leading candidate for explaining why studies reach such different conclusions is that the statistical models used to estimate these relationships are misspecified and fail to account for the major determinants of achievement. Although Hanushek and Krueger disagree about the conclusions that may be drawn from meta-analytic studies, they agree on the importance of taking into account model specification in combining evidence across studies, As they note, estimating equations are often adopted with little theoretical justification, making it difficult to know whether the assumptions that underlie particular approaches are reasonable. According to Krueger (2003), 'There is no substitute for understanding the specifications underlying the literature.'

This paper focuses on the problem of how to specify and estimate a production function for cognitive achievements in a way that is consistent with theoretical notions that child development is a cumulative process depending on the history of inputs applied by families and schools as well as on children's inherited endowments. We develop a general conceptual framework for estimating the relationship between achievement outcomes and family and schooling inputs and consider how to implement this framework under different kinds of data limitations. For each estimator, we discuss (i) its identifying assumptions, (ii) the plausibility of these assumptions in view of the conceptual model, (iii) the data needed to implement the estimator, and (iv) conditions under which some of the assumptions of the estimation method can be tested. In surveying different estimation approaches, we interpret many of the specifications used in the ECD and EPF literatures in terms of restrictions they place on the production technology. Some of the specifications that are commonly adopted can only be justified under highly restrictive assumptions on unobservables and on the parameters of the production function.

Our discussion is also relevant to the recent literature on outcome-based approaches for assessing the effectiveness of teachers. For examples, statistical models such as the Tennessee Value-Added Assessment System (TVAAS), are increasingly being used by school systems in evaluating teacher performance and ultimately in making promotion, hiring, and salary decisions.[4] We offer an

[4] TVAAS does not use a 'value-added' specification as defined previously. The terminology 'value-added' in the context of TVAAS is used to refer to the marginal contribution of a teacher. The type of specification adopted by TVAAS will be discussed in Section 2.3.4.

interpretation of the parameters estimated in statistical models such as TVAAS and consider the conditions under which they provide estimates of production function parameters.

Before turning to the topic of estimation, however, the next Section considers the broader question of what are the parameters of interest in studying the determinants of cognitive achievement outcomes. We distinguish between two types of parameters that have been the focus of empirical education research: policy effects and production function parameters (Heckman, 1992, 2000). We discuss why studies of the effects of school quality that are based on experimental data generally identify policy effects while studies based on nonexperimental data usually identify parameters of the education production function. Because the two types of analyses estimate different parameters, there is no reason to expect experimental and nonexperimental estimates to agree. We describe the advantages and limitations of experimental and nonexperimental evidence for answering different types of policy relevant questions.

1. What are the Goals of Estimation? Policy Effects vs. Production Function Parameters

Empirical studies of the link between achievement outcomes and school inputs can be broadly classified into two main types: nonexperimental and experimental. Nonexperimental studies are based on observational data, where a reasonable assumption is that the inputs into the education production process are subject to choices made by parents and schools. The fact that inputs are chosen purposefully would not necessarily pose a problem in estimating a production function for achievement if data on all relevant inputs as well as child endowments were observed; but, it does pose a problem when data on relevant inputs and endowments are missing. Thus, an important question considered in detail later in the paper is how to account for unobservables and for the potential endogeneity of observed inputs in modelling the relationship between cognitive achievement and school and family inputs.

In experimental studies, the values of at least a subset of the inputs are chosen by random assignment and are therefore not subject to choices made by parents or schools. For example, in the Tennessee Student/Teacher Achievement Ratio (STAR) experiment, children were randomly assigned to small or regular size classes.[5] Random assignment creates exogenous variation that, under ideal conditions, allows certain policy effects to be identified even in the presence of missing data problems.

As we make precise below, the parameters estimated in experimental studies, and also in most studies based on so-called 'natural experiments', typically differ from those estimated in nonexperimental studies and one type of evidence does not substitute for the other. We believe the difference has not been fully

[5] Students were assigned to the following categories: small class (13–17 students), regular-size (22–25), regular class with teacher's aide. See Finn and Achilles (1990) Moesteller (1995) Krueger (1997, 1998, 2003).

appreciated in the education production function literature, as comparisons are often drawn between experimental and nonexperimental estimates of school input effects under the presumption that the two types of studies are estimating the same parameter (Krueger, 2000; Hanushek, 2002).[6]

1.1. *A Simple Model for the School-Quality–Achievement Relationship*

In this Section, we present a simple model of achievement that we use to define and interpret various parameters of interest that have been the focus of nonexperimental and experimental studies. Later, we build on this framework in discussing alternative approaches to estimating cognitive achievement production functions.

First we need to define some notation. Let $t = 0$ correspond to the time interval prior to the age that the child enters school, $t = 1$ correspond to the first year of school, and $t = 2$ to the second year. A_1 denotes the child's achievement level at a point in time just prior to entering the first year of school. Let F_0 represent family inputs into the cognitive achievement production; process during the $t = 0$ (preschool) interval. For simplicity, for now, we abstract from the problem for unobservable data on inputs. Let μ be a measure of the child's endowed ability or mental capacity, assumed to be determined at the time of conception. Achievement at the time of school entry depends only on family inputs and ability:

$$A_1 = g_0(F_0, \mu).$$

Family inputs in the preschool period are assumed to be determined by the family's permanent resources, W, and the child's endowment.[7] Family choices about where to live and about whether to send children to public or private schools partly determine the level of school inputs the child experiences. If school inputs were solely a function of the family's location decision, then their determinants would be the same as those of family inputs. However, at the time of the location decision, parents have incomplete information about what the level of school inputs will be when their child is attending, and, even with complete information, the level of inputs applied to their child may differ from the aggregate school level. Therefore, we draw a distinction between the school level inputs, denoted by $\bar{S}_1$, chosen by the family at the time of the location (or private/public) decision, and $S_1 - \bar{S}_1$, the deviation between the actual level relevant to their child and the school-level.

We assume achievement at the start of the second year of school depends on the entire history of family inputs (F_0 and F_1) and school inputs (S_1) as well as on endowments:

$$A_2 = g_1(S_1, F_1, F_0, \mu).$$

[6] However, the difference between treatment effects estimated by experimental studies and structural parameters estimated by nonexperimental studies in general is recognised in the evaluation literature. See, for example Heckman (1992, 2000) and Heckman and Vytlacil (2000). For specific cases in which the distinction has been drawn, see Rosenzweig and Wolpin (1995) and Wolpin (1997).

[7] For simplicity, we consider the case of a family with a single child, but our basic points are not sensitive to this assumption.

Along with the technology for combining inputs to create achievement outcomes, there is a decision rule for both parents and schools that determines the level of inputs. The family decision rules concerning direct family inputs and the school inputs associated with their location decision are given by:

$$F_1 = \phi(A_1, W, \mu, S_1 - \bar{S}_1)$$
$$\bar{S}_1 = \theta(A_1, W, \mu),$$

where W represents the family's permanent resources. Notice that family input decisions are assumed to be made subsequent to the actual realisations of the school inputs applied to their children.

We assume that the school chooses input levels for a particular child purposefully, taking into account the child's achievement level and the endowment. For example, a child who enters first grade able to read may receive different kinds of school inputs than a child who is not able to read. At higher grades, schools sometimes use prior achievement levels or placement exams to 'track' students. The school's input decision rule is therefore given by $S_1 = \psi(A_1, \mu)$, which does not depend directly on the level of family resources. At the beginning of the second year, the family makes a new decision about where to live, governed by the child's achievement at the end of the first year, the family's resources and the child's endowment.

Within this simple framework, we can now consider different kinds of parameters of interest in empirical research. A typical question of interest is the following:

(*Q1*) *How would an exogenous change in class size, holding all other inputs constant, affect achievement?*

Knowledge of the production technology would suffice to answer this question. The goal of most observational studies has been to uncover features of the production technology, given above by g_0 and g_1.

1.2. *What Do We Learn from Experiments?*

An alternative question that may also be of interest is the *non-ceteris-paribus* effect of changing school inputs:

(*Q2*) *What would be the total effect of an exogenous change in class size on achievement, that is, not holding other inputs constant?*

The total effect includes both the *ceteris paribus* effect holding other inputs constant as well as any indirect effects that operate through changes in the levels of other inputs. In our simple model with only one school input, the total effect of an increase in the school input in the first year, S_1, on achievement in the second year, A_2, is given by[8]

$$\begin{aligned}\frac{\mathrm{d}A_2}{\mathrm{d}(S_1 - \bar{S}_1)} &= \frac{\mathrm{d}A_2}{\mathrm{d}S_1} \\ &= \frac{\partial g_1}{\partial S_1} + \frac{\partial g_1}{\partial F_1}\frac{\partial F_1}{\partial(S_1 - \bar{S}_1)}.\end{aligned} \tag{1}$$

[8] The experiment is conceived as an unanticipated change in the school input, that is, as a change in $S_1 - \bar{S}_1$. However, given the production function, a change in S_1 has the same effect.

Knowledge of the technology is not sufficient to answer (Q2), because the production function provides $\partial g_1/\partial S_1$ and $\partial g_1/\partial F_1$, but not $\partial F_1/\partial(S_1 - \bar{S}_1)$. To get the last term requires knowledge of the family input decision rule. The production parameters and the input decision rules could be estimated from nonexperimental data and, in principle, be used to obtain (Q2). As seen in (1), the total effect does not correspond to any single parameter of the technology or to any other parameters of the decision rules, such as those reflecting underlying preferences.

An experiment that randomly allocates class size (as in the STAR experiment) provides an answer to (Q2) but not (Q1). A comparison of the outcomes for children randomly allocated to different levels of school inputs gives the average effect of a change in resources on achievement for the group of persons participating in the experiment, which we can write as

$$\mathrm{E}\left(\frac{\mathrm{d}A_2}{\mathrm{d}S_1}\right) = \int \left[\frac{\partial g_1}{\partial S_1} + \frac{\partial g_1}{\partial F_1}\frac{\partial F_1}{\partial(S_1 - \bar{S}_1)}\right]_{A_1,W,\mu} f(A_1, W, \mu)\mathrm{d}A_1\mathrm{d}W\mathrm{d}\mu \tag{2}$$

where $f(A_1, W, \mu)$ is the joint density of A_1, W and μ. The randomised experiment uncovers a 'policy effect' that is often of interest – the average total effect of the change in school resources on achievement for the children in the population participating in the experiment, also known in the evaluation literature as the effect of *treatment-on-the-treated*.[9]

One limitation of experiments that is apparent in (2) is that extrapolating the results of the experiments to other populations is only valid if the indirect effect (given by the second RHS term) are expected to be the same on average in the new population. An assumption sufficient to justify such an extrapolation is if $\partial g_1/\partial S_1$ and $\partial g_1/\partial F_1$ are independent of μ and $\partial F_1/\partial(S_1 - \bar{S}_1)|_{A_1,W,\mu}$ is also a constant, independent of A_1, W, and μ, in which case

$$\mathrm{E}\left(\frac{\mathrm{d}A_2}{\mathrm{d}S_1}\right) = \frac{\partial g_1}{\partial S_1} + \frac{\partial g_1}{\partial F_1}\frac{\partial F_1}{\partial(S_1 - \bar{S}_1)}.$$

This assumption is stronger than necessary as the requirement that indirect effects be the same in the new population need only hold on average.

The average policy effect defined above can either be larger or smaller than the *ceteris paribus* effect. For example, in the Tennessee class-size experiment, families whose children were assigned to small classes may have spent less time teaching their children at home, that is, if school and family inputs are substitutes in producing achievement. In that case, although the *ceteris paribus* effect of class size is $\partial g_1/\partial S_1 > 0$, $\int(\partial g_1/\partial F_1)[\partial F_1/\partial(S_1 - \bar{S}_1)]|_{A_1,W,\mu} f(A_1, W, \mu)\mathrm{d}A_1\mathrm{d}W\mathrm{d}\mu < 0$, so the average policy effect measured by the experiment would be less than the *ceteris paribus* effect. Alternatively, the effects could reinforce each other.[10] This could happen, for example, if families were encouraged to apply more family inputs by their child's

[9] The effect of treatment on the treated is only one of many different kinds of policy effects that may be of interest. For a discussion of different kinds of policy relevant treatment effect parameters, see Heckman (2000, 2001) and Heckman and Vytlacil (2001).

[10] If one interprets the evidence that the experimental effect of class-size tend to be larger than those obtained from production function estimates, then one could conclude that as class-size is reduced, additional inputs are provided to children that reinforce the class-size effect.

greater learning at school, that is, school and family inputs are complements. Only under the assumption that families do not take into account changes in school characteristics in choosing family input levels $\left(\text{i.e., } \partial F_1/\partial(S_1 - \bar{S}_1) = 0\right)$ would the experimental estimate correspond to a production function parameter.

The fact that experiments do not generally recover parameters of the production function is not necessarily a limitation, as the total policy effect is precisely the desired effect for answering the policy question posed in (Q2). A key advantage of experiments, when they are properly implemented, is that they provide a way of estimating policy effects without additional assumptions. However, they do not answer questions such as (Q1) and it is a limitation that the evidence from experiments cannot necessarily be generalised to other populations of interest.

In addition to experimental studies, a few studies in the EPF literature obtain estimates of school quality input effect using so-called 'natural experiments'. These studies are also based on observational data, but they differ from nonexperimental studies in that there is no attempt to incorporate in the model all the determinants of cognitive achievement and the studies make use of a variable that arguably provides a source of exogenous variation in school input levels analogous to that provided by randomisation. For example, Angrist and Lavy (1999) make use of Maimonides' rule, a rule that partly determines class sizes in Israel, as an instrumental variable in analysing what they term the 'casual effect' of class-size on student achievement. As in consistent with their aim, their analysis does not hold constant family inputs or other school inputs in estimating the class-size–achievement relationship, so the 'casual effect' that is estimated corresponds to a total policy effect in our terminology. This policy effect, however, corresponds in the evaluation literature to the local average treatment effect (or LATE), which differs from the treatment-on-the-treated parameter if the population induced to receive treatment by the instrument differs from the population under a randomised experiment; see Heckman and Vytlacil (2000) and Imbens and Angrist (1994). Natural experiments share many of the advantages and limitations of randomised experiments.

To summarise, nonexperimental and experimental studies generally answer different questions of interest, so there is no reason to expect estimates of school input effects based on experimental studies to match those from nonexperimental studies. Notably, experiments do not generally estimate production function parameters, so they do not solve the problem associated with estimating education production functions, as is sometimes presumed in the literature. Therefore, in the remainder of the paper, we turn to ways of modelling and estimating cognitive achievement production functions.

2. Modelling the Production Function under Different Kinds of Data Limitations

A basic tenet of our approach is that estimating the cognitive achievement production function in school-age children requires taking into account both school and family inputs, current and past. The problem of missing data on inputs and on endowments and the related problem of imprecisely measured inputs present

major obstacles in estimation. In ECD studies, data are often available on family inputs but are lacking on school inputs. EPF studies typically use data sets gathered at schools, so information pertaining to the current school inputs are available but information on family inputs and historical school inputs is often very limited. In fact, sometimes the only variable available EPF studies related to the family is the percentage of students in the school participating in a free lunch programme, usually cited to be a proxy for family wealth (which is not itself an input).

Confronted with what are sometimes severe data limitations, empirical researchers have pursued a variety of estimation strategies to overcome them. One approach explicitly recognises the presence of omitted variables and develops estimators that allow for them. Another common remedy is to use one or more proxy variables that are not considered direct inputs into the education production process but are included as conditioning variables under the presumption that their inclusion will alleviate omitted variables bias because they are correlated with omitted inputs. Variables such as race or family income could be considered such proxy variables. Below, we consider the question of whether or not to include proxy variables and conclude that sometimes it may be better to refrain from using them, because their inclusion confounds the interpretation of other input effects and may even exacerbate biases.

Tables 1 and 2 describe representative subsets of studies from the ECD and EPF literatures in terms of the types of variables included in the analysis. As seen in Table 1, all ECD studies shown in the table include contemporaneous inputs, i.e., inputs that are close in time to the achievement measure. They also all include some inputs that are removed in time from the measure of achievement (historical inputs), although they differ in the specific measures used and inputs are not always treated symmetrically over time. For example, the study by Parcel and Menaghan (1994) includes separate contemporaneous and historical measures of maternal employment, but only includes a contemporaneous measure of family inputs, even though historical measures were available in the dataset. None of the studies has data on school inputs, which is problematic for ECD studies that focus on school-age children. For example, Baharudin and Luster (1998) and Crane (1996) analyse effects of family inputs on cognitive achievement of school-age children, ignoring the contribution of schools. Murnane *et al.* (1981) address the problem of missing school inputs through the use of school fixed-effects that assumes that children within the same school receive the same school inputs ($S_1 - \bar{S}_1 = 0$). Rosenzweig and Wolpin's (1994) within-family estimator addresses the same problem through the use of sibling differences, under the assumption that siblings experience the same quality of schooling.

In all ECD and EPF studies, children's heritable endowment is an important, unobserved determinant of cognitive achievement outcomes. It is common practice in the ECD literature to use as a proxy for children's inherited endowment a measure of parental ability, such as the mother's AFQT score. This practice is less common in the EPF literature, possibly because such data are less often available.

Table 2 summarises a variety of modelling approaches adopted in the EPF literature. Early studies tended to include only contemporaneous inputs (Hanushek, 1986). More recent research adopts value-added specifications to mitigate the

Table 1

Early Child Development (ECD) Studies

		Parcel and Menaghan (1994)	Baydar and Brooks-Gunn (1991)	Baharudin and Luster (1998)	Murnane *et al.* (1981)	Rosenzweig and Wolpin (1994)
Question of Interest		effect of early parental work, family social capital on cognitive outcomes	effect of early maternal employment on cognitive outcomes	factors related to quality of home environment and influence on cognitive ability	effect of mother's education and home resource environment on children's achievement	effect of mother's education on children's cognitive achievement
Outcome measure		PPVT	PPVT	PIAT-R, RC,M	Iowa Test of Basic Skills, Vocabulary Subset	PPVT, PIAT-R,M
Sample		3–6 years-olds	3–4 year-olds	6–8 year-olds	children grades 3–6	3–8 year olds, 1st and 2nd children only
Data Source		NLSY children	NLSY children	NLSY children	children from families in the Gary Neg. Inc. Tax Experiment	NLSY children
Contemporaneous/ Non-varying Covariates	parental characteristics	yes (focus on labour supply, occupational complexity)	yes (focus on labour supply of mothers)	yes (marital status, self-esteem)	yes (focus on education)	yes (includes labour supply)
	parental ability	yes (mothers AFQT)	no	yes (mother's AFQT)	no	yes (mother's AFQT)
	child characteristics	yes (incl shyness, health problems)	yes	yes (age, gender)	yes (gender)	no
	family characteristics	yes (does not include income)	yes (includes income)	no	yes (incl. income, expenditures, rented housing)	no
	home environment	yes (home score)	yes	yes (home score)	yes (encyclopedia)	no
Covariates pertaining to earlier time periods	parental characteristics	yes	no	no	no	yes (incl. labour supply)
	child characteristics	yes (low birthweight)	no	no	no	yes (birthweight)
	family characteristics	yes	no	no	no	yes
	home environment	no	no	yes (home score)	no	yes
	lagged test score	no	no	no	yes	no
Estimation Method		OLS	OLS	OLS (stepwise forward selection)	Value-added model est. by OLS with school fixed effects	GLS w/mother fixed effects

Table 2

Educational Production Function (EPF) *Studies*

Study		Goldhaber and Brewer (1997)	Hanushek *et al.* (1998)	Krueger (1998)	Ludwig (1999)
Outcome measure		NELS Math Achievement Test	TAAS(*) scores in reading, math	NAEP(†) test score	NELS(‡) Match and Reading Achievement Tests
Sample		10th graders	multiple cohorts of 4th graders	9, 13, and 17 year-olds	8th graders, 3 time periods
Type of Data/ Data source		NELS(‡) 1988 data, micro-level test and resource data		aggregated, time series data	NELS(‡) 1988 data, micro-level test and resource data
Contemporaneous/ Non-varying Covariates	school characteristics	multiple teacher characteristics, teaching styles, school and class size, school location	class size, %teachers with grad. degree, teacher experience aggregated to grade level	spending per student, model (aggregated at national level), Indicators for math, reading tests	student-teacher ratio, lowest full-time teacher salary, teacher education, teacher experience, % students on free lunch, racial characteristics
	child char.	sex, race	no	no	sex, race
	parental char.	parental education, family structure, family income	no	no	SES (socioeconomic status), education, income
	home env.	no	no	no	parents help with homework
Covariates pertaining to earlier time periods	school char.	no	no	no	no
	child char.	no	no	no	no
	parental char.	no	no	no	no
	home env.	no	no	no	no
	lagged test score	yes	yes	no	yes
Estimation Method		random and fixed-effects for teachers and schools	difference-in-differences (equivalent to value added with individual fixed effects)	OLS	IV estimator applied to model that relates changes over time in test scores to changes in resources (preferred specification)

*TAAS - Texas Assessment of Academic Skills; †NAEP - National Assessment of Educational Progress; ‡NELS - National Education Longitudinal Study of 1988.

need for data on historical inputs and endowments. Below, we show that the value-added specification places strong restrictions on the production technology. Also, the use of this type of specification does not eliminate the need for measures of contemporaneous family inputs, which are often lacking in EPF studies.

2.1. *How Can Inclusion of Proxy Variables Confound the Interpretation of the Effects of Other Inputs and Increase Bias?*

When precise measures of desired input variables are unavailable, researchers often substitute for the missing variables one or more proxy variables. If proxies measure desired variables up to random error and there is no measurement error in other input variables, then including the proxy variables reduces omitted variables bias relative to excluding them, as shown in McCallum (1972). In practice, though, researchers are often faced with a more difficult choice – whether to include proxy variables that are not simply imperfect measures of specific inputs. For example, to compensate for missing data on family inputs that might affect children's achievement, researchers often include family income as a proxy, the presumption being that families with higher income purchase more such goods. However, the use of this proxy is problematic in that an increase in the amount of a purchased input, such as the number of books available to the child, holding income constant must imply a reduction in expenditures on other goods (for example, Murnane *et al.* (1981) condition in their analysis both on ownership of an encyclopedia and family income). To the extent that these other goods (e.g., educational toys or tutors) also affect child achievement, the effect of an increase in the number of books on achievement will be confounded with the effect of the reduction in these other goods. Unlike the case of 'close' proxy, where inclusion is unambiguously beneficial, the inclusion a 'crude' proxy that is related to included and omitted variables, for example through decision rules, as in the case of family income, often confounds the interpretation of observed inputs and can actually lead to greater bias in model coefficients (Wolpin, 1995, 1997).

The use of proxy variables similarly may confound the interpretation of estimated model coefficients in EPF studies. Consider, for example, a model that relates achievement to a school input such as average class size. To compensate for missing data on other school inputs a researcher might include a variable such as expenditure per pupil, analogous to the inclusion of family income in ECD models. However, it must be the case that schools with identical expenditures per pupil, but smaller average class sizes, necessarily spend less on some other unobserved inputs (e.g., having less experienced teachers).[11] Thus, the measured effect of class size on achievement, conditional on per pupil expenditure, is net of the effect of the unobserved inputs. If class size and the unobserved inputs were uncorrelated, so that the omitted variables bias would have been zero in the model without the proxy, by including expenditure in the regression the researcher ensures that movements in class size are now confounded with movements in the

[11] Rosenzweig and Wolpin (1995) make this same observation with respect to family income. Krueger (2000) has recently make it also with respect to school expenditures.

unobserved inputs. Thus, including proxy variables can actually lead to greater bias.[12]

An alternative use of proxies such as family income or expenditure per pupil, because they should have no effect on achievement if all inputs are controlled, is as a diagnostic tool for assessing the importance of omitted variables. Omitted variables must exist if these kinds of proxies affect achievement net of included inputs. If it is found that the effect of included inputs changes substantially when such proxies are added to the regression, the researcher should be agnostic as to which estimate is of lower bias (Wolpin, 1995). In some sense, the problem of whether or not to include proxy variables is insoluble, because it involves a comparison between two unknown biases. Nonetheless, the fact that including them can make estimates difficult to interpret suggests that they should be used cautiously. Additionally, the relationship of proxy variables to measured and unmeasured inputs must be understood in the context of a behavioural decision model in order to analyse their likely impact on biases.

2.2. *A Conceptual Framework*

We next lay out a statistical model for cognitive achievement that assumes that children's achievement, as measured by test performance at some particular age, is the outcome of a cumulative process of knowledge acquisition. After presenting the model in its most general form, we use it to interpret the types of restrictions that commonly used estimating equations place on the production technology.

Let T_{ija} be a measure of achievement for child i residing in household j at age a. We conceive of knowledge acquisition as a production process in which current and past inputs are combined with an individual's genetic endowment of mental capacity (determined at conception) to produce a cognitive outcome.[13] As described in Section 1, we assume that inputs reflect choices made by parents and schools. Denote the vector of parent-supplied inputs at a given age as $\boldsymbol{F}_{ija}$, school-supplied inputs as $\boldsymbol{S}_{ija}$ and the vectors of their respective input histories up to age a as $\boldsymbol{F}_{ij}(a)$ and $\boldsymbol{S}_{ij}(a)$.[14] Further, let a child's endowed mental capacity ('ability') be denoted as μ_{ij0}, where there is an implicit assumption that there is only one kind of mental capacity relevant to acquiring all types of knowledge.[15] Then, allowing for measurement error in test scores, denoted by ε_{ija}, the production function is given by

$$T_{ija} = T_a\left[\boldsymbol{F}_{ij}(a), \boldsymbol{S}_{ij}(a), \mu_{ij0}, \varepsilon_{ija}\right]. \tag{3}$$

[12] Wolpin (1997) provides a similar example, although in a different context, where the inclusion of race in a production function may lead to greater bias in the estimates of production function parameters.

[13] This conception was first formally modelled by Ben-Porath (1967) in the context of an individual decision-maker choosing the level of (time and money) resources to devote to human capital investments. It has since served as the basis for much of the literature on skill acquisition in economics. Liebowitz (1974) was the first to extend this conception to home investments in children.

[14] The test measurement at age a is assumed to be taken after the age a inputs are applied.

[15] The assumption of only a single general intelligence factor, about which there is considerable debate, is made only for notational convenience. Allowing for different mental capacity endowments for different cognitive skills, creates no additional issues.

The a subscript on $T_a(\cdot)$ allows the impact of inputs and of the genetic endowment to depend on the age of the child.

2.3. *Alternative Specifications and their Identifying Assumptions*

The empirical implementation of the model described by (3) has founded on two main problems: (i) the genetic endowment of mental capacity is non-observable and; (ii) data sets on inputs are incomplete – in particular, they have incomplete input histories and/or missing inputs. To understand the manner in which the EPF and ECD literatures have dealt with these problems, in what follows we inventory the commonly used specifications and discuss the identifying assumptions that are required under those specifications for (3) to be consistently estimated. Table 3 provides an overall summary of the different specifications.

2.3.1. *The contemporaneous specification*

The 'contemporaneous' specification relates an achievement test score measure solely to contemporaneous measures on school and family inputs. The following assumptions on the production technology and on the input decision rules would justify its application.

ASSUMPTIONS

(i) *Only contemporaneous inputs matter to the production of current achievement.*
or
(ii) *Inputs are unchanging over time, so that current input measures capture the entire history of inputs.*
and, in addition to (i) or (ii),
(iii) *Contemporaneous inputs are unrelated to* (*unobserved*) *endowed mental capacity.*

The contemporaneous specification is usually adopted when there are severe data limitations in that little or no data are available on historical input measures or test score measures. We can write the contemporaneous specification as

$$T_{ija} = T_a(\boldsymbol{F}_{ija}, \boldsymbol{S}_{ija}) + \varepsilon'_{ija}, \tag{4}$$

where ε'_{ija} is an additive error. In such a specification, as seen in comparison to the true technology function (3), the residual term includes all the omitted factors – the history of past inputs, endowed mental capacity and measurement error. Clearly, in this setting, the assumptions necessary to obtain consistent estimates of the impact of contemporaneous inputs, the only observable data, are quite severe. Neither of the sets of assumptions given above is very plausible. Most theories of child development posit important links between experiences during infancy and early childhood and later childhood cognitive, social and behavioural outcomes.[16] Moreover, many inputs of potential importance in the development of cognitive skills vary temporally and may vary for systematic reasons with the child's age (e.g.,

[16] Probably the most well known theory of cognitive development is Piaget's theory, in which children are conceived as passing through stages with specific developmental characteristics that build sequentially on each other. See Case (1992).

Table 3

Identifying Assumptions and Data Requirements for Different Estimators

(Under assumption that coefficients on input effects are constant across ages, but are allowed to depend on the length of time since the input)

Model specifications	Eqn. no.	Estimation method	Key assumptions required for consistency	Minimal data requirements for implementing estimator	Test of specification relative to model III.3,III.4
I. Contemporaneous					
1. without endowment input correlations	(4)	OLS	only current inputs matter or inputs constant over time (to justify omitting past); A1*	one period of data on achievement test score and current values of inputs	Test whether coefficient on lagged inputs = 0
2. with parental endowment-input correlations		sibling fixed effects	inputs uncorrelated with sibling error terms (i.e. parents input decisions do not depend on innovations on sibling outcomes); A1*	one period on test score and current values of inputs on at least one sibling	same test as above
3. with child endowment-input correlations		child fixed effect	current inputs uncorrelated with error term from earlier age (i.e. parents' input decisions do not depend on innovations from earlier outcomes); A1*	two periods data on achievement and current values of inputs	same test as above
II. Value-Added					
1. gain formulation with no endowment	(7)	OLS	effects of inputs same across all ages	data on test score and on lagged test score, current values of inputs	Lagged input coefficients equal to current input coefficients
2. modified gain formulation with no endowment (lagged test score is a regressor)	(6)	OLS	effects of inputs decays at a constant rate over time	same as above	Ratios of coefficients on current inputs to lagged inputs are the same across all inputs
3. modified gain formulation with parental endowment	(10)	sibling fixed effects	effect of inputs decays at a constant rate over time; inputs uncorrelated with sibling error terms (i.e. parents input decisions do not depend on innovations on sibling outcomes)	one period data on test score, lagged test score, current values of inputs on siblings	same test as above

Table 3

Continued

Model specifications	Eqn. no.	Estimation method	Key assumptions required for consistency	Minimal data requirements for implementing estimator	Test of specification relative to model III.3,III.4
4. modified gain formulation with child endowment	(10)	child fixed effects	effect of inputs decays at a constant rate over time; current inputs uncorrelated with error term from earlier age (i.e. parents decisions do not depend on innovations from outcomes)	two periods data on test score, lagged test score, current values of inputs	same test as above
III. Cumulative					
1. without endowment-input correlations		OLS	A1* only	one period data on test score and on current and past inputs	Hausman–Wu test for existence of fixed effect
2. with parental endowment	(8″)	sibling fixed effects	inputs uncorrelated with sibling error terms (i.e. parents' input decisions do not depend on innovations on sibling outcomes); A1*	one period data on test score and on current and past inputs for siblings	Hausman–Wu test comparing estimates from the model with parental endowment to estimates from the model with child endowmment
3. with child endowment	(8″)	child fixed effects	current inputs uncorrelated with error term from earlier age (i.e. parents' input decisions do not depend on innovations from earlier outcomes); A1*	two periods data on test score and on current and past inputs	Test relative to model III.4 estimated by method using Hausman–Wu test
4. within child endowment model, allowing for unobserved inputs (current or lagged)	(8′)	IV estimator applied to child fixed effect model	instruments are valid (can use lagged values of own inputs and of sibling inputs as instruments, for time periods earlier than earliest test score, under the assumptions that inputs made at times before the test score was realised do not respond to the innovation on the test score); A1*	same as above, in addition data on inputs prior to earlier time period for individual and for sibling (used as the instruments)	Can be tested if there are over-identifying restrictions

* Assumption A1: any omitted inputs orthogonal to included ones.

maternal employment), or be specific to particular ages (e.g., maternal alcohol use during pregnancy).

Assumption (iii) – that inputs and endowed ability are uncorrelated – is also inconsistent with economic models of optimising behaviour, as is clear in the model of Section 2. Economic models in which parents care about a child's cognitive development imply that the amount of resources allocated to the child, in the form of purchased goods and parental time, will be responsive to the parent's perception of a child's ability, for example, Becker and Tomes (1976).

Thus, while the contemporaneous specification can be implemented with only limited data, strong assumptions are required to justify its application.

2.3.2. *Value-added specifications*

The lack of data on input histories and on endowed capacity has led researchers to adopt what has been called the *value-added* approach to estimating achievement production functions. In its most common form, the 'value-added' specification relates an achievement outcome measure to contemporaneous school and family input measures and a lagged (baseline) achievement measure. Thus, it differs from the 'contemporaneous' specification only in the inclusion of the baseline achievement measure, which is taken to be a sufficient statistic for unobserved input histories as well as the unobserved endowment of mental capacity. Evidence based on the value-added specification is generally regarded as being better (i.e. more convincing) than that based on a contemporaneous specification. See, for example, discussion in Hanushek (1996, 2003) and Krueger (2000). However, as we show below, the value-added formulation also imposes strong assumptions on the underlying production technology, and the inclusion of a lagged test score as a conditioning variables makes the model highly susceptible to endogeneity bias when data on some of the relevant inputs are missing, even if the omitted inputs are orthogonal to the included inputs.

To simplify notation, let $\boldsymbol{X}$ denote the vector of family and school input and $\boldsymbol{X}(a)$ their input histories up to age a, the value-added specification assumes that (3) can be written as a function only of a baseline test score and contemporaneous inputs (inputs applied between the baseline measure and a current measure). Without loss of generality, assume the baseline test is administered at $a-1$, in which case the value-added model assumes

$$T_{ija} = T_a\left\{\boldsymbol{X}_{ija}, T_{a-1}\left[\boldsymbol{X}_{ij}(a-1), \mu_{ij0}\right], \eta_{ija}\right\}. \tag{5}$$

Value-added regression specification usually treat the arguments in (3) as additively separable and the parameters as non-age varying, which leads to the estimating equation:

$$T_{ija} = \boldsymbol{X}_{ija}\boldsymbol{\alpha} + \gamma T_{ij,a-1} + \eta_{ija}. \tag{6}$$

A more restrictive specification sometimes adopted in the literature sets the parameter on the lagged achievement test score to one ($\gamma = 1$) and rewrites (6) as

$$T_{ija} - T_{ij,a-1} = \boldsymbol{X}_{ija}\boldsymbol{\alpha} + \eta_{ija}, \tag{7}$$

which expresses the test score gain solely as a function of contemporaneous inputs.

To understand the restrictions the value-added formulation implies for the true technology function, consider the regression analog of the true technology (3), namely

$$T_{ija} = \boldsymbol{X}_{ija}\boldsymbol{\alpha}_1 + \boldsymbol{X}_{ija-1}\boldsymbol{\alpha}_2 + \cdots + \boldsymbol{X}_{ij1}\boldsymbol{\alpha}_a + \beta\mu_{ij0} + \varepsilon_{ija}, \tag{8}$$

which imposes the assumption that

(i) *The $T_a(\cdot)$ function is non-age-varying, at least over the ages used in implementing the value-added model.*

Equation (8) does, however, allow the effects of inputs to vary with the temporal distance between the time the inputs were applied and the time of the test score measure. Subtracting $\gamma T_{ij,a-1}$ from both sides of (8) and collecting terms gives,

$$\begin{aligned} T_{ija} - \gamma T_{ij,a-1} = {} & \boldsymbol{X}_{ija}\boldsymbol{\alpha}_1 + \boldsymbol{X}_{ija-1}(\boldsymbol{\alpha}_2 - \gamma\boldsymbol{\alpha}_1) + \cdots + \boldsymbol{X}_{ij1}(\boldsymbol{\alpha}_a - \gamma\boldsymbol{\alpha}_{a-1}) \\ & + (\beta_a - \gamma\beta_{a-1})\mu_{ij0} + \varepsilon_{ija} - \gamma\varepsilon_{ij,a-1}. \end{aligned} \tag{9}$$

We are interested in determining the conditions under which (9) reduces to (6). Two conditions in addition to (i) must be met:

(ii) *Input coefficients must be geometrically (presumably) declining with distance, as measured by age, from the achievement measurement, i.e. for all j, and the rate of decline must be the same for each input.*

(iii) *The impact of the ability endowment must be geometrically declining at the same rate as input effects, i.e., $\beta_a = \gamma\beta_{a-1}$.*

For the value-added specification based on the gain in achievement (7) to be appropriate, we require in lieu of (ii) and (iii):

(ii)′ *The effect of each input must be independent of the age at which it was applied ($\alpha_j = \alpha_{j-1}$) and*

(iii)′ *The effect of the ability endowment must likewise be independent of the achievement age ($\beta_a = \beta_{a-1}$).*

With respect to estimation, if the restrictions in (9) that lead to the gain specification (7) are valid, OLS estimation of (7) would provide consistent estimates of input effects. With $\gamma \neq 1$, as in (6), however, in order that OLS estimation be consistent, the measurement error in test scores must be serially correlated and the degree of serial correlation must exactly match the rate of decay of input effects (that is η_{ija} is an iid shock). Moreover, in the more likely case that measurement error is iid, the estimate of γ will be downward biased given the necessarily positive correlation between the lagged test score measure and its measurement error.

Without additional data, because the estimate of γ is downward biased, it is not possible to determine whether the gain specification (7) or the more general value-added specification (6) is correct. γ can be estimated consistently, under assumptions (i), (ii) and (iii), if earlier observations on inputs (or test scores if

measurement errors are not serially correlated) are available to serve as instrumental variables.

If we drop the assumption that the impact of the mental capacity endowment declines at the same rate as the decay in input effects (given above by (iii)), then the error in (6) would include the endowment, i.e., assuming that $\beta_a - \gamma\beta_{a-1} = \beta'$ is a constant independent of age, yields

$$T_{ija} = \boldsymbol{X}_{ija}\boldsymbol{\alpha} + \gamma T_{ij,a-1} + \beta'\mu_{ij0} + \eta_{ija}, \tag{10}$$

instead of (6). This specification is consistent with the requirement that two sufficient statistics are necessary to fully describe the impact of past inputs and of endowment. Specifically, we can, under this assumption write, (3) as:

$$T_{ija} = T_a\{\boldsymbol{X}_{ija}, T_{ij,a-1}[\boldsymbol{X}_{ij}(a-1), \mu_{ij0}], \mu_{ij,a-1}[\boldsymbol{X}_{ij}(a-1), \mu_{ij0}], \eta_{ija}\}. \tag{11}$$

In this formulation of the technology, one of the sufficient statistics is, as before, the measure of achievement at the baseline age. The second is mental capacity at the baseline age, which can be given either of two equivalent interpretations. One is that mental capacity is a quasi-fixed input that may differ from the endowment at conception.[17] The other interpretation is that mental capacity is non-malleable (fixed for life at conception), but has an age-varying impact on achievement that reflects changing efficiency in the use of mental capacity, i.e., in a more general formulation may be age-specific and may explicitly depend on input histories. These two interpretations are observationally equivalent given the non-observability of capacity.[18]

Estimation of (10) by OLS is problematic. As with the contemporaneous specification, one requirement for OLS to be consistent is that contemporaneous inputs and unobserved mental capacity be orthogonal.[19] However, even if that orthogonality condition were not violated, OLS estimation of (6) would still be biased, because baseline achievement must be correlated with endowed mental capacity. Thus, any value-added model that admits to the presence of unobserved endowments must also recognise that baseline achievement will then logically be endogenous. If the endogeneity is not taken into account, then the resulting bias will not only affect the estimate of γ but may be transmitted to the estimates of all the contemporaneous input effects.

In an optimising behavioural model, we would expect family and school input choices to be affected by baseline achievement, particularly if, as in (10), baseline achievement has persistent effects on achievement in future time periods.[20] It is

[17] Conditional on a given genetic endowment of mental capacity, experiences within the womb and post-birth can subtract from mental capacity, even permanently (for example, as is the case in foetal alcohol syndrome), or possibly enhance it through environmental stimulation.

[18] This representation is consistent with information processing theories found in the development psychology literature. For example, Case (1992) postulates that knowledge increases the efficiency with which capacity is utilised and also that the growth in capacity is the result of neurological maturation.

[19] If mental capacity is malleable, then this specification requires that contemporaneous input levels be uncorrelated with contemporaneous mental capacity. To the extent that input prices, wages and income have some permanence, we would expect input choices to be correlated over time and also to be related to mental capacity through optimising behaviour.

[20] For example, schools often use achievement scores in deciding whether to allocate students to learning-disabled or gifted classes.

possible to obtain consistent estimates of the parameters in (10) if there exists a third (earlier) observation on achievement, along with the data on the input set $\boldsymbol{X}_{ij,a-1}$. In that case, with the assumptions already embedded in (10), namely that

(iii) *mental capacity is not malleable and its effect is the same at all ages, and*
(iv) *input effects are not age-specific,*

a simple differencing procedure can be used to consistently estimate (10).

Value-added specification in the presence of omitted variables. So far, we have assumed that there are no missing contemporaneous inputs. However, suppose instead that ε_{ija} contains unmeasured contemporaneous inputs. Further, to make the argument most strongly, suppose that the missing inputs are orthogonal to the included inputs. In this case, neither OLS estimation of (6) nor applying OLS to a differenced form of (10) will provide consistent estimates of input effects. Recall that the residual in (6) or (10), as derived from (8), is a composite of the underlying current and baseline period residuals in (9), $\eta_{ija} = \varepsilon_{ija} - \gamma\varepsilon_{ija-1}$.

When ε_{ija} contain omitted inputs, baseline achievement, $T_{ij,a-1}$, will likely be correlated with the composite residual for two reasons. First, baseline achievement, $T_{ij,a-1}$, must be correlated with its own contemporaneous omitted inputs contained in ε_{ija-1}. Second, to the extent that omitted inputs are subject to choice, optimising behaviour will create a correlation between the contemporaneous omitted inputs and baseline achievement. For example, parents may respond to realised poor achievement by increasing family inputs (such as parental time or providing tutors).

In the form most commonly adopted, the data requirements of a value-added specification are only slightly more demanding than those of the contemporaneous specification. One additional variable – a baseline test score – is all that is required. Evidence based on value-added specifications is generally regarded as being superior to that based on contemporaneous specifications; however the benefits of a value-added approach seem less clear when the potential for omitted data on inputs and endowments is taken into account.

2.3.3. *Estimation of the cumulative specification*

Direct estimation of the cumulative specification given by (3) requires data on both contemporaneous and historical family and school inputs. We next consider several different approaches to directly estimating the model given in (3), assuming that we have data on current and past inputs but do not observe endowment. Table 3 summarises the different estimators. In the discussion that follows, we impose the following assumption concerning omitted inputs:

(i) *any omitted inputs and measurement error in test scores are uncorrelated with included inputs*

Under this assumption, the problem in estimating (3) is that behaviour in the choice of inputs may induce correlations between the observable inputs and unobserved child endowments. A number of methods have been employed in the literature to deal with endowment heterogeneity, although few studies have incorporated contemporaneous and past inputs and none simultaneously with child, parental and school inputs. One method makes use of observations on multiple children within the same household or family (siblings or cousins) and the other makes use of multiple measurements for the same child at different ages.

In describing these methods and the properties of estimators based on them we take the following version of (8) as the baseline specification:

$$T_{ija} = \boldsymbol{X}_{ija}\boldsymbol{\alpha}_1^a + \boldsymbol{X}_{ija-1}\boldsymbol{\alpha}_2^a + \cdots + \boldsymbol{X}_{ij1}\boldsymbol{\alpha}_a^a + \beta_a\mu_{ij0} + \varepsilon_{ij}(a). \tag{8'}$$

Equation (8′) generalises (8) in that input effects vary not only with the distance between the application of inputs and the achievement measure (as indicated by the parameter subscripts), but also with age itself (as indicated by the parameter superscripts).[21] As the notation indicates, the residual includes all current and past unmeasured factors.

The cumulative model without endowment-input correlations. If there were no link between input choices and unobserved endowments, then OLS estimation of (8′) would be consistent under assumption (i). However, optimising behaviour on the part of parents and schools suggests that investments in children are likely to be correlated with child endowments. The estimators we consider next allow in different ways for such correlations.

Within-estimators. A class of estimators used to 'control' for permanent unobservable factors, such as endowed mental capacity, makes use of variation across observations within which the unobservable factor is assumed to be fixed. Two such 'fixed effect' estimators that are prominent in this literature use variation that occurs either within families (across siblings) or within children (at different ages). It is useful in what follows to rewrite (8′) for two different ages, a and a':

$$\begin{aligned} T_{ija} &= \boldsymbol{X}_{ija}\boldsymbol{\alpha}_1^a + \boldsymbol{X}_{ij,a-1}\boldsymbol{\alpha}_2^a + \cdots + \boldsymbol{X}_{ij,a'+1}\boldsymbol{\alpha}_{a-a'}^a \\ &\quad + \boldsymbol{X}_{ij,a'}\boldsymbol{\alpha}_{a-a'+1}^a + \boldsymbol{X}_{ij,a'-1}\boldsymbol{\alpha}_{a-a'+2}^a + \cdots + \boldsymbol{X}_{ij2}\boldsymbol{\alpha}_{a-1}^a + \boldsymbol{X}_{ij1}\boldsymbol{\alpha}_a^a + \beta_a\mu_{ij0} + \varepsilon_{ij}(a) \\ T_{ija'} &= \boldsymbol{X}_{ij,a'}\boldsymbol{\alpha}_1^{a'} + \boldsymbol{X}_{ij,a'-1}\boldsymbol{\alpha}_2^{a'} + \cdots + \boldsymbol{X}_{ij2}\boldsymbol{\alpha}_{a'-1}^{a'} + \boldsymbol{X}_{ij1}\boldsymbol{\alpha}_{a'}^{a'} + \beta_{a'}\mu_{ij0} + \varepsilon_{ij}(a'). \end{aligned} \tag{12}$$

It can be seen from (12) that in this general formulation, input effects differ both by the age at which the input is applied and by the distance in time from the achievement measure. Specifically, the parameter α_x^a indicates the effect of an input on an achievement measure taken at age a that is $x-1$ periods removed from age a. Note that what we call contemporaneous inputs are, as a matter of

[21] Note that the effect of the capacity endowment, while allowed to vary with age, does not depend on current or past inputs.

convention, zero periods removed from the age of the measurement although they may be thought of as being applied between a and $a-1$.

As an example, suppose we are looking at a child's achievement at age $a=6$. In (12), the effect of reading to a child at age 3 on the child's achievement at age 6 may differ from the effect of reading to the child at age 2 ($\alpha_4^6 \neq \alpha_5^6$) because the inputs differ in their distance from the achievement measure. In addition, the effect of reading to the child at age 3 on achievement at age 5 may differ from its effects at age 6 ($\alpha_3^5 \neq \alpha_4^6$), again because it is more distant from the achievement measure, and also may differ from the effect of reading to the child at age 4 on reading achievement at age 6 ($\alpha_3^5 \neq \alpha_3^6$) because the efficacy of reading to a child on subsequent performance may depend on the child's age.

Within-family estimators. Within-family estimators exploit the fact that children of the same parents (or grandparents) have a common heritable component. In particular, assume that endowed mental capacity can be decomposed into a family-specific component and an orthogonal child-specific component, denoted as μ_0^f and μ_0^c. Thus, siblings (or cousins) have in common the family component, but have their own individual-specific child components.

Rewriting (8′) to accommodate this modification yields

$$T_{ija} = \boldsymbol{X}_{ija}^a \boldsymbol{\alpha}_1 + \boldsymbol{X}_{ija-1}^a \boldsymbol{\alpha}_2 + \cdots + \boldsymbol{X}_{ij1}^a \boldsymbol{\alpha}_a + \beta_a \mu_{ij0}^f + \beta_a \mu_{ij0}^c + \varepsilon_{ija}. \tag{8''}$$

Now, suppose that longitudinal household data on achievement test scores and on current and past inputs are available on multiple siblings. We distinguish between two types of data. In the first case, data are available on siblings at the same age. Notice that unless the siblings are twins, the calendar time at which achievement measures are obtained must differ. In the second case, data are available on siblings in the same calendar year, which generally means that they will differ in age.

Case I. Data collected on siblings (or cousins) at same age, different calendar time. The within-family estimator is based on sibling differences, which eliminates the family-specific component of endowment but not the child-specific component. Consider the estimator in the case of two siblings, denoted by i and i' observed at the same age a.

Differencing (8″) yields

$$\begin{aligned} T_{ija} - T_{i'ja} = {} & (\boldsymbol{X}_{ija} - \boldsymbol{X}_{i'ja})\boldsymbol{\alpha}_1^a + \cdots + (\boldsymbol{X}_{ij1} - \boldsymbol{X}_{i'j1})\boldsymbol{\alpha}_a^a + [\beta_a(\mu_{ij0}^c - \mu_{i'j0}^c) \\ & + \varepsilon_{ij}(a) - \varepsilon_{i'j}(a)]. \end{aligned} \tag{13}$$

In estimation the residual term will include all the terms within the square brackets. Consistent estimation of input effects, therefore, requires that inputs associated with any child not respond either to own or sibling child-specific endowment components:

(ii) *Inputs choices may depend on family-specific endowments, but must be unresponsive to child-specific endowments*

Furthermore, given that achievement is measured for each sibling at the same age, the older child's achievement observation (say child i) will have occurred at a calendar time prior to the younger sibling's observation. Thus, the older sibling's achievement outcome was known at the time input decisions for the younger child were made, at the ages of the younger child between the older and younger child's achievement observations. Thus, consistent estimation of (11) by OLS requires the following assumption, in addition to (i) and (ii):

(iii) *Input choices are unresponsive to prior own and sibling outcomes* (*otherwise the realisations of* $\varepsilon_{ij}(a)$ *will affect some of the inputs to sibling* i')

In essence, intra-household allocation decision must be made ignoring child-specific endowments and prior outcomes of all the children in the household. The within-child estimator considered below relaxes this assumption.

Case II. Data collected on siblings (or cousins) at same calendar time, different ages. The within-family estimator based on siblings of different ages can be viewed as a special case of the within-child estimator based on test scores of the same child measured at different ages.

Within-child estimators. Within-child estimators are feasible when there are multiple observations on achievement outcomes and on inputs for a given child at different ages. Consider differencing the achievement technology at the two ages as shown in (13). This procedure, after grouping inputs applied at the same age, yields

$$\begin{aligned} T_{ija} - T_{ija'} = {} & \boldsymbol{X}_{ija}\boldsymbol{\alpha}_1^a + \boldsymbol{X}_{ija-1}\boldsymbol{\alpha}_2^a + \cdots \boldsymbol{X}_{i,j,a'+1}\boldsymbol{\alpha}_{a-a'}^a + \boldsymbol{X}_{ija'}(\boldsymbol{\alpha}_{a-a'+1}^a - \boldsymbol{\alpha}_1^{a'}) \\ & + \boldsymbol{X}_{ij,a'-1}(\boldsymbol{\alpha}_{a-a'+2}^a - \boldsymbol{\alpha}_2^{a'}) + \cdots + \boldsymbol{X}_{ij2}(\boldsymbol{\alpha}_{a-1}^a - \boldsymbol{\alpha}_{a'-1}^{a'}) \\ & + \boldsymbol{X}_{ij1}(\boldsymbol{\alpha}_a^a - \boldsymbol{\alpha}_{a'}^{a'}) + (\beta_a - \beta_{a'})\mu_{ij0} + \varepsilon_{ij}(a) - \varepsilon_{ij}(a'). \end{aligned} \tag{14}$$

Without any restrictions on the relationship among parameters, the within-child estimator recovers (a) age-specific input effects for the inputs that are applied between the two age observations and (b) differences in parameters that depend on both age and time from the achievement measure for inputs applied contemporaneously or prior to the earlier achievement observation.

The parameters of (14) can be consistently estimated under the following assumptions, in lieu of (ii) and (iii). The first is that

(iv) *the impact of the capacity endowment on achievement must be independent of age* ($\beta_a = \beta_{a'}$) *in which case the differencing eliminates the endowment from* (14)

In that case, orthogonality between input choices and capacity endowments need not be assumed. Second, similar to the within-family estimator based on sibling observations at the same age, because any prior achievement outcome is known when later input decisions are made, it is necessary to assume.

(v) *later input choices are invariant to prior own achievement outcomes*

The difference between the within-child estimator and the within-family estimator based on observations for siblings that differ in age is that in the latter only the family component of the endowment disappears from (14). Thus, consistency of that estimator requires the same behavioural assumption (given by (v)) with respect to intra-family allocations as did the within-family estimator based on sibling observations at the same age (it requires the earlier assumption (iii)).

As seen in (14), if the within-child estimator is necessary to obtain consistency, then coefficients associated with inputs applied at or before the age of the earliest test score observation will not be identified. Suppose, however, that the researcher is willing to impose the restriction that

(vi) *input effects are age-invariant, i.e.,* $\boldsymbol{\alpha}_x^a = \boldsymbol{\alpha}_x^{a'}$

as is often assumed in the application of fixed-effects estimtors. Then, (14) can be rewritten as

$$\begin{aligned} T_{ija} - T_{ija'} = {} & (\boldsymbol{X}_{ija} - \boldsymbol{X}_{ija'})\boldsymbol{\alpha}_1 + \cdots + (\boldsymbol{X}_{ija'} - \boldsymbol{X}_{ij1})\boldsymbol{\alpha}_{a'} \\ & + \boldsymbol{X}_{ij,a-a'-1}\boldsymbol{\alpha}_{a'+1} + \cdots + \boldsymbol{X}_{ij1}\alpha_a + \varepsilon_{ij}(a) - \varepsilon_{ij}(a') \end{aligned} \tag{15}$$

which would allow all the of the parameters of the technology to be identified.[22]

Instrumental variables within-child estimators. We now consider ways of relaxing assumption (v) that maintained that input choices do not respond to previous realisations of achievement. If the residuals in (15) consist only of unforeseen factors (e.g., randomly being ill or randomly drawing a bad teacher) and if the impact of these factors on achievement has limited persistence, then input levels prior to the earliest achievement observation (a') can serve as instrumental variables in estimating (14) or (15). For example if the achievement tests are taken at ages 8 and 5, then perhaps the set of inputs at ages earlier than age 3 might satisfy this requirement. However, even if that were the case, there are more parameters in (15) than instruments – at least as many as the number of measured inputs – so identification cannot be achieved with these orthogonality conditions alone. However, it is also the case that inputs associated with the child's siblings applied at a time sufficiently prior to the earliest observation used to implement the within-child estimator could also be used as instrumental variables.[23] Thus, when we relax

[22] It also may appear from (13) that the effects of time-invariant inputs are also identified (for example, the effect of maternal age at conception which is obviously fixed for the child). This identification is, however, illusory, stemming from the additive linear nature of the specification of the cumulative technology. A variable like maternal age at birth is not an input that is applied each period and, as such, would enter the cumulative specification only one time, although possibly in the more general case, like the capacity endowment, as having an age-varying effect. Specified in this manner, time-invariant variables would, as is usual, not be identified from within-child estimators. On the other hand, it is also evident from (13) that the effects of variables that might be unchanging over time but applied as an input in each period, e.g., if maternal hours of work was the same in every year, would be identified, but only at ages prior to the initial achievement observation.

[23] This kind of informational constraint has been used previously by Rosenzweig and Wolpin (1988, 1995) to estimate birth weight production functions.

assumption (v), we can still possibly estimate (14) or (15) consistently using own prior and sibling inputs as instrumental variables.

Some researchers have used cross-sectional variation in prices and other location-specific characteristics as instrumental variables to estimate human capital production functions.[24] One potential problem with that approach if applied directly to the baseline specification (6′) is that state-level variation in the capacity endowments of its residents will plausibly be correlated with the demand for different market or politically supplied services and products, e.g., school inputs. If applied to the within-sibling specification given by (13), such instruments will be valid to the extent that location decisions are independent of child-specific (though not necessarily family-specific) endowments and are also independent of the actual achievement realisations of the siblings.[25] Applying the same approach to within-child specification (14) or (15), given that the sample includes children who have lived in different locations, avoids the biases from omitting child-specific endowments, but would still be subject to the potential problem that families may change locations to find more suitable schools for their children based on their prior achievement.

Finally, none of the IV approaches are valid if omitted inputs are not orthogonal to the included ones. Omitted inputs that reflect choices are as likely to be correlated with an instrumental variable as are included inputs. Thus, any instrument that has power will also not be valid. It is therefore important to have data that contains a large set of inputs spanning both family and school domains.

2.3.4. *An application to interpreting statistical models for assessing the effectiveness of teachers*

In the education literature, there has been a recent movement towards using estimates based on statistical models that purport to measure the effectiveness of schools and teachers as one of the criteria used in hiring, promotion, and salary-setting decisions. The goal of statistical models such as the Tennessee Value-added Assessment System (TVAAS) is to measure the marginal contribution of a particular teacher to their students' performance, that is, to estimate the production function parameters associated with the teacher input.

Let i denote the child, g the grade-level, k_g the particular teacher the child experiences in grade g, and K_g the total number of teachers at grade g. For the purposes of this example, we assume all children attend the same school. Letting T_{igk} denote the test score in a particular subject, a simple specification of the education production function that captures the essence of TVAAS, for tests taken at two different grades is,

[24] For example, Rosenzweig and Schultz (1982) estimate the birthweight production function using state level variables such as tax rates on cigarettes and measures of the extensiveness of health facilities.

[25] Altonji and Dunn (1996) use differences in location-specific aggregate variables as instruments for differences in school inputs for siblings who attended school in different locations.

$$T_{igk} = \mu_g + \alpha_{gk} + \sum_{h=1}^{g-1} \sum_{l=1}^{K_g} \alpha_{ghl} 1(k_h = l) + \varepsilon_{igk}$$
$$T_{ig-1,k'} = \mu_{g-1} + \alpha_{g-1}, k' + \sum_{h=1}^{g-2} \sum_{l=1}^{K_g} \alpha_{g-1,hl} 1(k_h = l) + \varepsilon_{ig-1,k'} \tag{16}$$

where $1(\cdot)$ is an indicator variable that takes the value 1 if the condition in parentheses is true.[26] The second term on the right-hand-side captures the influence of past teacher assignments. The measured gain in learning between grades g–1 and g associated with teacher k is given by $\mu_g - \mu_{g-1} + a_{gk}$.

Because specification (16) includes both contemporaneous and historial inputs, it can be viewed as an example of the cumulative specification. However, comparing (16) to the specification of the cumulative production function (8′), several differences emerge. The TAAVS specification excludes all family inputs, contemporaneous and historical, all school inputs other than the teacher indicator variables (contemporaneous and historical) and the child-specific endowment. Given these differences, several assumptions are necessary to obtain consistent estimates of the teacher effects in (16).

As in the class size experiment discussed in Section 1, the response of families to teacher assignments of their children will be included in the teacher effects. Therefore, contrary to the goal of the teacher assessment model, the estimates of teacher effects will correspond to policy effects rather than production function parameters. For example, if families hire outside tutors for their children to compensate for a poor teacher, the teacher effect will be misstated. A similar problem may arise if other school inputs are differentially available to teachers in a particular grade. For example, if a popular teacher has higher enrollments and class-size is omitted from the specification, then the estimated teacher effect will include the impact of class size on performance.

Finally, there is an implicit assumption of random assignment with respect to unobserved characteristics of children that are permanently related to performance (child endowment). If a particular teacher were systematically assigned to children with high endowments, the influence of the teacher on performance would be overstated. Averaging measured teacher gains over time, as is the practice in implementing TAAVS, will not eliminate this bias. However, it would seem possible to circumvent this problem by augmenting the specification to include child-specific fixed effects.

3. Model Specification Tests

The last Section described the assumptions that are required to justify different empirical approaches for estimating the cognitive achievement production function given by (3). Depending on the type of data available, some of the

[26] The actual TAAVS model, as usually implemented, allows for a general variance-covariance structure, in particular, for random coefficients associated with teacher effects. Shrinkage estimators are also commonly employed to improve the robustness properties of the estimated teacher effects. See Sanders *et al.* (1997).

assumptions are testable. We next describe some model specification tests that can be used in choosing among the alternative competing estimators. We consider two main types of tests. One is a general test of the null hypothesis that the model is correctly specified against the composite alternative hypothesis that it is misspecified. The other type of test is a standard Hausman–Wu test (Hausman, 1983; Wu, 1973; Godfrey 1990) that compares the null hypothesis model against a specific alternative model.

3.1. *General Tests of Specification Against a Composite Null of Misspecification*

In our earlier discussion of proxy variables, we already suggested a diagnostic test for the presence of omitted variables based on certain kinds of proxy variables. Because a variable such as per pupil expenditure should have no effect on achievement outcomes once all the relevant input variables are controlled, a test for omitted variables can be performed by including per pupil expenditure in the estimating equation and testing for whether its associated coefficient is non-zero.

By similar reasoning, we can construct a test of the contemporaneous specification by including historical input measures (that do not belong under the null model) and checking whether their associated coefficients are significantly different from zero. The identical test can also be applied to versions of the contemporaneous model that allow for family or child endowments, as described in Table 3.

Section 2 considered four different variations of the value-added specification (summarised in Table 3) and corresponding estimators. The key assumption of the value-added specification is that the lagged test score is a sufficient statistic for historical inputs and, in the versions of the model that do not incorporate endowments, the lagged test score is also taken to be a sufficient statistic for endowments. A simple test of the first assumption is performed by including lagged input measures in the value-added specification, which should have no additional explanatory power under the sufficiency assumption. A way of testing the second assumption, regarding endowments, is described below.

We can similarly test the assumption maintained by the cumulative-within-estimators that input choices do not respond to past achievement realisations (assumptions (iii) and (v) at the end of Section 2.4). One test can be based on the observation that inputs chosen after the date achievement is measured, i.e. future inputs, should not enter the current period achievement production function.[27] When data on future inputs are available, such a test can be applied to specifications (8″), (13) and (15). If the test rejects in the most general specification (15), which allows input choices to be correlated with child-specific endowments, then the remaining option is to estimate the model by one the IV strategies described above the would allow current input choices to depend on earlier achievement realisations.

[27] This type of test is analogous to the causality test introduced by Sims (1972) in the context of aggregate time series data.

3.2. *Test of the Null Specification Against a Specific Alternative*

In addition to the general tests described above, we can also test the null hypothesis that a particular model is correctly specified against an alternative specification, using either direct tests of estimated model coefficients or a Hausman–Wu test. For example, in Section 3 we showed that the value-added specification placed restrictions on model coefficients in the cumulative specification. (The restrictions are summarised in Table 3.) These restrictions can be tested directly, assuming that enough data are available to estimate the cumulative specification.[28]

A Hausman–Wu test requires that under the null, both the null hypothesis estimator and the alternative estimator are both consistent, while under the alternative only the alternative estimator is consistent. For example, in Table 3, the cumulative model that allows for child or family endowments nests the cumulative model without endowments (or with endowments that are orthogonal to included inputs). Under the null that endowments are uncorrelated with inputs, the OLS estimator is consistent. Under the alternative, OLS applied to (8) is inconsistent. Therefore, a test can be based on a comparison of estimated model coefficients under the two different models, denoted by III.1 and III.2 or III.3 in Table 3.[29] The test-statistic is given by

$$N(\hat{\beta}_{H_A} - \hat{\beta}_{H_0})'(\hat{V}_{H_A} - \hat{V}_{H_0})(\hat{\beta}_{H_A} - \hat{\beta}_{H_0}) \sim \chi^2(k),$$

where k is the dimensionality of $\hat{\beta}_{H_A}$ and $\hat{\beta}_{H_0}$, N is the sample size, and $\hat{V}_{H_A}$ and $\hat{V}_{H_0}$ are the components of the variance-covariance matrix associated with $\hat{\beta}_{H_A}$ and $\hat{\beta}_{H_0}$.[30]

We can similarly test the hypothesis that input choices do not respond to earlier achievement realisations through comparisons of model coefficients estimated by within estimators, obtained with and without using instrumental variables.[31]

4. Conclusions

Early test score measures are important, from a economic point of view, because they have been shown to be strongly related to measures of later adult success. Murnane *et al.* (1995), Neal and Johnson (1996), Keane and Wolpin (1997), Cameron and Heckman (1993), Currie and Thomas (1999) and Dustman *et al.* (2003) all provide empirical evidence on the importance of early child development in explaining differences in schooling and adult labour market outcomes in the US and in Great Britain. This paper considers the problem of how to estimate the determinants of cognitive achievement in a way that is consistent with

[28] Ludwig (1999) uses location-specific instrumental variables to test some forms of the value-added specification.

[29] The test can only be performed on the set of model coefficients identified under both models. Fewer coefficients will be identified under the within-model as described in Section 2.

[30] The test statistic only takes this form when the estimator under the null is efficient. See Godfrey (1990).

[31] When there are more instruments than required for identification, it is possible to test over-identifying restrictions. However, a test for the validity of instruments can only be constructed under the assumption that a set of instruments is valid.

theoretical notions that child development is a cumulative process that depends on the history of family and school inputs as well as on inherited endowments.

First, we contrasted estimates of school input effects obtained from experimental and nonexperimental studies. We showed that randomised experiments and so-called 'natural' experiments generally recover policy effects and do not recover parameters of education production functions. Experimental evidence is useful for understanding the effects of particular policy interventions, but does not solve the specification problem in modelling the production of cognitive achievement. For example, experimental evidence cannot generally be used to understand the *ceteris paribus* effect of a change in class-size on achievement, to the extent that other school and family input change as a result of the experiment. Given that experimental and nonexperimental studies identify different effects, it is not surprising that experimental and nonexperimental evidence reported in the literature on the effects of school inputs, such as class size, often differs.

We then considered a variety of nonexperimental approaches to estimating the production function for achievement. A modelling framework enough to accommodate many of the estimating equations used in both the child development and education production function literatures was illucidated. We discussed strategies for dealing with different kinds of data limitations, such as proxy variables and missing data on inputs and endowments, and made explicit the identifying assumptions required to justify commonly used specifications. Many of the common specifications impose stringent assumptions on the production technology, which led us to suggest ways of relaxing these assumptions and to suggest statistical tests that can help guide the choice among competing specifications. Accounting for the variety of estimating equations adopted in both the ECD and EPF nonexperimental literatures, it is easy to see how studies, even those based on identical datasets, draw different conclusion.

University of Pennsylvania

References

Altonji, J. G. and Dunn, T. A. (1996). 'Using siblings to estimate the effect of school quality on wages', *Review of Economics and Statistics*. vol. 78, no. 4, (November) pp. 665–71.

Angrist, J. and Lavy, V. (1999). 'Using Maimonides' rule to estimate the effect of class size on scholastic achievement', *Quarterly Journal of Economics*, vol. 114, no. 2, (May), pp. 533–4.

Baharudin, R. and Luster, T. (1998). 'Factors related to the quality of the home environment', *Journal of Family Issues*, vol. 19, no. 4, (July), pp. 375–403.

Baydar, N. and Brooks-Gunn, J. (1991). 'Effects of maternal employment and child-care arrangements on preschoolers cognitive and behavioural outcomes: evidence from the Children of the NLSY', *Developmental Psychology*, vol. 27, pp. 932–45.

Becker, G. and Tomes N. (1976). 'Child endowments and the quantity and quality of children', *Journal of Political Economy*, vol. 84, no. 4, part 2, pp. S143–62.

Ben-Porath, Y. (1967). 'The production of human capital and the life-cycle of earnings', *Journal of Political Economy*, vol. 75, no. 4, part I, August, pp. 352–65.

Belsky, J. and Eggebeen, D. (1991). 'Early and extensive maternal employment and young children's socioemotional development: children of the National Longitudinal Survey of Youth', *Journal of Marriage and the Family*, vol. 53, pp. 1083–110.

Blau, F. D. and Grossberg, A. J. (1992). 'Maternal labor supply and children's cognitive development', *Review of Economics and Statistics*, vol. 74 (3)(August), pp. 474–81.

Burtless, G. (1996). 'Introduction and summary' in (G. Burtless, ed.) *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*, Washington, DC: Brookings Institution.

Cameron, S. and Heckman, J. (1993). 'The nonequivalence of high school equivalents', *Journal of Labor Economics*, vol. 11, pp. 1–47.

Case, R, (1992). *Exploring the Conceptual Underpinnings of Children's Thought and Knowledge*, Hillsdale, NJ: Erlbaum.

Coleman, J. S. (1966). *Equality of Educational Opportunity*, Washington, D.C.: Government Printing Office.

Crane, J. (1996). 'Effects of home environment, SES, and maternal test scores on mathematics achievement', *The Journal of Educational Research*, vol. 89, no. 5, pp. 305–14.

Crockett, L., Eggebeen, D. and Hawkins, A. (1993).'Father's presence and young children's behavioral and cognitive adjustment', *Journal of Family issues*, vol. 14, no. 3, pp. 355–77.

Currie, J. and Thomas, D. (1999). 'Early test scores, socioeconomic status, and future outcomes', NBER working paper, no. 6943.

Desai, S. Chase-Lansdale, P. L. and Michael R. T. (1989). 'Mother or market? effects of maternal employment on the intellectual ability of 4-year old children', *Demography*, vol. 26, pp. 545–61.

Dustmann, C., Rajah, N. and van Soest, A. (2003). 'Class size, education and wages', Economic Journal, vol. 113 (February), pp. F99–120.

Finn, J. D. and Achilles, C. M. (1990). 'Answers and questions about class size: a statewide experiment', *American Educational Research Journal*, vol. 27, (Fall), pp. 557–77.

Goldhaber, D. and Brewer, D. (1997). 'Why don't schools and teachers seem to matter? Assessing the impact of unobservables on educational productivity', *Journal of Human Resources*, vol. 32, no. 3, 505–23.

Godfrey, L. G. (1990). *Misspecification Tests in Econometrics*, Cambridge: Cambridge University Press.

Hanushek, E. A. (1986). 'The economics of schooling: production and efficiency in public schools', *Journal of Economic Literature*, vol. 24, no. 3, (September), pp. 1141–77.

Hanushek, E. A. (1996). 'School resources and student performance', in (G. Burtless, ed.) *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*, Washington, DC: Brookings Institution.

Hanushek, E. A. (1998). 'Conclusion and controversies about the effectiveness of school resources', *Economic Policy Review*, Federal Reserve Bank of New York, vol. 4, (1) (March), pp. 11–28.

Hanushek, E. A. (2002). 'Publicly provided education', NBER working paper, no 8799.

Hanushek, E. A. (2003). 'The failure of input-based schooling policies', Economic Journal, vol. 113 (February), pp. F64–98.

Hanushek, E. A., Kain, J. F. and Rivkin, S. G. (1998). 'Teachers, schools and academic achievement', NBER working paper W6691.

Hausman, J.J (1983). 'Specification tests in econometrics', *Econometrica*, vol. 46, pp. 1251–71.

Heckman, J. J. (1992). 'Randomization and social policy evaluation', in (C. Manski and I. Garfinkel, eds.), *Evaluating Welfare and Training Programs*, Cambridge, MA: Harvard University Press, pp. 201–30.

Heckman, J. J. (2000). 'Causal parameters and policy analysis in economics: a twentieth century retrospective', *Quarterly Journal of Economics*, vol. 115, pp. 45–97.

Heckman, J. J. (2001). 'Micro data, heterogeneity, and the evaluation of public policy: Nobel lecture', *Journal of Political Economy*, vol. 109, no. 4, pp. 673–748.

Heckman, J. J. and Vytlacil, E. (2000). 'Causal parameters, structural equations, treatment effects and randomized evaluations of social programs', manuscript, Univeristy of Chicago.

Heckman, J. J. and Vytlacil, E. (2001). 'Policy relevant treatment effects', *American Economic Review, Papers and Proceedings*, vol. 91, pp. 107–11.

Hedges, L. and Greenwald, R. (1996). 'Have times changed?' The relationship between school resources and student performance', in (G. Burtless, ed.) *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*, Washington, DC: Brookings Institution.

Hedges, L., Layne, R. and Greenwald, R. (1994). 'Does money matter? A meta-analysis of studies of the effects of differential school inputs on student outcomes', *Educational Researcher*, vol. 23, (April), pp. 5–14.

Imbens, G. and J. Angrist (1994): 'Identification and estimation of local average treatment effects', *Econometrica*, vol. 62, pp. 467–75.

Keane, M. P. and Wolpin, K. I. (1997). 'Career decisions of young men', *Journal of Political Economy*, vol. 105, (June), pp. 473–522.

Krueger, A. (1997). 'Experimental estimates of education production functions', NBER working paper no. 6051.

Krueger, A. (1998). 'Reassessing the view that american schools are broken', *Economic Policy Review*, Federal Research Bank of New York, March, vol. 4 (1), pp. 29–46.

Krueger, A. (2000). 'An economist's view of class size research', *Miliken Institute Award for Distinguished Economic Research paper*.

Krueger, A. (2003). 'Economic considerations and class size', ECONOMIC JOURNAL, vol. 113 (February), pp. F34–63.

Liebowitz, Arleen (1974): 'Home investments in children', *Journal of Political Economy.* vol. 82(2), Pt. I, (March/April), pp. S111–31.

Ludwig, J. (1999) 'School spending and student achievement: new evidence from longitudinal data', manuscript, Georgetown University.

McCallum, B. T. (1972). 'Relative asymptotic bias from errors of omission and measurement', *Econometrica*, vol. 40(4) (July), pp. 757–58.

Moesteller, F. (1995). 'The Tennessee study of class size in the early school grades', *The Future of Children: Critical Issues for Children and Youths*, vol. 5, (Summer/Fall), pp. 113–27.

Murnane, R. J., Maynard, R. A. and Ohls, J. C. (1981). 'Home resources and children's achievement', *The Review of Economics and Statistics*, vol. 63, no. 3, pp. 369–77.

Murnane, R. J., Willet, J. B. and Levy, F. (1995). 'The growing importance of cognitive skills in wage determination', *Review of Economics and Statistics*, vol. 77, pp. 251–66.

Neal, D. and Johnson, W. (1996). 'The role of pre-market factors in black-white wage differences', *Jounral of Political Economy*, vol. 104, no. 5, (October), pp. 869–95.

Parcel, T. L. and Menaghan, E. (1994). 'Early parental work, family social capital and early childhood outcomes', *American Journal of Sociology*, vol. 99, no. 4, pp. 972–1009.

Rosenzweig, M. and Schultz, T. P. (1982). 'Market opportunities, genetic endowments and intrafamily resource distribution: child survival in rural India', *American Economic Review*, vol. 72, no. 4, (September), pp. 803–15.

Rosenzweig, M. and Wolpin, K. I. (1988). 'Heterogeneity, intrafamily distribution and child health', *Journal of Human Resources*, vol. 23 (4) (Fall), pp. 437–61.

Rosenzweig, M. and Wolpin, K. I. (1994). 'Are there increasing returns to the intergenerational production of human capital? Maternal schooling and child intellectual achievement', *The Journal of Human Resources*, vol. 29, no. 2, (Spring), pp. 670–93.

Rosenzweig, M. and Wolpin, K. I. (1995).'Sisters, siblings and mothers: the effects of teen-age childbearing on birth outcomes', *Econometrica*, vol. 63, no. 2, (March), pp. 303–26.

Sanders, W. L., Saxton, A., and Horn, S. (1997). 'The Tennessee Value-Added Assessment System', in (J. Millman, ed.) *Grading Teachers, Grading Schools: Is Student Achievement a Valid Evaluation Measure*, California: Sage Publications.

Sims, C. (1972). 'Money, income and causality', *American Economic Review*, vol. 62 (4) (September), pp. 540–52.

Vandell, D. and Ramanan, J. (1992). 'Effects of early and recent maternal employment on children from low income families', *Child Development*, vol. 63, pp. 938–49.

Wolpin, K. I. (1995). *Empirical Methods for the Study of Labor Force Dynamics* (Fundamentals of Pure Applied Economic Series) vol. 60 (F. Welch, ed.), Luxembourg: Harwood Academic Publishers.

Wolpin, K. I. (1997). 'Determinants and consequences of the mortality and health of infants and children', in (M. Rosenzweig and O. Stark, eds.) *Handbook of Population and Family Economics*, Amsterdam: Elsevier.

Wu, D. M. (1973). 'Alternative tests of independence between stochastic regressors and disturbances', *Econometica*, vol. 41, pp. 733–50.