

Burgess, Simon

**Working Paper**

## Human Capital and Education: The State of the Art in the Economics of Education

IZA Discussion Papers, No. 9885

**Provided in Cooperation with:**

IZA – Institute of Labor Economics

*Suggested Citation:* Burgess, Simon (2016) : Human Capital and Education: The State of the Art in the Economics of Education, IZA Discussion Papers, No. 9885, Institute for the Study of Labor (IZA), Bonn

This Version is available at:

<https://hdl.handle.net/10419/141644>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

IZA DP No. 9885

**Human Capital and Education:  
The State of the Art in the Economics of Education**

Simon Burgess

April 2016

# Human Capital and Education: The State of the Art in the Economics of Education

**Simon Burgess**  
*CMPO, University of Bristol  
and IZA*

Discussion Paper No. 9885  
April 2016

IZA

P.O. Box 7240  
53072 Bonn  
Germany

Phone: +49-228-3894-0  
Fax: +49-228-3894-180  
E-mail: [iza@iza.org](mailto:iza@iza.org)

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

## ABSTRACT

### **Human Capital and Education: The State of the Art in the Economics of Education\***

This review describes the research frontier on human capital and education in economics research. It delineates what is known and largely agreed, and what are the most promising lines for future research. The approach will be to explain clearly and precisely the research evidence, in a way that makes this accessible to a wide audience. The survey has two particular aims. First, it draws out the implications for key education policy issues, highlighting which policy ideas can be supported by the economics research. To do this, my focus is on research that identifies causal effects. Second, it tries to identify some of the big open research questions and policy knowledge gaps in this field.

JEL Classification: I20

Keywords: education, human capital

Corresponding author:

Simon Burgess  
Centre for Market and Public Organisation  
School of Economics, Finance and Management  
University of Bristol  
2C1, The Priory Road Complex  
Priory Road, Clifton BS8 1TU, Bristol  
United Kingdom  
E-mail: [simon.burgess@bristol.ac.uk](mailto:simon.burgess@bristol.ac.uk)

---

\* This evidence review is produced for COEURE: Cooperation on European Research in Economics. "This is a cooperative action financed by the European Commission within the FP7, a forum which brings together the key stakeholders in the European economic research space. Its objective is to formulate an Agenda for Research Funding for Economics in Europe (ARFEE)." For more details, see: <http://www.coeure.eu/>. I would like to thank a number of people for their comments and suggestions. Most thanks of all to Susanna Loeb, who was particularly helpful and made a number of great suggestions to improve the first draft. Many thanks also to Paolo Battaglia, Lex Borghans, Gabriella Conti, Neil Davies, Matt Dickson, Emla Fitzsimons, Robin Naylor, Kjell Salvanes, Ismael Sanz, Paolo Sestito, Helen Simpson, Eleanor Stringer and Anna Vignoles. I would also like to thank Julia Belau, Marc Ivaldi and Wolfgang Leininger for comments and for organisational support. All errors and omissions are my own.

## Executive Summary

Education is crucially important for many of the policy outcomes that citizens and politicians care about. At an individual level, your education affects your earnings, your employability, and your chance of succeeding in life having started in a disadvantaged neighbourhood. It also affects your health, future family structure, intellectual fulfilment and other aspects of a good life. At a national level, a country's stock of skills matters hugely for its prosperity and growth rate. The distribution of skills is a big determinant of inequality, and the relationship of a person's skills to their background is central to the degree of social or intergenerational mobility.

While many disciplines contribute to our understanding of education, economic analysis can offer key insights for policy-makers. It has a strongly quantitative approach, and a clear framework for understanding the decisions and actions of all the relevant actors. Most importantly, it brings a relentless focus on trying to establish causal links between policy variables and outcomes. While descriptive studies can be invaluable for improving our understanding, policy decisions can only be securely made on the basis of causal relationships.

This paper reviews the current state of the evidence in the economics of education. The aim is to identify the frontier of knowledge in this field, to present what is largely agreed, and to sketch out what the open questions for research are. Additionally, the most promising policy options are highlighted.

The following issues have been covered in detail:

- Family background, family investment and the formation of human capital;
- The growth of, and interconnections between, cognitive and non-cognitive skills;
- Early childhood interventions;
- Challenges to the basic Mincer model of the demand for education;
- The role of school resources and particularly class size in changing educational outcomes;
- Measuring teacher effectiveness and identifying effective teachers;
- Policies to improve teacher effectiveness including hiring, retaining and performance pay;
- Policies to improve teacher effectiveness including teacher training and development;
- School choice, school competition and attainment;
- Tracking, elite schools and inequality;
- Neighbourhood schools and access to high-performing schools;
- School accountability;
- School autonomy;
- The rate of return to higher education including heterogeneous returns;
- Access to university and social mobility;
- Vocational education, training and adult skills.

Other topics have also been covered in less depth.

A few broad areas have been singled out as the best options for policy. This discussion can only be at a relatively high level of generality as each country of the EU will have its own institutions, laws, traditions and issues.

Having noted that, the most promising areas for policy are:

- Teacher effectiveness is by far the most important thing that matters in school. The difference in outcomes for pupils taught by effective or ineffective teachers is huge. More research is needed to better define promising policy options, but the two possibilities are to raise average effectiveness by selecting better teachers by changing hiring, retention and contract structure; or to raise the average by raising the effectiveness of each teacher through more focussed initial teacher training and by development once in post.
- Investment in the early years of child development is very important. It is essential that this not be understood to mean that investment in later years is pointless; that is clearly not the case. Policies to provide more high-quality child care, parenting programmes, and parental leave programmes are crucial, and we are beginning to learn what the best features of such schemes are. Such interventions are typically more important in disadvantaged neighbourhoods.
- A coherent market structure for schools to operate in is very important. This is a structure with strong accountability, largely autonomous schools, and centralised tests at the end of compulsory schooling. The content of the accountability criteria, the nature of schools' autonomy and the design of the tests are still being researched, but within and across country evidence shows the importance of having such structures.

The review also highlights some questions where more research would be particularly valuable. This is, of course, not to say that research in other areas of the economics of education would not also be useful. These include greater understanding of the formation of human capital, modelling the heterogeneity of returns to education, delineating the circumstances when extra resources do improve outcomes. There are also a number of topics under the heading of teacher effectiveness. Turning to school markets, further research would be very valuable on school accountability and autonomy, the costs and benefits of tracking, and in particular on the practices of effective schools. Finally, there is so much yet to learn in the economics of universities and even more still on vocational education and training.

Almost all of this work is empirical and so requires data. For many years, survey data – such as the Labour Force Survey – has been the mainstay of the economics of education. But in more recent years, researchers have exploited numerous other types of data to great effect, and indeed many of the most important recent breakthroughs have come from non-traditional data types. As discussed in the review, these include the use of birth cohort datasets, administrative or register data, internationally comparative datasets, data generated by field experiments, and psychological, neuroscientific and genetic data. An important role for research funding in this field is surely to support the creation, maintenance and use of such datasets throughout the EU.

# Contents

<b>Executive Summary .....</b>	<b>2</b>
<b>1. Introduction .....</b>	<b>6</b>
a) <i>Why are human capital and education important?</i> .....	7
b) <i>What can economics contribute to the study of education?</i> .....	8
c) <i>Organisation of this review</i> .....	8
<b>2. The aims of education – rates of return, inequality and social mobility .....</b>	<b>9</b>
a) <i>Overview: skills, inequality and social mobility</i> .....	9
b) <i>The return to education - Challenges to the Mincer model</i> .....	10
i. <i>Dealing with the endogeneity of schooling</i> .....	11
ii. <i>Heterogeneity in returns</i> .....	12
iii. <i>Life-course heterogeneity and cohort effects</i> .....	13
iv. <i>Uncertainty and the sequential revelation of information</i> .....	13
c) <i>The rate of return to going to university</i> .....	14
d) <i>Access to university and social mobility</i> .....	16
e) <i>Widening the concept of returns to education</i> .....	17
<b>3. The production of education – children and families; pupils, teachers and schools.....</b>	<b>18</b>
a) <i>The Formation of Human Capital</i> .....	18
i. <i>Multiple skills</i> .....	19
ii. <i>Modelling the dynamics of human capital formation and parental investment in their children</i> .....	20
iii. <i>Family background and the development of human capital</i> .....	23
iv. <i>Early childhood interventions</i> .....	24
b) <i>School effectiveness – resources, class size, teachers and pupils</i> .....	26
i. <i>School resources</i> .....	26
ii. <i>Teachers</i> .....	29
iii. <i>Pupil effort</i> .....	39
iv. <i>Peers</i> .....	41
c) <i>The supply of higher education</i> .....	43
i. <i>Competition</i> .....	43
ii. <i>Spill-overs</i> .....	44
d) <i>Training, vocational skills and adult skills</i> .....	45
<b>4. Education systems – assignment, admissions, accountability, autonomy.....</b>	<b>47</b>
a) <i>Assignment mechanisms</i> .....	47
i. <i>What are parental preferences for schools?</i> .....	48
ii. <i>Pupil assignment by school choice</i> .....	49
iii. <i>What is the effect of tracking and selective schools on the distribution of attainment?</i> .....	53
iv. <i>What are the effects of neighbourhood schooling assignment rules?</i> .....	54
v. <i>Assigning by income: private schools and the state sector</i> .....	55
b) <i>Accountability, autonomy and regulation</i> .....	56

i.	School Accountability .....	56
ii.	School Autonomy .....	59
iii.	School leadership .....	62
<b>5.</b>	<b>Data for education research .....</b>	<b>64</b>
a)	<i>Survey Data</i> .....	64
b)	<i>Birth cohort data</i> .....	65
c)	<i>International comparative education data</i> .....	65
d)	<i>Administrative or register data</i> .....	66
e)	<i>Data from field experiments</i> .....	66
f)	<i>Other data</i> .....	67
<b>6.</b>	<b>Education Policy and Future Research .....</b>	<b>67</b>
a)	<i>Putting it all together: Attainment, skills, inequality and sorting</i> .....	67
b)	<i>Policy</i> .....	69
i.	Improving teacher effectiveness.....	69
ii.	Ensuring high quality 'early years' opportunities.....	70
iii.	Designing the optimal school market structure.....	70
iv.	Resources for schools and class-size.....	71
v.	Classroom IT .....	71
c)	<i>Future research priorities in the economics of education</i> .....	71
i.	Formation of human capital:.....	71
ii.	Demand for education: .....	72
iii.	School resources: .....	72
iv.	Teachers and teaching: .....	72
v.	School assignment mechanisms: .....	73
vi.	School accountability: .....	73
vii.	School autonomy and the practices of effective schools: .....	73
viii.	Universities: .....	73
ix.	Vocational education: .....	74
x.	Data:.....	74
	<b>References.....</b>	<b>75</b>



## 1. Introduction

Education matters crucially for many of the things that citizens and politicians care about. It is important for individual earnings and for national prosperity. At a personal level, it matters for getting on in life, and at a national level it matters for social mobility. It is a central determinant of inequality and poverty. Of course, providing education costs a lot: on average in 2011 OECD countries spent over 6% of their GDP on educational institutions; and it accounted for almost 13% of total public spending in the OECD (<http://www.oecd.org/edu/Education-at-a-Glance-2014.pdf>), so governments are keen to make it as productive as possible. And schooling takes up a lot of time in young lives – if you're under 20 years old, being at school, thinking about school and doing school work take up a huge fraction of your time awake, on average perhaps around 10,000 hours in school over the OECD. And in older lives too, parents of school-age children also spend a lot of time, energy and stress worrying about their child's education.

Unsurprisingly then, there has been a lot of research on education. And there has been a lot of progress, a number of things that researchers are now fairly confident about. But there are also many open questions, and no doubt new questions yet to be asked, so a great deal of research is still needed. One of the corollaries of this is that more and different datasets are needed. New knowledge has been gained by using traditional surveys, including the difference in earnings that people receive for having higher skills. But increasingly, new data types are being exploited in this field and it is often these that are yielding the current big breakthroughs. This is set out below, but it includes administrative data from school records; field experiment data from randomised control trials undertaken in schools; and neuroscience and psychological data from specialist questionnaires and scans.

I take 'human capital' to mean the stock of skills, traits and knowledge that an individual possesses. It is important to be clear that there are multiple valuable skills, and that human capital does not just mean IQ. It is really only relatively recently that researchers have begun to map out the range of skills that can be considered part of human capital and we cannot yet determine precisely which types of human capital matter most in particular eras and contexts. Like other capital, human capital grows through being invested in, and that investment is called education. Not all education is done in schools; families are a very important part of the process. But education in schools is perhaps the primary lever for policies on human capital.

This review will describe the research frontier on human capital and education in economics research. It will delineate what is known and largely agreed, and what are the most promising lines for future research. The approach will be to explain clearly and precisely the research evidence, in a way that makes this accessible to a wide audience. The survey has two particular aims. First, it will draw out the implications for key education policy issues, highlighting which policy ideas can be supported by the economics research. To do this, my focus will be on research that identifies causal effects. Second, it will try to identify some of the big open research questions and policy knowledge gaps in this field to inform the future research funding agenda.

It is also important to be clear what this survey is not. It is obviously not a personal attempt to update all the 45 chapters and 2795 pages of the Handbook of the Economics of Education, in a few

months, all on my own. It is a summary of the state of the art in some key areas. While it will emphasise policy issues throughout, this is not an attempt to construct a detailed manual for policy-makers in education. That would be ambitious for just one country, let alone all countries in the EU with their widely disparate educational institutional structures, legacies and issues. However, the aim is certainly to highlight general areas of potentially fruitful policy, and areas which seem less useful. The idea is to make this as free as possible of national institutional detail and to stick to general principles.

In selecting some areas to focus on, I have had to leave others out. No doubt other researchers would have made different choices, but I believe that most of the issues I have focussed on here are consensual. The three topics I have devoted most space to are teachers; the 'early years' of human capital development, and the market structure for schooling. Topics that I have omitted include curriculum content, compulsory schooling ages, gender and race differences, the overall supply of teachers, the degree of centralisation of school funding, and over-education. These are far from unimportant as research topics or as policy issues.

Education policy-making in the European Union happens on different levels. Policy is determined at a national level (and see <https://webgate.ec.europa.eu/fpfis/mwikis/eurydice/index.php/Countries> for levels of devolution), but the European Commission offers support to its members addressing common educational challenges, primarily focussing on skills deficits. The relevant framework is "Education and training 2020" (ET2020; [http://ec.europa.eu/education/policy/strategic-framework/index\\_en.htm](http://ec.europa.eu/education/policy/strategic-framework/index_en.htm)). ET2020 has four common EU objectives: enhancing lifelong learning; improving the quality and efficiency of education; promoting equity and social cohesion; and enhancing creativity, innovation and entrepreneurship. In 2015, the EU set new priorities for education again around promoting employability and skills, increasing social mobility, and also this time aiming to counteract "fanaticism" and promote democracy.

## **a) Why are human capital and education important?**

There are three large policy domains for which education is critical. First, a country's stock of skills is central to the potential for economic growth in a highly competitive international environment. Second, the distribution of that human capital is a key determinant of income inequality, ever more important with a high wage premium for skills. Third, the link between a person's human capital and their background is a fundamental determinant of social mobility and the perpetuation of disadvantage. These are discussed in detail in section 2a below.

A child's family matters a great deal in understanding the formation of human capital, and this is discussed in detail. Beyond that, the provision of education is the main channel through which public policy can affect human capital and skills. The effectiveness of an education system influences the skills that all citizens acquire. The structure of the system affects the distribution of skills; for example, whether the system increases or decreases the inequality of educational outcomes. The system has to work well in a world of (very) imperfect information, with substantial variation in attributes and across children, allowing for a diversity of outcomes, whilst also respecting other goals such as children's wellbeing and rights, parents' rights, and democratic processes.

## **b) What can economics contribute to the study of education?**

The decisions by families and individuals on how much to invest in human capital are the standard types of decisions that economics can fruitfully analyse. They involve trade-offs between current costs and future benefits, inter-related dynamic decisions and risk. The education system, whether it is set up like a market or not, nevertheless has actors who have goals and constraints and who interact in some form or other of allocative mechanism. This is well-suited to an economic analysis. Researchers are using the tools of industrial economics to understand the incentives and constraints of all the different players in the market, and to analyse their interactions. Typically in Europe and the US, education does not function as a straightforward marketplace, so there has been interest in other forms of accountability to replace pure market discipline.

Another key contribution of economics is a strong quantitative approach. The majority of research in the economics of education is empirical, and uses a range of techniques including computable general equilibrium models and programme evaluation (see Meghir and Rivkin, 2011, for a review of methods in the field). However, perhaps the most important feature is an emphasis on trying to estimate causal relationships. Causality is not everything and descriptive studies can be extremely useful, for example in identifying need for action, but a policy discussion can really only take off from causal studies.

Of course, other disciplines bring insights to education also. In recent years, economists have started to combine effectively with psychologists and neuroscientists in the study of the development of cognitive and non-cognitive abilities and traits (for example, Cunha et al, 2006), with geneticists in studying the origins of traits and abilities (Benjamin et al, 2012) and also with behavioural scientists in trying to understand motivations and the best way to design incentives (for example Levitt et al, 2012).

## **c) Organisation of this review**

The main part of the survey is organised around three concepts. First are the goals of education, principally covering research on the rate of return to schooling and what follows from that such as inequality and social mobility; wider returns to schooling are also considered. Second, research on the production of skills, the formation of human capital, both through family investments and also through schooling. Whilst we have learnt a huge amount from recent research here, there remain large areas of uncertainty. Third, given the returns and the production function, I discuss the design of education systems, including assignment, accountability and autonomy. In each of these sections – returns, production, systems – I include post-compulsory as well as compulsory schooling. Finally, I conclude by gathering together the most promising policy ideas and identifying the areas where research funding could most fruitfully be aimed.

## 2. The aims of education – rates of return, inequality and social mobility

### a) Overview: skills, inequality and social mobility

Education is central to three very important policy domains. First, human capital and education are key, causal, drivers of growth and prosperity. Second, the distribution of human capital across people is an important determinant of income inequality, ever more important with a high wage premium for skills. Third, with higher inequality has come a renewed interest in social mobility, and the relationship between a person's human capital and their background is a major determinant of social mobility.

Starting with growth, Goldin and Katz (2008) write simply that higher levels of education lead to higher labour productivity, and that higher aggregate levels of education in a country support faster national economic growth. They explain why: "Economic growth ... requires educated workers, managers, entrepreneurs, and citizens. Modern technologies must be invented, innovated, put in place and maintained." (pp. 1-2). Recent cross-country analysis bears this out. Hanushek and Woessmann (2012) show that measures of cognitive skills are strongly associated with economic growth. Previous research had found mixed evidence of a role for education in influencing growth, but Hanushek and Woessmann argue that that work used the wrong measure of attainment, and that completed years of schooling or national enrolment rates in education do not capture skills. Instead they use direct measures of cognitive skills from international tests of maths and science abilities among pupils in 50 countries. The effect size is not trivial, since even small additions to a long-run growth rate are valuable. A quarter of a standard deviation rise in the cognitive skill score implies a higher growth rate of 0.3 to 0.5 percentage points; for comparison, the authors note that the difference between the US's PISA performance and the top performers is 40% of a standard deviation.

To establish that the relationship is causal, they implement an instrument variables strategy and use school institutional features (the presence of external exit exams, the share of privately operated schools, and the centralization of decision-making) as instruments. The implication is therefore that these policies are effective drivers of growth. They have since expanded the argument at greater length in Hanushek and Woessmann (2015), and quantified the very high cost of low skills on national income in Hanushek and Woessmann (2010).

Turning to inequality, Goldin and Katz (2008) write that we can think of earnings inequality and growth as the outcomes of a 'race' between education and technology. When the education system produces skilled people at a fast rate (at least keeping up with the increasing demand for skills from technological advance) then average income rises and inequality falls. For example, they argue that this picture characterises the US for the first three quarters of the twentieth century. But when the supply of skill slows behind technological advance, then inequality rises, distinguishing the time since the 1980s. They say "the skill bias of technology did not change much across the century, nor did its rate of change. Rather, the sharp rise in inequality was largely due to an educational slowdown." (p. 8). A lot of the foundational work understanding the sharp rise in inequality was carried out by Katz

and co-authors, summarised in Katz and Autor (1999). It has been established that the higher inequality is largely accounted for by a rising premium for skills, for education, from the 1970s. Whilst a lot of the early discussion focussed on technological change, it is now clear that the return to skills depends on both demand ('technology') and supply ('education').

One of the enduring concerns in developed economies is the question of how you get on in life. Getting an education has always been part of the answer, evidenced by innumerable stories from around the world. At an individual level, education can be seen as a way out of an unpromising start in life, an escape route. Over the last decade, policy makers have focussed on this, and comparing rates of intergenerational mobility between countries (Jäntti, 2006). Intergenerational mobility or social mobility is about where you end up in an economy relative to where you started; at root a correlation between the income of the present generation and their parents' income.

Black and Devereux (2011) see a substantial shift in emphasis in economists' studies of intergenerational mobility over the previous decade, away from refining measures of persistence towards understanding the underlying mechanisms that generate that persistence. Education, skills and (natural) abilities are at the heart of this. A very useful simple model by Solon (2004) considers intergenerational transmission as depending on parents passing on genetic endowments and investing in the education of their children, on the return to that education, and on the progressivity of government policy on education. Since heritability is fixed, Black and Devereux note that we can best understand differences in intergenerational mobility by focussing on "differences in the returns to skills ... and differences in government investments [in education]" (p. 1500). Evidence from international cross-sections (Ichino et al., 2009) and across US states (Mayer and Lopoo, 2008) backs up the idea that social mobility is higher when public education is better funded. Recently, Chetty et al (2014) use administrative data to characterise cities in the US as having high or low intergenerational mobility; they show considerable variation across the country, and one of the correlated factors is the quality of primary education. Gregg et al (2013) in an international comparison stress variations in the return to education as a driver of differences in intergenerational income persistence. A much more focussed version of essentially the same question is put by Dobbie and Fryer (2011) and Fryer and Katz (2013): is a high quality school enough to break out of a cycle of intergenerational poverty? Focussing on the Harlem Children's Zone schools and using quasi-experimental methods, they answer optimistically.

So education matters centrally in many of the biggest economic policy debates. Before moving on, it is worth noting that education has been shown to have impacts on other outcomes too: health, crime, household structure, and happiness. Last in this list, but first for some people, it is a source of personal fulfilment and inspiration.

## **b) The return to education - Challenges to the Mincer model**

The central basis of the demand for education is the return on the human capital accumulated during a pupil's time in school. Principally this is a private financial return in the form of higher earnings, but there is now a much wider concept of that return (see Dickson and Harmon, 2011). The starting point for all such analyses is the basic Mincer human capital earnings function (Mincer,

1974) and the work of Becker (1964). The Mincer model relates an individual's earnings (in logs) to her years of schooling and a quadratic function of her labour market experience. Lemieux (2006) calls this "one of the most widely used models in empirical economics". There are now a number of challenges to the assumptions of this standard model, as well as extensions, widening the concept of returns, and this is where the research frontier lies.

### **i. Dealing with the endogeneity of schooling**

The issue that has received by far the most attention to date is methods for dealing with the issue that some personal characteristics associated with the number of years of schooling chosen may also influence earnings. The obvious example is innate ability: theory says more able individuals will rationally choose more schooling and also likely earn more. This biases estimates of the rate of return to schooling. Responses have included using econometric techniques to deal with this (as discussed below, for example instrumental variables, IV) and finding datasets with matched identical twins where ability and upbringing can be differenced out (Ashenfelter and Rouse, 1998).

Different choices of instrumental variables define different parameters and can produce very different 'effects' of schooling on earnings even in the same data (Carneiro, Heckman and Vytlačil, 2005; Heckman, Urzua and Vytlačil, 2006). Instrumental variables may not estimate the policy-relevant parameters nor necessarily produce an economically interpretable parameter, and, like the Mincer model in general, do not (usually) estimate a rate of return but estimate some average growth rate of earnings with schooling.

The standard interpretation of instrumental variables assumes that there is only one response to the exogenous instrument: for example, all choose to have more years of schooling. This monotonicity assumption does not sit well with models incorporating more heterogeneity or dynamic discrete choice models. Models allowing for non-monotonic reactions to instruments have recently been emerging. These develop ordered choice models and general unordered choice models, and allow heterogeneity of response to treatments; see for example, Belzil and Hansen (2002, 2007), Heckman and Navarro (2006), Heckman and Vytlačil (2005), and Heckman, Tobias and Vytlačil (2001). Belzil and Hansen (2007) provide strong support for the correlated random coefficient wage regression model as a much more robust approach to estimating returns to education. They also suggest that relaxing the monotonicity assumption has a substantial effect on the instrumental variables estimates derived from a compulsory schooling reforms, returning much lower estimates than typically found in more standard Mincer models. Belzil and Hansen (2002) find the linear model to be severely mis-specified, with low levels of education having very low returns before a more convex shape of the education-earnings profile emerges as education increases. Ability bias is a non-trivial issue, with these two factors leading OLS estimates to seriously over-estimate average returns. Even instrumental variables estimates reported in the literature are subject to this erroneous linearity assumption leading Belzil and Hansen to call into question the majority of estimates in the empirical literature. Similarly Heckman and Vytlačil (2005) suggest that when the monotonicity assumption is relaxed so that the choices of treatment are allowed to be different, the method of instrumental variables breaks down and alternative methods are required. Heckman and Vytlačil (2005) propose the estimation of marginal treatment effects, which allows different treatment effects typically

estimated by researchers to be derived, and bridges the structural estimation approach and the recent literature on instrumental variables.

Since the 1970s, one challenge to the Becker/Mincer approach to education has been the signaling view – that schooling does not increase productivity, it merely signals natural ability. This argument is rarely raised in recent research and a number of papers including Chevalier et al (2004) and Arcidiacono et al (2010) provide convincing arguments in favour of the human capital approach.

## **ii. Heterogeneity in returns**

Mincer's model assumes a common impact of schooling on earnings, both across levels of education (linearity) and across people, i.e. that the growth rate of earnings with schooling is independent of the years of schooling. While this assumption makes life easier it does not derive from theory. In fact it is likely that other factors that affect the return to schooling are also correlated with the amount of schooling, namely that people select into schooling on the basis of their individual growth rate of earnings with schooling. It is also plausible that the growth rate of earnings with schooling declines with the level of schooling.

Card (1999) offers a development of the basic Mincer model which allows for heterogeneous returns to education, known as a correlated random coefficient wage regression model. This is the standard Mincerian log wage regression function in which the coefficients, in particular the return to schooling, may be individual-specific and arbitrarily correlated with schooling. It is also more general than the Mincer specification in a number of other ways, and provides a more realistic framework to evaluate returns to schooling. Belzil and Hansen (2002, 2007), Koop and Tobias (2004) and Aakvik, Salvanes, Vaage (2010) are examples using Card's model as basis for estimating random coefficient models of schooling allowing heterogeneity in the impact of schooling. Another line of attack in this area is to use quantile regression to consider the impact of schooling at different points in the conditional distribution of earnings (Brunello et al 2009).

If heterogeneity is likely, a description of the distribution of returns is needed. Research by Heckman (1990), Heckman and Smith (1998), Aakvik, Heckman and Vytlacil (2005), Carneiro, Hansen and Heckman (2001, 2003), Cunha, Heckman and Navarro (2005, 2006) and Cunha and Heckman (2006) take this step to evaluate joint counterfactual distributions of outcomes. These models seek to recover the whole distribution of the effect of schooling on outcomes. Cunha, Heckman and Navarro estimate the full distribution of earnings for college attendees under the counter-factual scenario where they do not attend college, only high school. Similarly, they estimate the distribution of earnings for high school graduates had they attended college. In both cases, the college earnings distribution is to the right of the high school distribution, indicating that many who quit education after high school would benefit from going to college. The typical high school graduate who continued into college would realize a substantial gain in lifetime earnings, though the gain above high school earnings for those who do actually go on to college (the effect of the 'treatment on the treated') is double the gain of the typical high school graduate, indicating the importance of selection. They also highlight that the difference is not just at the mean but present over the entire distribution. Other papers by combinations of these authors present similar estimates under different scenarios regarding the labour market conditions and the information available to individuals making their schooling decisions, but in each case the full counter-factual distributions of



earnings for different education levels are presented allowing analysis of the impact of education on different groups, thereby recovering different policy parameters.

### **iii. Life-course heterogeneity and cohort effects**

The Mincer model has an implicit stationarity assumption, that is, that cross-section log-earnings-experience profiles are common across cohorts. Since several studies show that wage patterns have changed substantially over time across cohorts (see e.g. MaCurdy and Mroz, 1995; Card and Lemieux, 2001), cross-sections no longer approximate the life cycle earnings or schooling returns of any particular individual (Heckman, Lochner, Todd, 2006). Accounting for the non-stationarity of earnings over time has empirically important effects on estimated rates of return to schooling. For example, Heckman, Lochner and Todd (2006) find that in the US, the return to college education falls steadily from around 12% in the mid-1960s down to 8% at the end of the 1970s if estimated under the assumption that log-earnings-experience profiles are common across cohorts. However, estimated using panel data that allows cohorts to be considered separately and does not impose the common profile assumption, the return is actually increasing from 8% in the early 1960s through to around 12% in the 1980s.

The use of data that follows actual cohorts over the life cycle is therefore essential to accurately measure their true earnings pattern and estimate the education premiums experienced by individuals. This is likely to be an important future direction for research, and Bhuller, Mogstad and Salvanes (2014), Buscha and Dickson (2015) attempt to make progress on these issues.

### **iv. Uncertainty and the sequential revelation of information**

The Mincer model assumes individuals have perfect information on the future earnings return to schooling. This is often false: Jensen (2010) shows that perceptions of the rate of return to schooling can be significantly wrong, and providing correct information raises the demand for school years. The total gain from schooling can be thought of as a component known at the time schooling choices are made, and a remainder, introducing ex ante and ex post distinctions into the analysis of the earnings functions. From an estimation point of view there are two elements to this. First, there is variation in the return to education across individuals that they know about ex ante but which are not observable to the researcher. As a result, even if each individual were to receive their expected return to education with absolute certainty, returns would still appear to the researcher vary randomly across the population. Second, there will be random variation in returns across individuals that they do not know about ex ante arising from individuals receiving good or bad draws from their (individual specific) wage distributions at a given point in time. Though this strand of research is still relatively new, several studies demonstrate the importance of uncertainty: for example, Harmon, Hogan and Walker (2003), Palacios-Huerta (2003), and Heckman, Lochner, Todd (2008). Understanding this uncertainty can partly explain why many individuals apparently drop out of education whilst there is still a high return – schooling decisions are made ex ante whilst estimates in the literature typically recover the ex post return. If individuals are risk averse, then the uncertainty around returns at the time when decisions are made can lead to this under-investment.



Belzil and Hansen (2002) and Heckman, Lochner, Todd (2006) also offer models that allow for the sequential revelation of information and for this to affect decision-making. They find that part of the economic return to finishing high school or attending college includes the potential for completing college and securing the rewards associated with a college degree. Both sequential resolution of uncertainty and non-linearity in returns to schooling can contribute to sizeable option values.

Finally, there are more technical challenges to the basic Mincer model relating to functional form. Specifically, numerous papers reject linearity of earnings in schooling and also reject the multiplicative separability of earnings and experience (see the surveys in Card (1999), and Heckman, Lochner, Todd (2006)). And the role of the “psychic costs” of schooling was largely ignored in the earlier literature, but this is of increasing importance and provides a link between this section and section 3.b.iii: heterogeneity in individuals’ psychological traits (for example here, conscientiousness) implies variation in the value of the effort cost of schooling, and therefore variation in the decisions to acquire schooling (see above, and Cunha, Heckman and Navarro (2005) and Cunha and Heckman (2007)). Similarly, individuals’ risk and time preference and debt aversion are elements of psychic costs which will differ among individuals and impact on their decision-making. Heckman, Lochner and Todd conclude that earnings return is not the only or possibly not even the main criterion that individuals evaluate when making schooling decisions and so the modelling assumption that psychic costs are zero is quite important.

### **c) The rate of return to going to university**

There are different perspectives on the rate of return to university and on access to universities. The rate of return question is also more important because students face a real decision on what to do after compulsory schooling; if the rate of return is not high then it may be better to get a job. This choice also makes estimating causal rates of return harder. The access question is the reverse of this: if the rate of return is high but access to university is limited, this is regressive.

The rate of return to education in general was discussed above in section 2.b and many of the same points apply here. One issue likely to be more important for higher education is the heterogeneity of returns. Students chose different subjects and different time frames for their degrees and go to different institutions. As with schools, students also differ themselves in the costs they face and the returns they may receive. This means we should focus on the distribution of returns. There are major national differences in many steps of this process, not least the cost to the student of the course and other aspects of admission, and the length of time of the course. A related factor is the availability and cost of credit if students have to pay upfront.

It is straightforward to produce descriptive estimates of the earnings gap between those with and without degrees, as this information can simply be taken from Labour Force Surveys. This is discussed in a useful review of the recent US evidence in Oreopoulos and Petronijevic (2013). This approach typically produces substantial earnings gaps. One of the important findings is the diversity of returns to university in different dimensions. Across subjects for example, Walker and Zhu (2011) show very different rates of return for different subjects and also by degree class; and also significant variation across people within subjects, some quantiles low or even negative. Typically

returns to degrees in Science, Technology, Engineering and Maths (STEM) subjects are the greatest. Across institutions, Brunello and Cappellari (2008) find substantial differences in earnings depending on which university is attended, although these differences are not persistent. Nor are they sufficient to generate switching by students to better-performing universities. Naylor et al (2015) show differing rates of return by degree class. Looking across countries, Psacharopoulos (2006) produces some estimates of this type and calculates rates of return to higher education (p. 122), with a 'world average' of around 20%. Another typical finding is that the returns to post-graduate degrees are substantial.

However, this approach cannot capture the causal return to a degree because attending university is self-selected by the students. Researchers have looked for circumstances in which we can get at a causal return by exploiting natural experiments or more recently by using a regression discontinuity design. An example of an earlier natural experiment approach was to use commuting time between home and college as providing natural variation in the cost of college. Oreopoulos and Petronijevic (2013) provide a discussion of this, and report that the studies find that each year of college yields a return of about 10%. Zimmerman (2015) uses the admission criteria of a relatively low-standards university in Florida to compare the earnings of students who just got in with those who just missed out. He found that a marginal admission gives earnings gains of 22% some ten years after high school graduation. This is a regular four year college so the alternative is probably either no university or a two year course. Similarly, Hastings, Neilson and Zimmerman (2014) use the same technique and a very extensive dataset for Chile covering over a thousand degree programmes. The fact that admissions are score-based with entry thresholds means that they can estimate the causal impact of a degree on subsequent earnings. They find that returns are variable, with large positive returns to some subjects (health, science, and social science) and to highly selective programmes. They also find some programmes with low returns, and argue that policies should try to guide students toward the higher-return degrees.

On top of these studies, Oreopoulos and Petronijevic (2013) suggest that one compelling piece of evidence for the positive causal return to university is the simple fact that the graduate premium has remained high or risen further despite a very substantial rise in the number of graduates. For example, Walker and Zhu (2008) investigate this for the UK, which has seen an almost-doubling of people going to university in the UK over the 1990s; despite this huge increase in supply, they find no fall in the graduate premium for men and a slight rise for women; Harmon, Hogan and Walker (2003) come to the same conclusion studying the variance in returns to university in the UK over time.

These regression discontinuity designs, particularly Zimmerman (2015), also address a final issue under the rate of return: comparing the average return to a degree to the marginal return. This group is particularly interesting as this is the group that will be affected by policy changes. By virtue of the econometric design, Zimmerman's evidence directly assesses the gain to the marginal student of attending college versus not, and he shows a substantial return.

A credible counter-argument is that the marginal students will have the lowest returns simply by the fact that they are marginal, almost indifferent to going or not. However, Zimmerman and other evidence suggests that the return to the marginal student is at least as high as the average return (the return takes into account the alternative, and high returns can be generated by having worse

alternatives). Also, as Oreopoulos and Petronijevic (2013) point out, a richer model with a second basis for deciding to attending university or not (for example a student's cultural or social context, see Hoxby and Avery, 2012) would imply that students can be marginal overall but still see substantial financial returns.

#### **d) Access to university and social mobility**

It has been demonstrated that average and marginal rates of return to a degree are substantial. This does not necessarily mean that they are substantial for everyone at any institution doing any course, as the work of Hastings et al (2014) illustrates. Nevertheless, it seems likely that there are people who would benefit from attending university who do not go. One very active research issue is trying to understand why, and to design policies to rectify it.

The first issue is identifying the key market or policy failures. Different countries have very different university student funding regimes, so there are a limited number of points that can be made at a general level. In some countries, students are required to pay upfront for a place at university. The issue then is about whether the student (or her family) can pay; typically high tuition costs are accompanied by the option to take a loan to cover the cost, perhaps at a subsidised rate. So the first bases for access are: parental income, available credit, and the cost of the loan. Kane (2006) surveys the empirical evidence on the effects of higher education financial aid policies. He finds that a number of state-based programmes (such as the Hope Scholarship program in Georgia and the Cal Grant program in California) have quite substantial effects. Though since Dynarski (2000) shows that these effects are mostly for middle class students, this will widen rather than reduce subsequent earnings inequality. There is dispute about the importance of credit constraints in preventing access to university; Carniero and Heckman (2002) argue that low parental income is correlated with lower ability children and that it is more the latter factor that drives lower university attendance among that group. Kane (2006) summarises this debate: "even though there are some pieces of evidence that would be consistent with borrowing constraints, it is difficult to find a definitive test of the existence of borrowing constraints in the literature." (p. 1396). Relatedly, young people may be debt averse; even though the net return to the investment is likely to be positive, simply holding debt is unattractive.

Chapman (2006) looks at evidence on income-contingent loans for higher education in a number of countries, although only Sweden and the UK from Europe. He considers the Australian scheme in some detail as it has been running since 1989 and has been much studied. He finds that the introduction of a graduate tax type scheme did not change the social mix of university students, and notes more broadly across a range of case studies that the socio-economic composition of university students seems relatively insensitive to the funding rules, broadly supportive of Carniero and Heckman (2002).

If there is an upfront cost to attending university, government assistance is often available to low income families, either as a grant or as a loan. Steiner and Wrohlich (2012) study the case of Germany and are able to separately identify the role of parental income and the means-tested aid

because of the strongly non-linear relationship of the latter to the former. Using the SOEP, they find a significant but small effect of the aid programme on university enrolment.

Of course, it may not be all about money. Hoxby and Avery (2012) demonstrate the difference in college application patterns between students and highlight the potential role of peer groups and role models. If student aid is available, sometimes the application process can be dauntingly complex. A few recent studies have looked at this process in the US. Bettinger et al (2012) set up a field experiment in which low income families were given help in filling out the application forms. They were also given further information on comparative university costs in their local area. The treatment led to substantially increased applications for aid, and subsequently higher enrolment at university. They showed that it was the assistance not the information that was the driver of the change. The fact that the US state programmes noted above have much less cumbersome and complex procedures and are relatively effective, which constitutes a further piece of evidence that the Pell Grant programme and the accompanying financial aid application form (FASFA) are not well designed. Hoxby and Turner (2014) evaluate a field experiment that provided information on both the college application process and on the net costs, and also waived the application fee. They show positive results on the calibre of college applied to and on subsequent performance.

Typically, the rates of return noted above apply to students who complete their university course. So while enrolment is important, degree completion is also part of the story. One way that this manifests itself is through slow or delayed degree completion, which can in turn pose problems for the universities. Garibaldi et al (2012) consider the idea of higher tuition fees for delayed graduation to reduce this problem. They use data from Bocconi University and employ a regression discontinuity design across the different fees there to study this, and show a substantial impact. In other institutional structures, the alternative to not graduating on time is to drop out. Bettinger and Baker (2014) implement a field trial of a programme of individualised student coaching on the skills needed to succeed at university. Students assigned to the treatment programme were more likely to be still attending university a year after the programme ended.

## **e) Widening the concept of returns to education**

Findings that show a substantial causal impact of schooling on individual earnings are documented in various survey articles (Card in 1999 and 2001; Heckman, Lochner and Todd, 2006). Staying within the labour market, schooling also generates returns in the form of lower workplace injury rates and less night/evening work (Hamermesh, 1999), less unemployment, quicker job finding and lower welfare receipt (Oreopoulos and Salvanes, 2011).

There is also a considerable literature showing evidence of causal linkages between education and numerous non-pecuniary outcomes (Grossman, 2005, Oreopoulos and Salvanes, 2009; Lochner 2011). This includes improved health and increased longevity (Lleras-Muney, 2005; Oreopoulos, 2007) and greater life satisfaction (Oreopoulos, 2007, Oreopoulos and Salvanes, 2011). Teen pregnancy rates are reduced by education (Black, Devereux, and Salvanes (2005); Oreopoulos and Salvanes (2009, 2011)), and there are substantially lower divorce rates for the more highly educated

(Oreopoulos and Salvanes (2009, 2011)). Oreopoulos and Salvanes (2011) also argue that education raises patience (lower discount rates) and trust.

Schooling may also improve parenting (Del Bono et al 2015, Oreopoulos and Salvanes (2011)) and this is part of the wider research showing that education has inter-generational impacts, implying that some of the return to education is enjoyed by the children of those receiving additional schooling (see Plug, 2004; Oreopoulos, Page and Stevens, 2006; Black, Devereux and Salvanes, 2008; Dickson et al (2015)).

There is also evidence that education has broader social returns, including a lower rate of participation in crime (Lochner and Moretti, 2004; Machin et al 2011), and higher civic participation (Dee, 2004; Milligan, Moretti and Oreopoulos, 2004). Glaeser, Ponzetto, Shleifer (2005) also highlight a link between education and the strengthening of democracy relative to dictatorship.

### **3. The production of education – children and families; pupils, teachers and schools.**

The formation of human capital starts early in a person's life, even *in utero*. And it can last throughout life, though generally the most significant acquisition of skills occurs before the age of 25. There are some commonalities in skill formation across all that time-span, including the nature of investment decisions under uncertainty, the development of the brain, and the behavioural science of how people actually learn something. But there are also differences, not least different lead actors. In the earliest stages, the child is essentially a passive recipient and processor of what others provide, parental investment or the care of others. Soon enough, teachers become a key part of skill formation, and the focus is on interactions between teacher and child, parent and child. Later still, the child's own decisions are paramount, whether to study hard, what to study, recognising what her strengths are, what her preferences are; at this point too, friends and peers become important actors. This section sets out what we know, roughly following these stages, after setting out some general approaches first.

#### **a) The Formation of Human Capital**

There is a very substantial literature examining the role of family in the formation of human capital in children. Excellent recent surveys are Björklund and Salvanes (2011), Currie and Almond (2011), and Heckman and Mosso (2014). There is also theoretical work modelling the decisions that families make to invest in the human capital of their children. Much of this is based around the work of Heckman and co-authors, following Todd and Wolpin (2003), setting out in detail a model of family decisions, and collating a large body of analysis and evidence. Heckman and co-authors provide a number of recent summaries and overviews of this literature, including Heckman and Mosso (2014) and Kautz et al (2014).

This research has a number of facets, all of which rely on the others. Accordingly, there is no obvious first way in. Starting with the core investment model would make sense, but this model is built around the relatively recent idea of multiple and very different skills. So this is what I present first, to set the stage for the investment model.

### **i. Multiple skills**

One major advance in the economics of education in recent years is the recognition that there are multiple valuable skills. Typically, previous work has focussed on a single measure of ability, implicitly intellectual ability, often discussed as IQ, though Bowles et al (2001) is an early statement of the importance of other attributes. Researchers have now distinguished such cognitive skills from non-cognitive skills. The latter confusingly have different collective labels in the literature: non-cognitive skills or non-cognitive abilities, soft skills or socioemotional skills, or personality traits or even 'character'. These labels matter more than just in disciplinary disputes: character or trait implies fixed and unchangeable, whereas skills and abilities can be learned or improved, discussed further below. Once we look under the collective label, different researchers have different lists of individual skills/traits: for example, Heckman (2004) talks about 'perseverance, motivation, self-control' (p. 180); Duckworth et al (2007) analyse a trait they call "grit", variously defined as involving perseverance, motivation, persistence, zeal and tenacity. Heckman and Kautz (2012) cite "conscientiousness, perseverance, sociability, and curiosity"; Heckman and Mosso (2014) focus on conscientiousness ("tendency to be organized, responsible, and hardworking", Heckman and Kautz, 2012, p. 5). In fact, conscientiousness is perhaps the most useful psychological construct here as it is the most strongly correlated with attainment and labour market outcomes, see Conti and Heckman (2014). In the sense of "working hard" it is also of course related to motivation and the potential power of incentives for pupils. Brunello and Schlotter (2011) review the labour market impacts of non-cognitive skills and personality traits.

This focus on non-cognitive capacities makes a link to personality traits and the research of psychologists, reviewed in Borghans et al (2008) and Almlund et al (2011). Among other things, they show that preference parameters familiar in economics such as risk aversion and time preference may be related to psychology's 'Big 5' personality traits (Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism). But this is not a straightforward link and Almlund et al (2011) and Heckman and Kautz (2012) set out measurement and conceptual differences between the two approaches that need to be resolved before the evidence can be truly pooled. One aspect of this is the question of whether personality traits are stable over time. This is not resolved, though Heckman and Kautz (2012) argue that the bulk of the evidence suggests stability. To be precise, they argue that traits are stable, but evolving slowly and malleable. The last characteristic is clearly key: they argue that public policy interventions can shape non-cognitive skills.

There is evidence (see Almlund et al (2011), Cunha et al (2006)) to suggest that pupils' scores on achievement tests are influenced by their non-cognitive skills as well as their cognitive ability or IQ. In some ways, this is not surprising; conscientious pupils will study hard and do better. But for those considering achievement tests as pure measures of IQ, this is more surprising. Borghans et al (2011) use the NLSY and other datasets to show that a high fraction of the variation in scores in

achievement tests, can be explained by personality variables. Duckworth et al (2007) show that “grit” was predictive of success measures over and above IQ and conscientiousness.

Furthermore, there is evidence (see Almlund et al, 2011) that both cognitive and non-cognitive skills are valued in the labour market conditional on achievement tests (qualifications). That is to say, qualifications do not capture all of the aspects of an individual’s value to future employers. Again, this does not seem surprising, but it carries an important point. Much of the policy interest in education is centred on improving attainment in order to raise earnings, and the importance of non-cognitive skills gives another, separate, lever on that.

I return to discuss evidence on policy interventions below, but note here the evidence specific to non-cognitive skills. Heckman and Kautz (2012) argue that “personality traits can change and be changed over the life cycle but through different mechanisms and to different degrees at different ages” (p. 4). The key item of evidence is the Perry Preschool Program, a small experimental intervention targeting disadvantaged children aged three to four for two years in the US; follow-up data were collected for both treatment and control groups until age 40. Heckman, Pinto and Savelyev (2013) show that the treatment group experienced significantly better adult outcomes including higher attainment and higher earnings. The programme induced substantial but ephemeral increases in IQ and also raised in a long-lasting way non-cognitive attributes and also enhanced academic motivation. Chetty et al (2011) have found similar evidence for the long-term effects of Project STAR, discussed in more detail below (section 3.b.i).

Recent evidence from Fiorini and Keane (2014) explores the development of cognitive and non-cognitive skills of young children within the household by using time diaries. They show that the child’s cognitive skills are best developed by time spent in educational activities, particularly with her parents. However, time use appears to matter much less for the enhancement of non-cognitive skills, but maternal parenting style is very important.

The evidence on what schools more broadly can do to enhance key non-cognitive abilities is not settled, and this is surely one key area for future research. Blazar and Kraft (2015) extend the teacher effects literature discussed below to estimate teacher impacts on a range of non-cognitive outcomes. They show substantive relationships of a similar magnitude to teacher effects on attainment, only weakly correlated between cognitive and non-cognitive skills.

One of the reasons for policy interest in conscientiousness and grit is the relationship to pupil ‘effort’ in raising attainment, and the study of pupil motivations. I discuss this below, along with models of incentives. While IQ, cognitive skills and raw abstract reasoning capacity give some measure of potential ability, measures of attainment also depend on the degree of deployed effort, and that in turn on the motivations of pupils, potentially opening a role for incentives, another promising area for future research.

## **ii. Modelling the dynamics of human capital formation and parental investment in their children**

The parental investment decision is dynamic because later choices may depend on the outcome of earlier investments. The initial condition comprises the endowments that the child begins life with,

including the genetic endowment. One of the key themes is the interaction between private and public investment; that is, whether the pupil's and parents' chosen inputs are complementary to the state's, or substitutes.

The theoretical model is set out in Cunha and Heckman (2007) and estimation is discussed in detail in Cunha, Heckman and Schennach (2010); there is also a useful exposition in Currie and Almond (2011). The centre-piece of the model is a vector of skills of the child at each age. These are later interpreted in the empirical model as cognitive and non-cognitive skills. The model has a recursive structure as each period's skill set depends on the previous age's skills and investment. The initial conditions of the model are the child's initial endowment, including genetics, and her parents' characteristics, their own cognitive and non-cognitive skills and education attainment, discussed below. The technology of skill formation determines how current skills derive from past skills, new investment chosen optimally by the parents, and the environment as summarised by parental characteristics.

The model is largely standard dynamic decision making, so the key factors are the assumptions made about the technology function. The first is 'self-productivity': skills produced at one stage augment skills attained at later stages. The idea is that different skills are reinforcing and cross-fertilising, for example, a higher stock of non-cognitive skills at one age raises the stock of cognitive skills at the next (perhaps because conscientiousness improves learning). The second key characteristic of the technology is dynamic complementarity: that is, skills produced at one stage raise the productivity of investment at later ages, and so make further investment more valuable. Together, self-productivity and dynamic complementarity produce multiplier effects reflecting the idea that 'skills beget skills'.

With the technology fixed, parents optimise their investment profile over their child's growing up. Information is obviously important for this: Dizon-Ross (2014) uses an experimental setting in Malawi to show that parents' adjust their investment in their child depending on their beliefs about her academic performance. The parameters of the production function influence the optimal early years/late years investment ratio. The assumed properties of the skills production function are enough to mean that earlier investments have greater returns than later returns. As noted, there are differences in the optimal investment profile between cognitive and non-cognitive skills. But on top of that, some of these models include a role for 'sensitive periods' – particular ages when investment is more valuable, that is, more productive in creating new skills – largely derived from the field of neuroscience. After birth, children do not acquire many more neurons, but during early childhood there are dramatic changes in the connectivity between them. The idea of sensitive periods is that gross changes in connectivity are particularly sensitive to the environment (in the broadest possible sense) during those times. In fact, neuroscience suggests (see Howard-Jones and Washbrook, 2011) that the concept of sensitive periods might help understand abnormal development rather than provide a rationale for greater investment – that is, they are periods when great damage can be done, but possibly not great good. Similarly, Doyle et al (2009) motivate the optimal timing of early childhood intervention through a focus on adverse risk factors during the antenatal and early childhood periods. The science here is still at a relatively early stage and it may be that there are other sensitive periods beyond early childhood (see Blakemore's 2010 discussion of the re-organisation of adolescents' brains). Carlsson et al (2015) report that cognitive skills are malleable among 18 year-olds, simply through spending longer in school.



Cunha, Heckman and Schennach (2010) estimate a constant elasticity of substitution version of the model above on a sample from the Children of the NLSY/79. They find strong evidence in favour of self-productivity and dynamic complementarity. They also find higher substitutability for non-cognitive skills between early and late investment, and lower substitutability for producing cognitive skills, and evidence of productive interaction between cognitive and non-cognitive skills, more for non-cognitive skills promoting cognitive development than the other way around. They also argue that their findings are consistent with the idea of a sensitive period for the development of cognitive skills in early years. These estimates of the skills production technology support the idea that investing early, particularly in the development of cognitive skills, is the most valuable approach.

In terms of education, schools are essentially absent from this analysis. Cunha and Heckman (2007) count “government inputs (e.g., schooling)” (p. 36) as subsumed as part of the parental investment decision. In the empirical work, Cunha, Heckman and Schennach (2010) note that the school measures they have access to in the CNLSY are poorly measured and show “no important effects of schooling inputs on child outcomes” (p. 904) and so are excluded from the main estimation results.

But for some families, what happens in school might be a lot of the totality of investment, undertaken by others on their behalf. And schools and teachers are not just for cognitive skills and knowledge acquisition. They also attempt to develop non-cognitive skills as parents do, teaching pupils to sit still, to pay attention, to wait for rewards, to treat their studies with respect, to be organised and to work hard. Differentiating between the contributions of ‘home’ factors and ‘school’ factors in explaining attainment, and their complex interactions, is a question going back at least to the always-cited Coleman Report (Coleman et al 1966). Hoxby’s (2001) insight that the relationship between parental activity and child outcomes depends on the nature of the schooling system is important but does not seem to have been much followed up. Part of the parental investment is picking a good school, and if for some reason only ineffective schools were available, it is not clear that any parental investment would be able to make up for that. As the rest of this survey shows (through randomly allocated schooling provision in many cases), there are considerable differences between the school effectiveness and teacher effectiveness that children experience, and this needs to be better integrated and understood in the framework for skills development.

One recent example that sheds new light on ‘home’ versus ‘school’ is Curto and Fryer’s (2014) investigation of urban boarding schools for the poor. SEED schools combine a “No Excuses” charter school with a five-day-a-week boarding programme. Since attendance in these schools exceeds capacity, a lottery is used to determine which of the applicants gets in; this allows estimation of a causal effect among applicants. They find very substantial effects, of 21% of a standard deviation in reading and 23% in maths. The latter is similar to those found in other outstanding charter schools (see section 4.b.ii below), but the reading impact is much higher. The fact that the impact of attending a charter school on maths does not depend on whether it is combined with a more supportive out-of-class environment or not, is suggestive of the importance of high school effectiveness. The differential impact on language scores is as interesting in reflecting that so much language learning happens outside class, and so a change in environment can make more of a difference. As Curto and Fryer note, boarding schools are very expensive and they postpone calculation of a rate of return until the first cohorts going through the system are older.

We know very little in terms of robust causal evidence about the role of schools in producing non-cognitive skills. One important direction for future research is a much greater focus on this, and on understanding better the interaction of school and home.

### **iii. Family background and the development of human capital**

The starting point is a simple description of the degree to which a child's educational attainment is influenced by their family background. Björklund and Salvanes (2011) review the techniques used to assess this, the evidence and the important caveats that come with it. Perhaps the most common approach is looking at the correlation of educational (or other) outcomes between siblings. This separates the factors that siblings share from their idiosyncratic characteristics. It is important to be clear that the shared things will include genes and upbringing, but also probably the same schooling and the same neighbourhood. These results typically suggest that between 40% and 60% of variation in attainment comes from something in the family background shared by siblings. Björklund and Salvanes (2011) provide a taxonomy of what these things might be. They also discuss a popular decomposition of attainment into genetic factors, shared environmental factors and idiosyncratic factors, "nature" versus "nurture", but note that the assumptions of a linear additive form and zero correlation between genes and environment raise strong doubts about the resulting estimates.

There has always been interest in the genetic basis for cognitive and non-cognitive abilities and 'intelligence', and in particular the degree of heritability of these. Typical findings are that varying but substantial fractions of the variance in measures of general intelligence can be explained by genetic factors, and researchers are beginning to identify individual genetic variants which associate with education. Branigan et al (2013) offer a meta-analysis of the heritability of attainment, making the point that the evidence shows heritability to be environmentally contingent, and Rimfield et al (2015) show that even performance in specific subject exams at the end of compulsory schooling is "substantially" heritable, even after controlling for general intelligence. However, Heckman and Mosso (2014) caution against laying too much stress on genetics and report studies arguing that the degree of heritability varies by socioeconomic status and more broadly that the expression of genes is dependent on the environment. Björklund et al (2006) using data on adopted children in Sweden show evidence of interactions between post-birth factors (environment) and pre-birth factors (genes).

There is now a substantial evidence base showing that early childhood environments have a strong impact on educational attainment and adult outcomes (Currie and Almond, 2011). Early childhood events include pre-natal events: Currie and Hyson (1999) show that low birth weight affects test scores at age 16 and many subsequent papers reinforce and extend that result. They also review the evidence on post-natal but early childhood shocks, typically health shocks, and argue that these are less powerful than foetal exposure, but nonetheless, mental health conditions can have large and persistent effects.

Gaps between children in cognitive skills are apparent very early in life and are strongly correlated with family background. One of the most studied aspects of family background is parental education, and the impact that this has on the child. Different methods have been used to move beyond correlations to causation including comparing the children of twin mothers, and the use of policy change that yields an exogenous change in parental education (see the methodological study in

Holmlund et al, 2015). Studies find positive causal effects of mother's education (Chevalier, 2004; Carneiro et al, 2013) or father's education (Black, Devereux, and Salvanes, 2005) or both (Oreopoulos, Page, and Stevens, 2006; Maurin and McNally, 2008). Carneiro et al (2013) show substantial intergenerational returns to education in the NLSY, with an additional year of mother's schooling raising attainment and reducing the child's behaviour problems. They argue that this is likely to work through income effects, delayed childbearing, and assortative mating. However, the intergenerational correlation of education captures only part of what siblings inherit from their parents, perhaps about a third. Other factors not correlated with parental attainment might include non-cognitive attributes, ambition and sources of income not derived from earnings. Björklund and Salvanes (2011) conclude that at most half of the intergenerational correlation of education can be considered causal, which reduces further the component of family background that has a causal effect on the next generation. Blau and Currie (2006) make the point that large fractions of children in developed countries spend many hours at young ages with people other than their parents, so the role of pre-school time in generating childhood skill development are also important.

#### **iv. Early childhood interventions**

There have been very many public programmes designed to alter children's developmental trajectories, many of them focused on disadvantaged communities. Some have been introduced in a way that is helpful to causal analysis; many have not. Blau and Currie (2006) and Currie and Almond (2011) provide extensive reviews of programme characteristics and evaluation evidence.

One of the key open questions is the interaction between public and private investment, whether they are substitutes or complements. There are contradictory findings in different contexts, and more research is needed to isolate the factors pushing in different directions. Gelber and Isen's (2013) results support complementarity: they showed that children randomly assigned to receive additional child care also saw a substantial increase in their parents' involvement (reading to the child, and just spending time) both during and after the child care intervention. On the other hand, Fredriksson et al (2013) find substitution: responding to bigger class-sizes (ie lower public investment) high-income parents provided more support to their children, and low-income families did not respond at all.

One obvious intervention is simply to provide poor families with more money. Given identical preferences and technologies, poor families will choose lower levels of investment in their children (and, as Currie and Almond point out, lower levels of consumption). In addition, they may have access to different production technologies (for example, available child care). Berger et al (2009) show that measures of the home environment and parenting skills are strongly correlated with income and that this accounts for a lot of the income-outcome relationship. Overall, Currie and Almond (2011) argue that the evidence about the effects of income is positive. Two further exacerbating issues for poorer families are first, that they tend to be subject to more or larger early-life shocks (Currie and Stabile, 2003), and second that any shock will have a greater impact as they have lower initial skills stock, that is, they are on a steeper portion of the production function.

An alternative to providing more money is to provide more time, for example with parental leave reforms. Carneiro et al (2015) use a reform in maternity leave in Norway to identify the impact on later outcomes of mothers' time with the child. Children of mothers eligible for more time had lower school dropout and higher wages at age 30; the effect was stronger for children of mothers who

would have taken very little leave before the reform. This makes clear the link between human capital development and family labour supply choices. Del Boca et al (2014) set out and estimate such a model using the PSID: combining standard family labour supply decisions with time and money inputs into the child development function. They find that both maternal and paternal time are crucial inputs in production, about equally productive, and that cash transfers to the household have only minor impacts on child outcomes.

A second approach is direct intervention in the lives of young disadvantaged children. Home visits, and specifically Nurse-Family Partnerships, are seen as promising both in the evidence that Currie and Almond (2011) review and also by Heckman and Mosso (2014). As Currie and Almond argue, “Home visiting programs can be viewed as a type of parenting program” (p. 1425) and given the importance of parenting styles, it makes sense that improving parent knowledge should be successful. However, they also argue “it is remarkably difficult to change parents’ behavior and that many attempted interventions are unsuccessful. The most successful parenting programs are those that combine parent education with some other intervention that parents want, such as visits by nurses [Olds et al, 2007]”. (p. 1425/6). Fryer et al (2015) report mixed results from a field experiment in Chicago incentivising parents to engage in positive parenting behaviours through attending a parent academy.

A third type of intervention is child care programmes. The two most cited are the Abecedarian Project and the Perry Preschool Project. They are rightly famous as they adopted an experimental approach: children were randomly assigned to treatment and control group, they had relatively low attrition, and continued to track the children’s outcomes over many years. Both found positive effects on schooling, bigger and more persistent effects on non-cognitive than cognitive skills, and have been shown to generate higher outcomes well into adulthood. The evidence also suggests stronger effects for children from disadvantaged backgrounds.

However, these programmes illustrate a problem in projecting from the strong success of small-scale intensive programmes to large-scale “business as usual” projects. The obvious comparison for the very small-scale Perry Preschool and Abecedarian projects in the US is Head Start, a long-running (since 1965) and very large-scale (about 800k children) preschool programme for disadvantaged pre-school children. Head Start is not of the same quality as the small-scale intensive projects, although often of higher quality than disadvantaged neighbourhood alternatives. Heckman Pinto and Savelyev (2013) note that about 30 percent of all Head Start centres offer a version of the Perry curriculum.

The results from Head Start showed an initial improvement in test scores that tended to ‘fade out’ for black children, but not for whites. Currie and Thomas (2000) showed that this is probably because the black children leaving Head Start went on to less effective schools than the white children graduating from the programme. This fits in well with the argument of Heckman and Mosso (2014) that while early investment is important, it only pays off if it is supported by on-going investment. Additionally, the finding of ‘fade out’ may have been a consequence of relying only on cognitive measures to measure impact.

In policy circles there seems to be a very strong distinction between ‘early years’ investment and ‘late years’ investment, with a view that the former is very valuable and the latter pointless. This view is undoubtedly stronger and less nuanced than would be held by the researchers advocating early investment, who would argue that the issue is more a question of emphasis and a re-balancing

of existing programmes than a wholesale shift. Also, as noted, it seems that early investment has to be supported by later investment for it to pay off.

We should be cautious in comparing the results of small-scale, very intensive ‘early years’ projects with the likely outcomes of national-scale, everyday ‘early years’ programmes. It is also worth remembering that there have been very successful ‘later years’ interventions both in relatively small-scale experimental settings (for example, Fryer, 2014, introducing charter school practices into regular public schools) and at a national-scale (such as in some cases of raising the school leaving age, Meghir and Palme, 2005). Finally, Carlsson et al (2015) identify a causal effect of extra schooling on cognitive skills for 18 year olds. They show that an extra ten days of schooling raises scores on crystallized intelligence tests by 1% of a standard deviation, while days not spent in school have no effect (so it is not just age). The effects are the opposite for scores on fluid intelligence: they do not increase with more days of schooling but do increase slightly with age.

## **b) School effectiveness – resources, class size, teachers and pupils**

### **i. School resources**

Does money matter for schools? Money matters for the provision of most things. In one sense, it does, since teachers’ jobs and careers depend on school funding, and tight budgets make running a school a lot harder. But it is not so clear that money matters for pupil attainment. In fact, while there is evidence on both sides, possibly the majority of researchers in this field would agree that increases in a school’s resources are unlikely to have a major effect on attainment. This is in large part the debate about the number of teachers a school has relative to its pupils (class size) but it is also broader than that. The empirical work to date has not produced uniform results and this is not yet a settled question. Some reasons for this are given below, principally about the difficulty in establishing a robustly causal link from resources to attainment.

#### *International perspective*

At a macro level, a simple comparison of attainment and school expenditure is revealing. The internationally comparable attainment data in PISA is one obvious route to doing this. This exercise shows clearly that among developed economies (PISA suggests a threshold of about \$20k GDP per capita), there is no relationship at all between the two; see for example, the Figures in OECD (2012) which use average reading performance and average spending per student from the age of 6 to 15. Summarising their findings, the OECD states “The bottom line: Money alone can’t buy a good education system. ... When it comes to money and education, the question isn’t “how much?” but rather “for what?”” To take one national example, over the period 2000 – 2010 real per-pupil spending rose by 68% in England, with nothing like a commensurate rise in attainment.

#### *Micro studies*

The leading proponent of the idea that resources do not matter for attainment is Hanushek, see for example Hanushek (2003, 2006). For example in table 3 of his 2006 survey, Hanushek categorises 376 studies in the US linking measures of resources with attainment and shows that the majority of studies for all measures are insignificant. This finding is stronger in models of value-added. Despite highlighting some concerns about the statistical quality of the studies (with worries about the nature

of the funding regime and missing variables in the analysis), his view is that the overall finding is valid.

The main reason why it is difficult to determine the causal effect of school expenditure on pupil attainment is that there is a second mechanism at work. Funding is certainly not distributed randomly to schools, and many countries allocate more funding to schools in more disadvantaged neighbourhoods. Since these neighbourhoods are almost certainly going to generate lower attainment on average, this funding mechanism sets up a negative relationship between funding and attainment that works against any positive relationship. Finding a way to deal with this – to essentially generate a situation in which it is as if the funding were distributed at random – is difficult and such studies are consequently rare. This partly lies behind Krueger's (2003) counterpoint that the evidence at his time of writing was not really capable of making a causal statement. A different and more aggregated approach, pioneered by Card and Krueger (1992) related later life earnings to measures of state expenditure on schools, and found that some aspects of school characteristics were associated with higher returns.

In a useful review of more recent studies, Gibbons and McNally (2013) argue that the bulk of the evidence now suggests a more nuanced picture than Hanushek's. They note a few recent studies showing significant impacts of expenditure using natural exogenous variation in expenditure to identify the effect. Their survey also argues that when effects are detected for resources, they tend to be larger for pupils from poorer communities. This includes their own work (Gibbons and McNally, 2011) and Holmlund et al (2010). Two other recent papers show that schools can achieve substantial attainment gains from more money; the common theme of these two studies is that schools are free to use the money as they wish. de Haan (2015) shows that additional funds for secondary school for low-ability pupils significantly increases the probability that pupils pass the exit examination. Jackson et al (2014) show that school finance reforms in the US in the 1970s and 1980s raised overall educational spending in disadvantaged neighbourhoods and leads to a substantial and significant effect on years of schooling and earnings. Leuven and Oosterbeek (2007) consider a number of policies focussed on disadvantaged schools in the Netherlands giving extra resources to schools for personnel and for IT equipment, and also explicitly to reduce class sizes. They review evaluations of these policies that were able to use quasi-experimental methods and so mimicked random assignment. They find no effect on attainment for any of them, and are able to rule out any substantial positive impact.

The main issue in the discussion over school resources is about class size, though there are other factors, noted below.

#### *Impact of class size on pupil attainment*

Researchers have adopted different approaches to identify a causal effect of class size on attainment and on subsequent outcomes, using natural variation, administrative rules and experiments. Rockoff (2009) describes experiments from the early twentieth century to determine the impact of class size on attainment – this is not a new research topic.

The first approach follows a school through time and exploits the idea that exogenous changes in demographics will lead to small changes in class size (Hoxby, 2000). Using this (and maximum class sizes), she found no effect of class size, and ruled out even modest effects (2-4% of a standard deviation) for a 10% fall in class size.

Angrist and Lavy (1999) used administrative limits on class size in their study. The ‘Maimonides rules’ are strict limits on class size, meaning that a small change in total school enrolment can mean very different class sizes (for example, given a maximum size of 30 two schools with 58 and 63 pupils respectively will have class sizes of 29 and 21). They find significant and sizeable effect sizes of around 17% of a standard deviation for a class size reduction of 8 pupils (as achieved in STAR, see below). Gibbons and McNally (2013) cite Piketty (2004) exploiting a similar rule in France and also finding sizeable effects. Gary-Bobo and Mahjoub (2013) use both the ‘Maimonides rules’ and the Hoxby approach and find significant but small effects of class size.

Perhaps the most famous analysis in this topic is the Tennessee STAR class size reduction experiment. This involved 11,000 pupils (and their teachers) being randomly assigned to different class sizes from kindergarten to third grade from 1985. Small classes (13–17 pupils) were compared to regular classes (22–25 pupils), and also to regular classes with a teacher’s assistant. While the random nature of the experiment has been questioned, Kreuger’s (1999) analysis deals with non-random attrition. He shows significant and sizeable test score gains in the first year that pupils are in the smaller classes, and smaller but non-zero subsequent gains. Again, effects are larger for poorer students.

More recently, Chetty et al (2011) have published a long-term follow up on Project STAR. The pupils in the original experiment are tracked into their late 20s using administrative data, to yield a number of findings. While class size impacts on earnings at age 27 were insignificant (though imprecisely measured), pupils in small classes were more likely to attend college; they also had positive impacts on other outcomes. Fredriksson et al (2013) also study long-term effects of smaller class sizes following a Swedish reform. They show that children in smaller classes at the end of primary school complete more years of education, and have higher earnings at ages 27 to 42.

More research on the impact of class size would be useful. But also useful would be a recognition that class size differences may not be simple and symmetric. Schools with more autonomy can be more imaginative with class sizes. For example, a school with three classes of (20, 20, 20) might make an asymmetric change to (40, 5, 5) rather than a symmetric change to (15, 15, 15, 15). This gets closer to the idea of schools themselves creating a way to engage in high-intensity tutoring for students who would benefit from this.

#### *Other school resources*

Schools can spend their money on factors other than just teacher quantity. One obvious alternative is paying more to attract more effective teachers. Higher teacher quality is very effective in raising attainment (see section 3.b.ii), but it is less clear yet how important salary is in selecting such teachers. Another related factor is spending more money on hiring teaching assistants, classroom aides. There is less robust evidence on this topic, although Kreuger (1999) shows in the Project STAR study that the experimental treatment of regular class plus full-time teacher aide showed no impact on test scores. Blatchford et al (2009) in a non-causal study also found no effect of teaching assistants on attainment.

Information technology in schools is often discussed in policy circles as a great investment but in fact there is very little strong evidence of a positive impact on attainment (though see Machin et al 2007). Leuven et al (2007) actually estimate a causal negative return to IT funding for primary schools. Faben et al (2015) use spatial data to identify as good-as-random variation in internet

speeds and show that this has no significant effect on the time students spend studying or on their performance.

Neilson and Zimmerman (2014) consider capital expenditure – the implementation of large school re-building project in the US, and find positive effects of construction expenditure on attainment. It is not clear how this might come about, but the authors speculate about effects on pedagogy, and on pupil and teacher motivation.

### *School responses to budgets*

One of the interesting and relatively unexplored questions in this field is how schools make their financial decisions. If spending does only have low or no effect on attainment, why do schools spend their money in this way? Is it bureaucratic or legal or union constraints? Is it lack of information? Allen et al (2012) explore schools' financial decisions and show that in a period of budget increases there was no commonality of action among schools on how to spend the increase. It may also be that part of the reason for lack of consistent findings on resources and class size is that schools may adopt different coping or implementation strategies when faced with big or small changes (Gibbons and McNally, 2013).

## **ii. Teachers**

Research on the economics of teachers and teaching has grown dramatically in recent years. This follows in part from a realisation that this is an area of great policy promise for raising attainment. This means that there is a lot of new research to cover, but also means that there a number of very helpful recent reviews, that I will draw on here including Jackson et al. (2014) and Koedel et al (2015).

Economists define teacher effectiveness (or teacher “quality”) precisely but narrowly. It is based on the progress in academic achievement that a pupil makes over her time with the teacher, typically measured by standardised tests at the end (and ideally at the beginning too). A teacher’s effectiveness is the average of this pupil progress measure across all the pupils she teaches. The early studies were interested primarily in the overall distribution of teacher effectiveness, but later work has become focussed on the effectiveness of individual teachers in the context of performance management and reward. More recent work has also broadened the sense of teacher effects to include their impact on non-cognitive outcomes as well as test scores.

### *Teacher effects*

While economists’ studies of teacher effectiveness started some while ago (eg Hanushek, 1971), recent analysis has been hugely helped by the use of administrative data. The key to any measure of teacher effectiveness is a class list – which pupils were taught by which teacher – and this is typically only available in administrative data. The papers that provided the foundation for the subsequent work were Rockoff (2004), Rivkin, Hanushek and Kain (2005) and Aaronson, Barrow and Sander (2007); the early work is surveyed in Hanushek and Rivkin (2010). These first few studies are worth considering in some detail as much subsequent work uses similar techniques and produces similar findings.



Rockoff (2004) estimates teacher effectiveness using data from two school districts in New Jersey over the years 1989–90 to 2000–01 covering grades 2–6. The data allow individual teachers to be matched with their pupils for each year of the study. Rockoff finds that a one standard deviation increase in teacher quality results in a 0.11 standard deviation increase in reading and writing test results. Teacher experience is found to have a significant positive effect on maths and reading exam results, but no other observable teacher characteristics are found to have significant effects. The data used by Rivkin et al. (2005) also covers early grades, half a million students across more than 3,000 schools in Texas. Their data does not have class lists and so cannot match individual students to individual teachers, only to a set of teachers in a grade within a school, which is likely to attenuate estimated teacher effects. Their lower bound estimate implies a one standard deviation increase in teacher quality is associated with 0.11 and 0.095 standard deviation increases annual growth in achievement in maths and English respectively in grade 4. They find a significant negative effect of inexperience in maths teachers, and a smaller negative effect for English teachers. However the qualifications of teachers were found to have no significant effect. Aaronson et al. (2007) study ninth-grade maths scores in one school district in Chicago over a 3-year period. They can link students with the actual teacher that taught them, and they also have prior attainment data, which they assume absorbs student heterogeneity. They find that an increase in teacher quality of one standard deviation above the mean is associated with 0.15 standard deviation increase in the maths test score. Clotfelter et al (2007) and Slater et al (2012) are able to use point-in-time fixed effects to control very fully for observed and unobserved student ability. Clotfelter et al. (2006, 2007), take a different approach and directly regress student outcomes on teacher characteristics.

These test score outcomes have been shown to be correlated with human capital growth and so influence outcomes of interest. Chetty et al (2014b) find that pupils taught by highly effective teachers earn more, are more likely to go to university and live in richer neighbourhoods. Similarly Jackson (2013) shows that teacher effectiveness predicts high school dropout rates and college plans.

Note that research often reports within-school differences in teacher effectiveness (Aaronson et al, 2007, look at teacher mobility but have few transfers for identification). This is because overall school effects are confounded with school mean teacher effectiveness and a measure of total variation in teacher effectiveness could only be derived from tracking a large population of teachers moving frequently between a fixed set of schools. This is only very recently becoming available (Jackson, 2013, and Mansfield, 2015); Mansfield uses administrative data from North Carolina yielding 3500 teacher transfers; see further discussion below. Total variation in effectiveness, including both within- and between-schools, is greater than within-school but further research is required to determine by how much.

A lot of studies have followed these early breakthroughs. In his 2011 survey, Hanushek notes “Literally hundreds of research studies have focused on the importance of teachers for student achievement” (p. 467). The typical result is remarkably consistent: a one standard deviation in teacher effectiveness yields a 10% - 20% standard deviation change in pupil attainment. This appears to be true across different stages of school, different subjects (though typically greater in Maths) and (to the extent we have evidence) across different countries. An alternative way of expressing the difference is in terms of years of achievement gain, the difference equivalent to “some teachers

producing 1.5 years of gain in achievement in an academic year while others with equivalent students produce only 1/2 year of gain” (Hanushek 2011, p. 467).

As the interest in non-cognitive attributes has risen, researchers have also considered whether teacher effectiveness influences these. Jackson (2013) shows that teacher impacts on non-test behaviours such as absences and grade progression also predict later educational outcomes. Blazar and Kraft (2015) also estimate teacher effects on non-cognitive outcomes. They find effects of the same size as teacher effects on attainment, and note that effects are only weakly correlated between cognitive and non-cognitive skills.

### *Can we trust the standard estimates?*

The main threat to the validity of these estimates is that there is unmeasured variation in mean pupil characteristics between teachers. Suppose that for some reason a particular teacher was assigned a group of ambitious, highly motivated pupils; if that is not measured in the dataset, then the higher test scores achieved by those pupils will be incorrectly assigned to the teacher’s effectiveness. This is the line taken by Rothstein (2009, 2010) who argues that there is strong non-random sorting within schools. He makes the case by showing that pupils’ future teacher assignments are conditionally correlated with their current performance, suggesting such sorting is important.

This has generated a number of responses, some statistical and some quasi-experimental. Chetty et al (2014a) and Koedel and Betts (2011) show that persistence in class membership, correlated measurement error and mean reversion can account for Rothstein’s finding. Goldhaber and Chaplin (2012) use simulation methods to show that the falsification test performs poorly in small samples. Experimental evidence from Kane and Staiger (2008) and Kane et al (2013) reinforces the view that conventional estimates do not suffer from strong bias. They estimated teacher effectiveness from past classes and then randomly assigned high and low effectiveness teachers to new classes. The historical estimates predicted student progress well. This suggests that when students’ prior ability is sufficiently accounted for, non-randomized studies can recover similar estimates of the variability of teacher effectiveness to random assignment. Chetty et al (2014a) also implement a quasi-experimental test for bias, comparing cross-cohort variation in estimated teacher effects and cross-cohort variation in attainment; they find no evidence for bias. This debate is on-going: Rothstein (2014) questions the validity of the test, but Chetty et al (2015) confirm its validity. Jackson (2014) does however suggest that accounting for pupil tracking reduces estimates of persistent teacher effectiveness.

These results give us confidence that estimates of the distribution of teacher effectiveness are plausible and not simply driven by unobserved classroom sorting. Of course, this does not mean that all such estimates are equally free of bias: other contexts with different teacher assignment rules and different datasets will produce different outcomes. Controlling for pupil characteristics through prior achievement and/or point-in-time pupil fixed effects is the best method for ensuring unbiased estimates.

### *Correlates of teacher effectiveness*

The early research found few teacher characteristics that were correlated with their effectiveness (see Gordon, Kane and Staiger, 2006). This matters because a strong predictor of effectiveness would be very helpful for schools in decisions to hire and to retain teachers. Rockoff (2004) and

Rivkin et al (2005) and other researchers found nothing but teacher experience to be significant correlates of effectiveness.

Arguably the most important characteristics for policy and school practice are whether teacher effectiveness is correlated with teaching experience, the teacher's own educational achievement, and with teacher training or certification, their route into teaching. I deal with these in turn. Other studies (Dobbie, 2011; Rockoff, et al 2011; Hill et al, 2015) have considered other teacher characteristics, not typically recorded in research datasets but possibly available to panels hiring teachers; these are discussed below under policies on teacher recruitment.

As Papay and Kraft (2015) note, the relationship between effectiveness and experience has been considered settled: effectiveness was markedly lower in the first few years of work, improved substantially over the first 3 years but thereafter was essentially constant. Recently, however, Wiswall (2013) and Papay and Kraft (2015) show teacher effectiveness increasing much later into the job. The argument is a simple and plausible one. The research to date has almost universally used cross-sectional data, comparing the distribution of effectiveness among groups of teachers at different levels of experience; this has been repeatedly shown to be flat. This ignores differential attrition; this flat profile could derive from different combinations of selective attrition from teaching (or at least from the dataset) and within-teacher improvement in effectiveness. Wiswall (2013) and Papay and Kraft (2015) both find continuing increasing returns to experience within-teacher offset by negative attrition as the higher quality teachers are more likely to leave. This implies a "decreasing average innate quality for those teachers who stay in teaching" (Wiswall, 2013, p. 5). Both of these aspects are of great policy interest: the increasing effectiveness matters for teacher retention decisions (below), and the fact of a higher loss rate for more talented teachers is relevant for pay decisions among other factors. Of course, it remains true in a cross-section that experience *per se* does not predict teacher effectiveness.

The improvement in effectiveness is substantial; Wiswall finds it to be 0.75 of a standard deviation higher for a teacher of 30 years' experience than for one of 5 years'. Intriguingly, it is restricted to maths – neither study finds any improvement for English. Papay and Kraft conclude that the debate on teacher experience is now "at least not settled". This analysis links to research on teacher careers that I discuss below.

The second key characteristic of interest is the teacher's own academic background. Here there seems to be consensus that this is largely uncorrelated with effectiveness. Ladd and Sorensen (2015) note that most of the papers in this field find no effect of the teacher herself holding a Master's degree; Hanushek (2011) concurs: "Most documented has been the finding that master's degrees bear no consistent relationship with student achievement" (p. 467). Ladd and Sorensen's own work re-affirms this relationship for high school teachers, accounting for the selection bias on which teachers get a Master's degree and including teacher fixed effects.

Third, the link between teacher-training or the 'route' into teaching is clearly important for policy: do some methods of selecting and training teachers produce more effective teachers? This is increasingly important as recent times have seen an increase in the availability of new pathways into becoming a teacher (see Grossman and Loeb, 2010, for the US case, and Boyd et al, 2008, specifically for the wide variety of teacher training in New York City). Research to date has found mixed evidence on effectiveness from different teacher certification programmes. Kane et al (2008) show

that traditionally certified teachers, Teach for America (TFA) teachers and uncertified teachers have the same distribution of teacher effectiveness. However, Boyd et al (2011) are more positive about the role of differences teacher certification, noting that the influx of more able TFA teachers substantially improved student attainment in middle and high school. Boyd et al (2012) study a new certification path introduced in New York City, 'Math Immersion' teachers, and find that these teachers yield smaller pupil gains than regular teachers and substantially smaller gains than TFA teachers. Goldhaber and Liddle (2011) find no effect on student attainment from teachers trained in one state (Washington) and outside. Harris and Sass (2011) find no consistent relationship between formal professional development training and teacher productivity.

This is surely a key area for further research: whether teacher training programmes can be evaluated and modified to raise teacher effectiveness. There are obviously a number of complexities, not least the issue, largely unconsidered in research so far of the endogenous selection of trainees into programmes and then the endogenous recruitment from particular programmes into particular schools. Mihaly et al (2013) raise other econometric issues.

### *Teacher labour markets*

The teacher labour market is of course connected to the wider professional labour market; nevertheless it makes sense to talk about it as a separate entity as it relates to one occupation, one type of employer and a (somewhat) controlled mode of entry. The individual decisions of teachers, would-be teachers, and schools aggregate up to hiring rates and separation rates, job tenure and time in occupation, and flows into and out of the profession. It is the latter that has attracted more attention in the past under the general heading of 'teacher supply' (for example, the survey in Dolton, 2006) but the recent research on models of teacher hiring and selection, and performance-based retention has focussed more on school- and individual-level decisions. This research is making use of the administrative datasets now available. For example, Boyd et al (2013) study the teacher-school matching process and aim to disentangle their separate preferences in forming that match. Allen et al (2012) characterise teacher tenure and turnover using a new census administrative dataset in England. They show that schools in disadvantaged neighbourhoods do have slightly higher turnover but that this is largely explained by the fact that they hire younger teachers; that in turn derives either from the preferences of young teachers or the market opportunities of schools in such neighbourhoods. An alternative, longitudinal, aggregation is to characterise typical and atypical teacher careers, but this does not appear to have been studied much yet in the economics literature.

Does teacher turnover matter? Ronfeldt et al (2011) use school-by-grade level teacher turnover and investigate the impact on the test scores of 850,000 students in New York. They find that students in grade levels with higher turnover score lower in both maths and English, with larger effects for low-performing students. Teacher quits can be prompted by a difficult teaching environment (Barbieri et al, 2011, using Italian school transfer application data; and Falch and Strøm, 2005 using Norwegian panel data), or by a school with low performing students (Falch and Rønning, 2007, using the same data; Hanushek et al, 2004 using Texas data; Goldhaber et al, 2011). Goldhaber et al (2011) also suggest that more effective teachers are more likely to stay in their initial schools, though whether that is because they initially place better, or some other mechanism is unclear. A group of recent studies agree that higher pay in a specific school reduces teacher quits from that school: Clotfelter et al (2006), Falch (2011) and Hendricks (2014). This might provide a policy lever for one particular

school, but it is not clear what the general equilibrium effects might be if a lot of quits are school-to-school.

### *Policies to improve average teacher effectiveness*

Given how powerful differences in teacher effectiveness have been shown to be, a clear goal for education policy is to find ways to raise it. This topic has generated a lot of research, but a clear and consistent policy tool has so far proved more elusive than might have been hoped.

The key question is: are effective teachers born or made? That is, is the ability innate, not really teachable ('born'), or are there reliable ways to improve effectiveness of current and nascent teachers ('made')? Undoubtedly, the answer will be somewhere between the two extremes but the two extremes illustrate the two main tracks that the research has taken. If effective teachers are 'born' then the issues are about selection, identifying that ability, hiring and retention; if effective teachers are 'made' then we should focus on the best training, mentoring and feedback mechanisms. There is a third dimension wherever the ability comes from, namely teacher effort; here the policy interest is about enhancing that effort typically with performance pay. I consider the evidence on these three policy areas.

### *Teacher selection: identification, hiring and retention*

The first step is to find a reliable measure of teacher effectiveness. Whilst value-added is the standard choice, there have been concerns about its stability over time and across different types of class, grade and pupil. However, Jackson, Rockoff and Staiger (2014) report work showing that the correlation between just one year's teacher performance and performance over the rest of her career (which is what a retention decision is about) are high enough to be useful; for example "over three quarters of teachers at the 25th percentile of one-year value added have career value added that is below average" (p. 13).

The MET (Measures of Effective Teaching) project is an influential and large-scale programme that has explored different measures of effectiveness. Kane and Staiger (2012) discuss the collection of three different types of evidence on teacher effectiveness, based on the value-added measures discussed above, on classroom observations, and on seeking student views through surveys. In the final report, Kane et al (2013), they randomly assign teachers to classes, and then compare the test score gains as predicted by a composite measure of effectiveness with the actual scores. They found that the composite measure on historical data did correctly identify the high-performing teachers and also got the magnitude of the differences right. This provides a good deal of validation for the measure, though Rothstein and Mathis (2013) offer a sceptical review. Mihaly et al. (2013) also favour a composite indicator, arguing that an equally weighted average of the indicators provide a better overall predictor of multiple outcomes. Relatedly, the work of Kane et al (2010) matches detailed classroom observation data with measures of pupil progress and shows that one can relate high pupil progress to observable practices of teachers.

Clearly, these measures can only be collected on people who teach. They are not available for hiring decisions, and this is one of the central problems in teacher selection: how should you hire teachers when you have little information to predict whether they will be good at the job? Angrist and Guryan (2008) show that a policy of requiring teachers to pass a standardised test does not raise mean teacher effectiveness, though it is associated with increases in wages. Other research has looked at other more commonly available indicators that would be available at interview. For

example, Rockoff et al (2011) and Dobbie (2011) look at a variety of widely used measures; they find that a composite of these can predict a (small) fraction of teacher effectiveness. Hill et al (2015) solicit other measures including subject-knowledge, self-reported effort, and mindset variables, and find 'mild' correlations with teacher effectiveness. Goldhaber et al (2014) and Jacob et al (2015) report some promising results from other pre-hire screens.

If informed teacher hiring is still an open issue, teacher retention or layoff has been addressed. Here obviously the performance measures can come into play. An influential and controversial contribution comes from Staiger and Rockoff (2012) who use the known facts about teacher effectiveness to simulate an optimal teacher retention policy. The policy is based on value-added information about a teacher's performance and the core tradeoff is essentially between laying off low value-added teachers and therefore needing to hire inexperienced teachers with the effectiveness penalty noted above. The result is very strong: given the widely estimated parameter values, schools should retain less than half of teachers after their first year of teaching. The fact that differences in effectiveness are large, persistent, and unknowable before hiring, means that even on the basis of one year's measure, the one-time cost of replacement is worth it to avoid a career-long tenure of an ineffective teacher. Rockoff et al (2012) show that simply providing school Headteachers with performance information leads to some Headteachers changing their views on their teachers and increasing turnover of low-performing teachers. Rothstein (2015) notes that the Rockoff and Staiger model assumes a perfectly elastic supply of teachers and adapts the model to allow for wage rises that would accompany such a retention policy. He finds that optimal layoff rates are lower, but still high and higher than current practice, and there is still a substantial gain to average value-added and in dollar terms.

There have been a few studies of actual layoff programmes and their relationship to teacher effectiveness. Seniority is typically the main criterion of who gets laid off, and Boyd et al (2011) show that the difference in student attainment outcome between this rule and an effectiveness-based rule was large and persistent. Kraft (2015) examines the impact of discretionary teacher layoffs in North Carolina Schools in 2009 and 2010, noting that while seniority was not the only criterion used, the decisions were not based on effectiveness, and that the loss of effective teachers was detectable in school performance measures. Dee and Wyckoff (2013) examine the results of the controversial IMPACT system of teacher evaluation in Washington DC, which entailed unusually high and credible threats of teacher dismissal for poor performance. They show that the strong dismissal threat increased quits by low-performing teachers by 50% and raised performance among low performers remaining by 27% of a (teacher-level) standard deviation. The regression discontinuity design means that these are estimates very close to the margin for dismissal, rather than the average treatment effect.

Given the variability, persistence and unpredictability of teacher effectiveness, the selection of teachers is a very important policy issue. This has been vividly illustrated by Hanushek (2011), and Chetty et al (2014b) among others. Replacing the lowest performing 5-10% of teachers with average teachers produces extremely large net present value calculations. Each effective teacher raises the lifetime earnings of very many pupils over her career; Hanushek (2011) suggests that using consensus parameters "replacing the bottom 5-8 percent of teachers with average teachers [would have] ... a present value of \$100 trillion". Chetty et al (2014b) similarly find that replacing the 5% least effective teachers with average teachers would yield around \$9000 per classroom per year.

### *Teacher improvement: training, mentoring and feedback*

The other route is to improve the effectiveness of existing teachers. The new results on experience noted above suggest that learning is possible, albeit that that refers to personal learning through own experiences, not ideas communicated by someone else. The challenge is to find effective ways to pass on that learning to new teachers and throughout careers.

There are several layers to improvement. First, there is the formal initial training, preparation or certification to become teacher. The work on this to date, discussed above, does not suggest any dramatic differences in effectiveness from different training routes although there are worries about the importance of the endogeneities in route selection, school selection and trainee selection.

Once a teacher is in post there are a number of potential channels for improving effectiveness, reviewed below by increasing degree of formality: informal learning, peer mentoring, peer coaching, and evaluation.

First, the most informal is simply learning on-the-job from other teachers around. Jackson, Rockoff and Staiger (2014) cite research showing that working with effective peer groups raises a teacher's own effectiveness. Jackson and Bruegmann (2009) use observational data with both within-school and within-teacher variation and find that a teacher's students have larger achievement gains in maths and reading when she has more effective colleagues. Papay et al (2015) run a field experiment and show significant improvements in teacher productivity when effective teachers work with a low-performing colleague in their school to improve that colleague's teaching skills. So for a school trying to assemble a group of effective teachers, there are in principle positive side-effects on mutual learning. From a school-wide perspective, Kraft and Papay (2014) show that the school context matters: teachers working in more supportive environments improve their effectiveness more than teachers working in less supportive schools. The nature of the professional environment was measured using state surveys on working conditions. They use teacher fixed effects and interact the professional environment with teacher experience. The non-experimental assignment of teachers to schools means that a causal interpretation is not secure.

Second, formal programmes of peer mentoring do not appear to generate strong or substantial impacts, see Rockoff (2008) and Glazerman et al (2010). Slightly different and more promising is Kraft and Blazar's (2014) study of personalised teacher coaching on classroom management techniques using the MATCH Academy approach. They find substantial, persistent and significant impacts on teacher practices, but as yet no evaluation of the impact on student attainment is available.

Third, teacher evaluation programmes can work to raise individual teacher effectiveness, as well as their role as a selection device noted above. As Taylor and Tyler (2012) note, this can work through two different channels: "by improving teacher skill, effort, or both in ways that persist long-run." (p. 3628). There are three studies suggesting that this is a promising approach. Allen et al (2011) use an RCT set-up to analyse a system of interventions focussed on ongoing, daily interactions of teachers with their students. This is measured through a particular framework or rubric, which makes it closer to the other evaluation studies discussed below than a simple professional development course. The intervention included initial workshop-based training, access to an annotated video library, and a year of personalized coaching followed by a brief booster workshop. This showed substantial impacts on student attainment, over 10% of a standard deviation, in the year after the intervention.

Taylor and Tyler (2012) also study the impact of an evaluation programme in Cincinnati on teacher effectiveness, using the Danielson framework. The analysis was able to track teachers and their pupils for some years after the evaluation. They show that teachers are more effective to a substantial and enduring degree, and that the biggest gains were in initially low-performing teachers. The findings are more impressive in that they relate to mid-career teachers whom might be expected to have hard-to-shift capabilities. Teachers were assigned to evaluation quasi-randomly and the intervention was a year-long classroom observation-based program. They find increases in test scores of 11% of a standard deviation in the years following the evaluation, suggesting that it largely derives from improvements in effectiveness and classroom skills than a one-off increase in effort due to the presence of the evaluators.

Steinberg and Sartain (2014, 2015) also analyse an evaluation intervention, comprising evaluation and discussion in Chicago, again based on the Danielson framework. They show that this raised school-average attainment during the evaluation period and persisting afterwards in the treatment schools. Unusually, a larger effect was found for English than maths. However, the findings also indicate the importance of support for the intervention since schools implementing the same program the following year with less support from the central office found no significant effects.

#### *Teacher effort: performance related pay*

Throughout the economy, performance pay is seen as a solution to the problem that a worker's effort is unobserved. The idea is that it aligns the incentives of the worker (the agent) towards the incentives of the owner (the principal). 'Effort' here need should not be taken too literally: even if the agent is working extremely hard, they may be spending a lot of time and energy on the wrong things (from the principal's perspective). So a performance pay scheme can be as much about increasing effort on the 'right' tasks as it is about simply increasing effort.

Performance-related pay (PRP) does two things: it raises individual productivity and it selects more productive workers. The former is clear; the latter is often neglected but can be very important in raising average productivity (Lazear, 2000). The idea is simple: faced with a choice between working in an organisation with PRP and one with common, constant pay, the more productive workers will choose the former and the less productive ones the latter. The basic model of optimal performance pay has been around a long time, setting out the trade-off between high-powered incentives and the consequential increased risk. In the canonical case the risk came from the weather affecting crop yields; for teachers the risk of associating their pay with test scores is the impact on those from events outside the classroom, and luck.

There are a number of extra issues that arise in performance pay for public sector organisations, see Dixit (2002) and Burgess and Ratto (2003) for surveys. These include multi-tasking and output measurement problems, the higher scope for implicit motivation, and the issue of multiple principals. All of these may mean that the optimal PRP rate for teachers is low. Enhancing implicit motivation is another approach; Teach First in the UK and Teach for America stress the social value of teaching in their recruitment drives.

Teachers' PRP is of most interest here when it includes a dependence on student attainment. Beyond that, there are very many design possibilities. PRP can be wholly objective (based on a given quantitative metric such as pupil progress), wholly subjective (based on a headteacher's appraisal) or a mixture. The last is most common as it incorporates the recognition that a teacher's job involves



multiple tasks some of which are very hard to measure. PRP can be on an individual teacher basis or a team basis, for example a Department, a year group or a whole school. Another key issue, considered in depth in Neal's (2012) survey is whether the reward should be for absolute performance (per fraction of pupils passing a test for example) or relative performance (being in the top half of the school's performance distribution). There are also arguments for and against threshold measures as opposed to continuous performance measures; a continuous measure provides an incentive throughout the performance range, but the threshold can focus it where it might be most needed (for example, all pupils gaining a basic competence).

The potential advantages of a PRP scheme are clear: increased 'effort' and better selection. Critics note a number of potential disadvantages. First, a riskier salary will in general require higher average pay to compensate. There are obvious measurement problems for many things teachers do, and it is not necessarily easy to disentangle the contribution of a single teacher to student progress. The system needs to be fair and accurate to be accepted. Some designs may also promote competition between teachers at the expense of cooperation, for example in lesson preparation. Another core focus is the likelihood of generating unintended consequences; see Neal (2012) again. Jacob (2005) studies the introduction of NCLB accountability in Chicago Public Schools and finds that teachers reacted strategically to the incentives, including substituting teaching time towards the tested subjects and away from others, and also by increasing special education placements. Teachers might focus narrowly on the rewarded measure, rather than a 'true' aim of education. This is often described as 'teaching to the test'. Of course, if the tests are well-designed then this might be appropriate. But often they are not, and the design of the outcome measure is key; Neal (2012) suggests that there should be one assessment system for PRP and a different one for measures of student achievement. While this is an important point, the PRP target needs to be an outcome that students and policy-makers care about as this is undoubtedly what teachers will focus on (see the classic paper by Kerr, 1975: "On the folly of rewarding A, while hoping for B"). At its extreme, the rewards may lead teachers into cheating on the tests – see section 4.b.i for further references to unintended responses to incentives. Finally, it is argued that providing monetary incentives for teachers can undermine their intrinsic motivation to do their best for their pupils.

It is, then, an empirical question whether the introduction of PRP schemes raises or reduces pupil attainment; see recent surveys in Neal (2012) and Jackson, Rockoff and Staiger (2014). The brief answer is that there are robust well-designed studies yielding both answers; this is not a settled question. For probably obvious reasons, it is very hard to be able to introduce a PRP scheme that admits an experimental evaluation. There are therefore relatively few studies credibly estimating a causal effect. We certainly do not know enough to be able to adjudicate all the different design issues noted above.

For example, Lavy (2009) finds a strong effect on maths and English scores of an experimental programme in Israel. The design was: individual teachers, monetary reward, and a tournament (relative) structure; the targets were high school graduation exams in English and maths. Lavy conducted a survey and found that the channels for the improvements were both changes in teaching methods, and additional schooling. In England, Atkinson et al (2009) also find a significant and substantial effect of PRP (albeit a poorly designed scheme) on pupil attainment. On the other hand, Fryer (2013) analyses the introduction of PRP in New York City and finds no effect at all. Here the design was team-based, based on thresholds, and the target was test score growth rates.

Overall, Neal (2012) argues that the balance of the evidence is positive. Muralidharan and Sundararaman (2011) show positive and significant impacts on test score gains from a bonus scheme in India. Woessmann (2010) finds strong supporting cross-country evidence, with the existence of bonuses for high performance being significantly associated with math, science, and reading achievement across countries. However, the key question is not so much “does PRP raise scores?” as “how best to design a PRP scheme to raise scores?”. As Imberman and Lovenheim (2012) put it, “the results from this analysis indicate that design features matter a lot in determining how effective an incentive system is in increasing student achievement” (p. 4)

There are three final design features to discuss. The team-nature of schooling is clear: by the time a pupil leaves a high school, it is quite likely that she will have been taught by every maths teacher in the school. Her final grade in maths is in some sense due to all of them. Furthermore, an individual scheme can lead to rent-seeking in terms of seeking to be assigned particular classes. However, economists are typically wary of group-based schemes because of the free-rider issue, so again it is an empirical issue. Lavy (2002) shows significant, positive effects from a group pay scheme in Israel, although smaller than in his 2009 study of individual pay. Other studies are able to test the free-rider hypothesis directly. Sojourner et al. (2011) find that districts in Michigan rewarding individual goals had large attainment gains, while whole school goals elicited no extra effort. Imberman and Lovenheim (2012) exploit naturally-occurring larger or smaller groups and find weakened incentives and lower attainment gains for the larger groups. Interestingly, their results imply “achievement-maximizing group sizes of between 3-5 teachers”.

Second, work by Fryer et al (2012) show that behavioural insights can add significantly to the power of incentives for teachers. They showed that invoking the power of loss aversion by paying the bonus up front shifts teacher PRP effects from small and marginally significant to much larger. Whether this is widely practical is not clear, as there will be situations where some low-performing teachers may not be able to repay the money.

Third, test scores are co-produced between teachers and the students (and possibly the parents too), so it may be that teacher incentives work better when the students are also incentivised. Both Behrman (2012) and Jackson (2010) find that schemes incentivising both pupils and teachers work better than schemes incentivising either separately.

I have devoted a lot of space in this review to teachers. This is because the research evidence clearly shows that teacher effectiveness is extremely important for pupil attainment. As Hanushek (2011) puts it: “No other attribute of schools comes close to having this much influence on student achievement.” (p. 467). Furthermore, while there has been a great deal of research in this field in the last decade, there is still a large amount we don’t know.

### **iii. Pupil effort**

One concrete expression of the demand for education is the effort put into co-producing it by a pupil. This is ‘effort’ in a rather broad sense – hours of study, but also concentration, a focus on the task in hand and a determination not to be distracted. The capacity to do this is related to the psychological traits discussed above, particularly conscientiousness, with the related traits of self-control, accepting delayed gratification, and a strong work ethic (see Heckman, 2011).

Conscientiousness has been shown to be an excellent predictor of educational attainment and course grades (Almlund et al 2011, Borghans et al 2011). This is an example of non-cognitive attributes enabling the development of more cognitive skills (see Cunha et al, 2010).

Bishop (2006) sets out a rich model of the decisions that pupils face, and highlights the role of school exit exams in signalling achievement and hence the input of the pupil. The role of pupil effort in producing educational attainment is only now being studied in a causal way. The empirical problem is finding some factor that exogenously causes the value of leisure time to change, thus changing the 'price' of time, energy and focus spent on school work.

There are a small number of such studies exploiting sources of natural variation. Stinebrickner and Stinebrickner (2008) found that randomly being assigned to a college roommate who has a video game console significantly reduces time allocated to studying (using self-completed surveys), which then negatively impacts on educational attainment. The estimated effect may include peer effects as well as the changing marginal value of leisure. Metcalfe, Burgess and Proud (2014) exploit the fact that the world's two largest football tournaments, the World Cup and the European Championship, overlap with the period of annual, compulsory, high-stakes exams in England. Since these tournaments occur every other year they have an alternating sequence of treatment and control years, with and without distractions, and test the impact on attainment. They find heterogeneous effects, but for some groups, particularly boys from disadvantaged families, the impacts are substantial and significant. Similarly, Lindo et al. (2012) show how college students in the University of Oregon are negatively affected in their exams when the college football team is doing well. Focussing on other distractions, Hill (2015) uses the friendship data in the Add Health study to look at the influence of the share of opposite-gender school friends. He shows that a high share reduces attainment quite substantially. He additionally shows that the effects work within the classroom through difficulties in concentrating and interacting with the teacher, and outside the classroom through spending time on romantic relationships.

A different approach is to estimate a causal effect by introducing an incentive programme for pupils. This explores the nature of pupil motivation and how much their input matters in the educational production function. Of course, as we have seen, acquiring qualifications earns a return and so is already incentivised. But it may be that not all pupils see this and for those the provision of an immediate incentive now, while studying, may be in Fryer's words "a way to straddle the perceived cost of investing in human capital now with the future benefit of investment." (Fryer, 2011, p. 1756). There has been a variety of research on such incentives in recent years. Some programmes have offered incentives based on enrollment and attendance, such as Progreso in Mexico (Behrman, Sengupta, and Todd, 2005) and a similar conditional cash transfer program in Columbia (Barrera-Osorio et al., 2008). Others have conditioned incentives on test performance, such as those using high school exit and achievement exams in Israel (Angrist and Lavy, 2009) and Texas (Jackson, 2010); standardized tests in Kenya (Kremer, Miguel and Thornton, 2009), Ohio (Bettinger, 2010) and New York City (Fryer, 2011); as well as assessment tests in Chicago (Levitt et al., 2012), Dallas (Fryer, 2011) and Houston (Fryer, 2012). There have also been programmes that reward college students in Canada (Angrist, Lang and Oreopoulos, 2009), the Netherlands (Leuven, Oosterbeek and van der Klaauw, 2010) and several U.S. cities (Barrow et al., 2014, Barrow and Rouse, 2013). These vary in the results they find; some showing strong positive effects (for example, Jackson, 2010; Kremer, Miguel and Thornton, 2009), and others finding no effect (Bettinger, 2010; Fryer, 2011) and others

finding very heterogeneous effects both positive and negative (Leuven, Oosterbeek and van der Klaauw, 2010). Burgess et al (2015) do something different by incentivising inputs, effort itself, rather than incentivising outputs. They find heterogeneous effects, ranging from zero for arguably already highly-motivated pupils, to substantial positive effects for other pupils. Levitt et al. (2012) find that offering students immediate incentives on a low-stakes tests improves scores ten to twenty percent of a standard deviation, purely through increased effort on the test. One concern in this literature is the impact of incentives on implicit motivation. In general the literature does not find that the former reduces the latter; and Borghans et al (2013) run in a laboratory experiment which shows the result that intrinsic and extrinsic motivations can be complements.

#### iv. Peers

There seems to be a widespread belief that peer effects are important in learning. This may simply be on an anecdotal level that many people feel that they have learnt from being with others. This interest is reflected in a large economics research literature on peer effects. There are a number of useful recent surveys available: Sacerdote (2011) principally focusing on the evidence, Epple and Romano (2011) providing a thorough analysis of the theory and econometrics, and Betts (2011) focussing on one of the main policy issues related to peer effects, namely tracking or setting in schools.

Peer effects in education come in many forms, from the simplest idea of direct learning from one's peers, through the influence of peers' backgrounds on the quality of learning resources, to much more indirect channels such as peers' families being very demanding of school performance. Other channels that have been suggested are based on peer behaviour (such as disruptive children in class, Lazear, 2001; Figlio, 2005) or on peer characteristics (such as the racial or gender composition of class, see Fryer and Torelli, 2010; Hanushek, Kain and Rivkin, 2009). Bursztyn and Jensen (2014) show that peer pressure to publicly conform to anti-study norms makes some students work less hard, so the composition of the peer group in a hard to observe dimension affects behaviour. This follows Fryer and Torelli's (2010) study of similar peer pressure against "acting white" among black students.

One key issue for econometric analysis and policy is understanding the mechanisms that create peer groups. Occasionally they are exogenously formed by administrative rules such as exploited by Sacerdote (2001) in his study of dorm- and room-mates at university. More typically, school peer groups are created by two endogenous forces. First, there is neighbourhood and school choice (see section 4.a.ii, and the survey in Nechyba, 2006), which can produce more or less segregated school intakes. Second, sorting within schools is created by tracking or setting – that is, assigning pupils to classes on the basis of attainment (see Betts, 2011). There is of course a further layer, taking peer groups to be friendship groups rather than class-mates, but research here is limited by the lack of datasets with friendship networks (Fryer and Torelli, 2010; Burgess, Sanderson and Umaña-Aponte, 2011). But where available, such network data are valuable: Bramoullé, Djebbari, and Fortin (2009) exploit assumptions on network structure to identify causal peer effects.

The econometric problems associated with endogenously formed peer groups are well known following Manski's (1993) foundational paper; see Epple and Romano (2011) for a thorough

explanation. To try to estimate causal effects, researchers have typically used exogenous assignment where available, or natural variation and school or student fixed effects to control for (self-) selection.

Sacerdote reviews the evidence on peer effects on attainment in schools and notes a wide variation in estimates of effect sizes, from zero to substantial. However, he argues that the results can be largely reconciled by considering the functional form adopted, and highlights a typology of peer effects models from Hoxby and Weingarth (2005). The key is a move away from a simple linear-in-means characterisation of peer effects. Such a model assumes that all pupils are equally affected by the class average ability. Apart from typically not fitting the data well, a linear model is not very interesting as it implies a zero-sum situation in which re-grouping pupils can have no effect on overall average attainment. They propose and characterise a number of non-linear models, in which the attainment interactions are richer and more complex. These models do have non-trivial policy implications for class allocations. For example, it may well not be the class average that matters as much as the number of high-performers, or the number of low-performers in a class. Hoxby and Weingarth find that both low-performing students and high-performing students benefit from having class-mates like them, rather than just having high performers. More generally too, studies have tended to find complementarities rather than linear-in-means models fit the data better, with low ability pupils harmed by adding high ability peers and removing lower ability peers, though Vigdor and Nechyba's (2007) results disagree.

This evidence favours the idea of tracking or setting – teaching a class of high-ability students and a class of low-ability students, rather than two mixed classes. Sacerdote (2011) notes that about half of the research addressing within-school tracking specifically finds positive effects from the policy. There has to be some reason why the practice has endured in schools, particularly given a general tendency in schools to favour egalitarian approaches. One recent example is Kremer, Duflo and Dupas (2011) who test models of tracking in a randomised control trial in Kenya. They find that average scores were 18% of a standard deviation higher in the tracked schools relative to non-tracking schools; furthermore, students in the top half of the ability distribution gained by 19% and in the bottom half by 16% of a standard deviation, and students throughout the distribution benefited from tracking.

For tracking, the key question is the relative value of having high-achieving peers in your class, versus the value of having peers like you in your class. The potential benefit of the former is clear; the potential benefit of the latter is allowing teachers to target their teaching style and content more precisely to their pupils' needs. If the former is more important then tracking increases inequality; if the latter is larger then it raises attainment for all and does not materially affect inequality (Figlio and Page, 2002).

Betts (2011) rightly contrasts within-school tracking, discussed here, with between-school tracking as practised in some European countries, discussed below. The argument is that different school 'types' – for example academic versus vocational – are very differently resourced, thus exacerbating inequality. For example, Brunello and Checchi (2007) show substantially worse pupil-teacher ratios in the more vocational schools in Germany, France, and Italy. It is widely argued that such a system increases inequality. I discuss this point more fully in section 4.a.iii on elite schooling systems.

## c) The supply of higher education

Universities are education institutions like schools. Many of the issues facing universities, lecturers and their students are the same as those facing schools, teachers and their pupils. However, there are also some new issues, and different perspectives on the common issues. While schools are often competing for pupils, the competition between universities is arguably more intense. This is partly because of its international nature, partly because of the emergence of new forms of provision of higher education, and also partly because this is post-compulsory education and students have the option, denied to school pupils, of not going at all.

With universities as initially with schools, the bulk of the research has been on the demand side, analysing rates of return, the impact of fees on attendance and so on. In fact, there are many open questions on the supply side too, and research here may illuminate some of the more puzzling issues in terms of university and college behaviour. For example, one question about changes in university fees illustrates the “demand side only” approach: if the price (tuition fee) goes up, why does quantity (attendance) go up not down? The counter argument is about the supply side: the supply of places at university also bites and extra funding for universities can relax the quantity constraint. This point is made by Chapman (2006) specifically in regard to Australia: “typically many public higher education systems are supply-constrained, and this was certainly the case in Australia at the time of the introduction of HECS [graduate tax]. The effect of the introduction of the scheme was to encourage the government to outlay substantially more resources for university places given the promise of higher future revenues” (pp. 1493/4)

The fact of a binding quantity constraint has another implication of course: there has to be some rationing system for the scarce places. Often this is accomplished by selection, by imposing some minimum score level on previous tests. McPherson and Schapiro (2006) focus on some supply side issues, including pricing and quantity decisions.

### i. Competition

The market for higher education has many suppliers. In some countries there is a tradition of going to university close to home, in others a tradition of moving away. In either case, the market is likely to be competitive. What is the impact of this? There are some interesting issues here. Not least because some of the suppliers are for-profit, some are charitable, and some are straightforward public institutions. All contain key staff (ie academics) with conflicting interests between teaching and research. Two early papers focussed on the joint production issue in universities. Del Rey (2001) and De Fraja and Iossa (2002) analysed the situation where competing universities had to choose the balance of resources between teaching and research. The latter show that the degree of student mobility is key (unsurprisingly: this is a measure of competition) and that at non-extreme levels only asymmetric equilibria may exist, which they characterise as the emergence of ‘elite’ institutions. More recently, Epple et al (2013) set out a general equilibrium model for the US higher education market.

A very useful paper from Aghion et al (2010) sets out and quantifies many of the issues. They are interested in the industrial organisation of the higher education market, and specifically the relationship between market structure, governance and productivity. They show strong positive correlations between measures of university autonomy and the level of competition, and productivity. To get at a causal relationship they exploit political appointments that generate exogenous funding changes for US universities. These show that funding produces more output (publications and patents) if the university is more autonomous and is in a more competitive environment.

In addition to competition from regular education institutions, universities might start to feel some competitive pressure from novel sources of supply. This includes specific professional training courses such as Knowledge Academy (<https://www.theknowledgeacademy.com/>) and Code Academy (<https://www.codecademy.com/>). Many of these courses are far from cheap (for example, €6595 for a website design course in Brussels at Knowledge Academy) and allow people to build up a significant and valuable skill set piece by piece. Obviously all of the discussion above about course fees, loans and access applies here too.

At the other end of the scale are MOOCs (Massive Open Online Course), free and available anywhere, much heralded as completely changing higher education (see Barber, Donnelly and Rizvi, 2013). So far, reality is a little different, with most users of these courses already highly educated. But there are studies showing that some students of these courses use them for professional or employability reasons. Walker and Brooks (2014) look at people taking 5 MOOCs from the University of Michigan and divide them into two groups: “strivers” (typically young, male, non-native English speakers, living outside the United States) mostly taking courses for professional and employment reasons, and “grazers” (generally living in the US, more likely to speak English natively, less likely to be working or studying in a related field) mostly taking the courses for enjoyment. Because of the novelty of MOOCs, there is very little research on their value, their cost and successful characteristics. It would be useful to know, for example, how valuable MOOC qualifications will be in the labour market; presumably if the medium is successful in transferring valuable skills reliably, then they will be valued.

## ii. Spill-overs

Part of the value of universities is that they are believed to yield spill-over benefits to the neighbouring region. This is in part the human capital spill-overs from having more highly educated people around (Moretti, 2011) deriving from better knowledge sharing and the faster adoption of new ideas (see also Goldin and Katz, 2008).

Neumark and Simpson (2015) review the evidence more broadly on spill-overs from universities, and note that belief in spill-overs leads policy-makers to invest in universities as a way to regenerate regions. If spill-overs increase with proximity then such investment might help to create a local technology cluster. The simple observation that the best universities are typically located in thriving and innovative regions does not of course provide causal evidence for the strength of spill-overs. The non-random siting of universities means that other evidence is needed. Abramovsky and Simpson (2011) consider whether the presence of leading university research departments is associated with the choice of location for firms’ R&D facilities. They show that this matters more for some industries

than others: pharmaceutical firms do tend to disproportionately place their R&D labs close to the best chemistry research departments.

Andersson et al (2004) study the case of Sweden, which undertook a substantial expansion of universities from 1987, and show subsequent positive effects on local labour productivity, with stronger effects for the newly established universities. They followed the analysis up (Andersson et al, 2009) and confirm the findings, also using an instrumental variables approach to deal with the potential endogeneity of the location. Veugelers and Del Rey (2014) provide a review of the evidence on the contributions of universities to innovation and employment in Europe.

#### **d) Training, vocational skills and adult skills**

Vocational education is not a topic in 'the economics of education' in the same way that 'teacher effectiveness' or the 'rate of return to education' is. It is a policy issue, and not one always one high up the agenda, that borrows analysis from other parts of the field. It is also at least as much the subject of more mainstream labour economics as it is education economics; talk of 'training programmes' recalls some very famous papers in labour economics (LaLonde, 1986, for example).

One of the initial issues is defining vocational education. The general idea is education or training directed at a particular occupation or skillset; this is as opposed to academic courses which are more general and more abstract. But on top of this is a sense that it is rather low-level. For example, training to be a doctor or a lawyer is not referred to as vocational education; people have in mind skilled jobs in catering or building or fashion or car maintenance. The fact that it is typically lower level skills means that it is often relatively neglected in policy discussions and that this neglect is problematic for less able pupils. Vocational education includes more skills-based courses within regular schools and also a concentration of such courses in particular types of school, for example technical schools in Germany and the Netherlands. It includes youth skills training programmes beyond compulsory schooling and also adult training.

There are important policy questions here. What is the return to vocational qualifications? How does this vary by type of vocation, and by type of institution? How should these courses be delivered: in regular schools, or in separate organisations closer to the labour market? How do these compare to apprenticeship programmes? How can low skill adults be encouraged to raise their skills?

There is little direct evidence on many of these questions. Taking the rate of return first, it is instructive to start with an example from the UK of what would clearly count as a 'vocational' course – the University of Birmingham's degree in Applied Golf Management Studies. As the UK Minister for Universities said in 2010, "... It's all too easy to scoff ... to regard [it] as just a "Mickey Mouse" degree". But he went on to note that "it had 109 applications for 25 places ... its graduates from 2009 all secured jobs ... at an average salary higher than the average starting salary of Birmingham graduates as a whole". So it is clear, firstly, that there are high value vocational qualifications, and secondly, that there are likely to be very heterogeneous rates of return to such courses. And of course vocational courses that are not thought of as vocational courses generate very high returns: Ketel et al (2015) show a very high rate of return to attending medical school in the Netherlands.



McIntosh (2006) and Dearden et al (2002) estimate rates of return to a long list of available vocational qualifications, albeit without a watertight argument for a causal interpretation. They show that many vocational qualifications do earn a rate of return, particularly for lower ability students, but that the rate of return is typically below that of academic qualifications of a notionally similar 'level'. Oosterbeek and Webbink (2007) find zero return to the lengthening by one year of schooling in basic vocational schools. They interpret this as showing that the extra year on a vocational course is no more valuable than spending that year gaining experience in work.

More broadly, Attanasio et al (2011) note that there is little reliable and consistent evidence on the impact of training on raising skills in developed countries, with mixed results for evaluations of government training programmes. Card et al (2011) concur, arguing that the US evidence shows the impact of job training programmes to be modest, and that the European evidence is very variable due to the lack of experimental studies. Attanasio et al (2011) and Card et al (2011) themselves provide evidence of effective job training programmes in Colombia and Dominican Republic respectively. Heckman LaLonde and Smith (1999) provide a relatively recent survey of this literature, and Kluge et al. (2007) offer a meta-analysis of results. A recent survey of apprenticeships in Wolter and Ryan (2011) focusses on the labour market links rather than the education links and highlights the huge variety of quality, quantity, and financing.

There seems to be little robust evidence on the best setting for such provision. To some extent this echoes the debate on tracking discussed above, between a within-school tracking approach in which more vocational courses are provided alongside more academic ones, against an across-school tracking approach in which there are specific vocational schools (see Brunello and Giannini, 2004, for a theoretical model of optimal school system design in terms of a split between vocational and academic education). This latter is more common beyond the compulsory school leaving age. In terms of vocational versus academic courses there is more of an argument for split institutions as the teaching skills and knowledge are more diverse than (say) teaching maths to both high-ability and low-ability groups.

Finally, the study of adult skills was given a tremendous boost by the publication of the first Survey of Adult Skills from the OECD's Programme for the International Assessment of Adult Competencies (PIAAC), covering 33 countries. The survey measures both cognitive and workplace skills needed for work and life. The first report on the findings from this survey, OECD (2013) highlighted significant problems across the OECD of low skills in literacy, numeracy and IT. Although the report was optimistic about the role of adult learning programmes in reducing the problem, as noted above, the causal empirical evidence on training programmes is not that promising. Brunello and Rocco (2015) analyse the impact of vocational training on adult skills in the PIAAC data and report mixed results – training at the lower levels is about as good for earnings as more academic education, but less so at higher levels. Albrecht, van den Berg and Vroman (2004) found that a large scale Swedish adult skills programme raised employment prospects but not income for men and had no effect for women. Oosterbeek (2013) reviews public policy towards adult learning across Europe. He notes a lack of rigorous evaluation studies on training funds, but that training subsidies carry a large dead-weight loss. A recent survey provides a very detailed overview of policies towards adult learning in Europe (European Commission, 2015).

## **4. Education systems – assignment, admissions, accountability, autonomy**

The human processes underlying the formation of human capital, the delivery of effective education, and the returns to schooling form the ‘fundamentals’. The systemic issue is to design an education system that facilitates the best outcome given those fundamentals. The important role of the family in early education suggests that an education system could be construed in a very broad sense to include areas of social policy. However, that is beyond the scope of this review and this section relates to the education system as typically understood, relating to schools and higher education. As a thought experiment, the ideal system could be considered relative to a system of complete centralisation with total state provision and funding, or relative to a pure market system. Given the diversity of education systems in the EU, this is not an approach that space permits here.

### **a) Assignment mechanisms**

At the heart of every school system is a set of rules to assign pupils to schools. Consider one city; there is a set of pupils with particular characteristics (location, parental background, ability) and a set of school places with characteristics (for example, a highly effective school or an ineffective school, a particular specialism, location). Assuming that there are enough places overall, the question is: which pupil goes to which school? What is required is a mapping that assigns each pupil to a school based on her characteristics and its characteristics. The dependence on characteristics can be null – for example, a simple lottery over all pupils into all schools. Or the function might assign pupils purely on location, or on ability, and so on. Another mechanism is choice: families list their preferred schools and this plus school priorities determines the assignment.

The assignment mechanism constitutes the main element in the ‘rules of the game’ in the education market. As such, it is part of the incentive structure of all the players, families and pupils, and schools. Different assignment mechanisms will generally yield different outcomes for a range of measures of interest: mean attainment, variation in attainment, school sorting, social mobility and inequality.

The most common assignment mechanisms are: neighbourhood schooling (each pupil goes to her local school); tracking or elite schooling (schools are allocated on the base of a test score); and choice-based schooling (school assignment depends on parental choice and school capacity). I also consider assignment based directly on income – the role of the private sector interacting with state schools. I discuss these in turn, and the evidence on how they affect outcomes. But first, I review evidence on parents’ preferences for schools.

### **i. What are parental preferences for schools?**

Preferences matter most obviously under a regime of school choice. But they also matter whatever the assignment mechanism as there will in general be strategies available to parents to raise their chance of getting their most preferred school. This includes moving house under a neighbourhood schooling rule, and intensive additional coaching under an exam-based assignment rule.

There are a number of empirical challenges in estimating preferences for schools, particularly around identification. It is generally impossible to know the pool of schools that parents consider when making their choice, so this has to be estimated. Also, given that admissions to popular schools have to be rationed, it can be difficult to disentangle parental preferences from school priorities. Finally, as I discuss below, it is not always optimal to put down the truly preferred school as the top choice and this also complicates the analysis.

Hastings et al (2008) use school choice data from Charlotte-Mecklenburg, North Carolina, to estimate a mixed-logit demand model for schools. They find that parents value school-home proximity and academic attainment highly. They also find considerable heterogeneity in preferences and for example show that the preference attached to a school's mean test score increases with household income and with the child's academic ability. They use their model to estimate the elasticity of demand for each school with respect to mean test scores in the school. They find that demand at high-performing schools is more responsive to increases in mean test scores than demand at low performing schools. Their model also implies a 'mobile' more affluent group of families exerting pressure on school performance, and a less mobile less affluent group essentially going to the local school.

Hastings and Weinstein (2008) make an important distinction between a family's preferences for school characteristics and the information they are able to access about the schools. Using a mix of field and natural experiments, they show that the provision of additional information on school characteristics does change school choices, particularly for disadvantaged families.

Burgess et al (2014) estimate the preferences of parents for schools in the UK, as a function of the school-home distance, the academic performance of the school, school composition in terms of socioeconomic status and ethnicity, and whether it is a faith school. They pay attention to the key method for rationing access to over-subscribed schools (distance) and define a set of schools for each family in the data that they could almost surely access. They also find academic performance and proximity to be highly valued; social composition is also valued, but ethnic composition has no effect. By comparing schools that are feasible by distance and the subset to which the family has almost-sure access, they show that the use of the distance rule for rationing access has strong regressive effects.

There are a few other studies using different techniques. Schneider and Buckley (2002) use an online schools database in Washington, DC to track parent search behaviour for schools as an indicator of parent preferences. They find that patterns of search behaviour depend on parent characteristics, and find a strong interest in the demographic composition of a school. Rothstein (2006) adopts a more indirect approach to evaluate the relative weight parents place on school effectiveness and peer group and finds little evidence that parents focus strongly on school effectiveness.

## ii. Pupil assignment by school choice

School choice has been much studied, principally but not only in the US. It is seen as a way 'out': a way of escaping a low quality local school and attending a better school further away. It is also seen as the basis of school competition, raising standards in all schools. Research has considered the outcome of this process, for individual pupils as well as at a systemic level.

It is useful to distinguish two senses of 'school choice': as a systemic market rule for assigning all pupils to schools; and as a specific individual entitlement to attend a different school to your current one, rather like a voucher. I deal with the systemic market rule first.

### *Does school choice as a systemic market rule raise attainment?*

The claim is that school choice induces competitive pressure on low-performing schools to improve (Hoxby, 2003). If parents value academic attainment and have the information to recognise it in a school, if their choices make a difference to the allocation, and if schools benefit and expand with greater demand, then the market should operate to raise the attainment. Low-performing schools with lose pupils, lose funding and so work to reverse this by raising their performance. A counter-claim is that enhanced choice results in greater sorting or stratification across schools in poor and affluent neighbourhoods. Academic and policy debates on school choice remain controversial and unresolved. A recent contribution from Avery and Pathak (2015) reminds us of the complexities of school choice programmes when residence is a choice variable, and that the distributional consequences can be surprising.

Research in this field proceeds by defining some measure of the degree of choice that families have, and relating this measure to attainment scores. As always, the key issue is to identify a causal effect; there are many studies reporting associations between the two (reviewed in Belfield and Levin (2003)) but plausibly exogenous differences in competition are much harder to find.

A market for school places is inherently spatial – you have to actually be in the school all day, so measures of competition are about geography. This includes the number of different school districts that are close together (so could be chosen whilst working in the same job for example), the number of schools within a short drive from home and so on. For example, a number of studies use the HHI (Herfindahl-Hirschman Index) of local districts as proxies for competition (Hoxby (2000) for example). Alternatively, Burgess et al (2007) combine school locations with a complete road map to define 10-minute drive time zones around each school, and then for each school count the number of other schools in the zone. Feasible school choice is almost always going to be higher in densely populated urban areas, which are of course different in many ways from rural or suburban areas. Consequently, simple associations between this measure of choice and attainment are likely to be biased by confounding variables.

The best known attempt to establish causality is Hoxby (2000), who uses natural landscape features to instrument for historical school district boundaries and the HHI. She shows that areas with more school districts – higher competition – raise attainment. The findings have been strongly questioned by Rothstein (2007), however, arguing that they are not robust to simple changes in data coding; taking these into account he finds no impact of competition. A more structural econometric approach is taken by Bayer and McMillan (2010) who adopt an equilibrium sorting model between neighbourhoods (see Bayer, Ferreira, and McMillan 2007) and use the slope of the school's demand

curve to measure the degree of competition each school faces. They find that a one standard-deviation increase in competition leads to a 0.1 standard-deviation improvement in attainment.

In the UK, there have been two attempts to estimate a causal effect, both yielding low to zero impacts of competition. Gibbons et al (2010) use the distance of a school from its nearest local authority boundary to instrument the amount of competition it faces; they find no overall effect of choice or competition on school performance. Burgess and Slater (2006) use the administrative boundary change of 1998 that split Berkshire into six local authorities to estimate the impact on pupil progress of possible falls in competition across the new boundaries. They also find no significant impact of these boundary changes on pupil achievement.

School choice as a systemic rule has a long history in the Netherlands and in Denmark, and something of a history in Sweden since 1992. A number of studies of the Swedish system are discussed below. In the Netherlands, parental choice of school has been in place since the early twentieth century. Dijkgraaf, Gradus and de Jong (2012) find that increases in competition as measured by the HHI are associated with a small decrease in attainment, but this is not a causal study. Competition specifically from Catholic schools also appears not to have an impact. de Haan, Leuven and Oosterbeek (2011) using a law change for identification find a negative effect from a city having more but smaller schools, although the effect disappears once school size is controlled for.

Lavy (2010) considers a school reform in Tel-Aviv that switched from an inter-district bussing programme to a school choice system. As this is not experimental variation, Lavy uses alternative identification strategies and comparison groups and shows that the choice system increases school completion and raises cognitive achievement (it also raises students' satisfaction with school).

### *Choice as a voucher*

The idea of an educational voucher is that it entitles a child to go to a different school than her default or 'normal' school. Details vary hugely by scheme, but in essence it is seen as an 'escape' from a low quality local school. This is generally a specific entitlement (for example, Figlio and Page (2003) consider a scheme in Florida in which students in 'failing' schools are given vouchers, which they can use to move to an alternative school) rather than a system-wide assignment mechanism, although it is sometimes combined in system-wide reforms as in Sweden. The outside option school can be a private school (as in Sweden, though with capped fees) or a charter school as is often the case in the US. The biggest voucher programmes are in Chile and Colombia (see Bettinger, 2011, for a survey) but they are also part of the system in Sweden and the Netherlands; and of course in the US.

In all of these cases, there are two main research and policy questions: what is the impact of the voucher on the individual who receives it? And what is the impact on the system as a whole, on those 'left behind' in the low-performing schools? There are also complex general-equilibrium theoretical issues in voucher schemes that are summarised by Epple and Romano (2012).

There appear to be no definitive answers to the two core empirical questions yet. In a substantial recent review, Epple et al (2015) argue that the bulk of the findings suggest no significant effect, yet "multiple positive findings support continued exploration". The task now seems to be the attempt to understand the role of the context in determining the variation in outcomes. In surveying work outside the US, Bettinger (2011) argues that evidence from Columbia on the impact on the voucher-

user is possibly the strongest, but it may not be causal. On the second question, there is some evidence that the system improved in Sweden (for example, Böhlmark and Lindahl, 2007, Björklund et al 2004) but it is difficult to single out the voucher component as many reforms were introduced together in 1992. More recently Böhlmark and Lindahl (2012) now find small positive results from competition and choice, ten years on from the reform.

Evidence from the US is also complex and contested, and generalising is difficult given the differences in design (Ladd, 2002) and in some cases small numbers. The evidence on the impact on the voucher-using student is mixed. Peterson et al (2003), for example, examine data from three privately funded school voucher programmes in New York, Washington DC and Dayton, Ohio. In all three schemes, a lottery is used to allocate vouchers among eligible (low income) families, and the voucher does not cover full costs. Test-score gains from switching to private schools are evident for African-Americans but not for students from other ethnic backgrounds. Hoxby's (2003a) review of the evidence from recent studies using randomised control groups of students from lottery allocation mechanisms shows the same. Cullen et al (2006) collect data from the lotteries used to allocate students to oversubscribed schools in the Chicago Public School (CPS) system; arguably in the CPS choice is essentially systemic as over half of pupils do not attend their default school. Cullen et al find that winning a lottery has no impact on test scores at ninth or tenth grade. They speculate why this might be, but it is not because the treatment is not a treatment, as the lottery winners did attend schools that are better across several dimensions. Nor is it that winners had longer school commutes and more disruption to their friendship groups. It is true that lottery winners have lower in-school rank than the losers, which may be a factor in greater school drop-out. They do find positive effects on non-academic outcomes and consider that this might be the reason that parents enter school lotteries rather than for attainment improvements. Howell's (2004) work on New York City also cautions that the final users of targeted vouchers may differ significantly from the average intended user: among targeted voucher schemes, those actually using them tend to be the better off in the group.

Turning to the question of the systemic impact, Hoxby (2003c) investigates the causal impact of three school choice reforms: vouchers in Milwaukee, charter schools in Michigan and charter schools in Arizona. In each case, state schools responded to competition from the choice programme by raising the achievement levels of their remaining students. This increase was sufficient to outweigh any negative allocation effects. Hoxby's analyses are not unchallenged, Ladd (2003) noting that the Milwaukee programme was part of a broader package. Bettinger (2005) challenges the findings for Michigan, and Bilfulco and Ladd (2006) find negative impacts in North Carolina.

### *Does choice raise sorting?*

Analysis of choice and sorting is complex with theoretical analysis as well as empirical work contributing insights. Hoxby (2003b) argues that there are no very general theoretical predictions about student sorting with choice. In particular, she argues that 'cream skimming' (schools actively selecting high ability students) is not a general prediction, but is more likely with broad eligibility for vouchers and a uniform value of vouchers; if vouchers are targeted, this will necessarily reduce sorting. Nechyba (2003a, 2003b and 2004) uses a theoretical approach to explore the complex 'spillover' effects of school choice and sorting. For example, Nechyba (2003b) shows that a pure state school system leads to more spatial segregation than a private system. Nechyba (2004) summarises work on income and ability sorting, discussing different channels of sorting. Similarly,

Epple and Romano (2003) analyse three different student assignment regimes: neighbourhood schooling (a strict residence requirement for admission); school choice with no choice costs; and choice over many school districts and show that different public policy regimes have dramatic effects on the nature of sorting. Neighbourhood schooling leads to strong income stratification across neighbourhoods, whereas costless, frictionless choice equalises peer groups across schools. Much of this theoretical work, however, analyses a system where individual schools can grow or shrink costlessly to accommodate the outcome of parents' choices; this flexibility is often lacking, in which case the theory is not so useful a guide.

Evidence from England, New Zealand, Sweden and the United States suggests that the degree of choice does influence the degree of sorting. For example, Burgess et al (2007) analyse student-level data from England, and show that the degree of student sorting in terms of ability and socio-economic status varies considerably across the country. Looking at choice, they measure the degree of choice as the number of schools that can be reached within a particular drive time and show that school sorting relative to residential sorting is considerably higher in areas where there is more choice. Cullen et al (2000) show that in the Chicago state school system, the exercise of parental choice leads to an increase in sorting by ability, although Hoxby (2003a) argues that Chicago does not have pure school 'choice' as money does not follow students, and schools cannot expand or contract much in response to the demand. Again it is worth noting that that last feature is not uncommon. Söderström and Uusitalo (2010) analyse student level data from Sweden, and compare student sorting before and after a significant reform to the school admission process in Stockholm, switching from a predominantly residence-based admissions system to an explicitly ability-based system. Unsurprisingly, they find a significant increase in ability sorting in schools, but no change in residential sorting.

Overall, the evidence suggests that, compared with neighbourhood schooling, parental school choice with supply-side flexibility reduces sorting. Parental choice plus poor flexibility on the supply side means that schools have to use some criteria to choose students. The evidence from a number of countries suggests that this combined process of choice by parents and proximity rationing schools leads to greater sorting.

#### *What are the best (truth-revealing) market mechanisms to implement choice?*

Allocations based on school choice need a way of aggregating parental choices and school priorities to yield an assignment. In turn, the nature of the mechanism will affect parents' school nominations (Roth, 1991). Ideally, that mechanism should have optimal properties, for example including the Pareto characteristic that there is no other assignment preferable to all; and whether it elicits parents' true preferences. Abdulkadiroğlu and Sömnöz (2003) set out the mechanism design approach to school assignment, and Abdulkadiroğlu et al (2005a, b) apply this approach to the Boston and NYC school districts respectively, and Pathak and Sömnöz (2008) and Abdulkadiroğlu et al (2008) subsequently update the design. These papers determine the properties of particular assignment mechanisms and whether they elicit true preferences from the participants. Revealing true preferences is a weakly dominant strategy in two common mechanisms, Student Proposing Deferred Acceptance (SPDA, Gale and Shapley, 1962, also called Student Optimal Stable Matching) and Top Trading Cycles. More recent refinements, for example restricting the number of schools that parents are allowed to nominate, show that when parents can make only limited nominations, truth



telling is not optimal in some circumstances even with an SPDA mechanism (Haeringer and Klijn (2009) and Calsamiglia et al. (2010)).

### **iii. What is the effect of tracking and selective schools on the distribution of attainment?**

An alternative way to assign pupils to schools is by a measure of ability, typically by setting an exam. This is the way that the school system works in a number of European countries. The public school system in Germany is tracked, with entry to the *Gymnasium* schools determined by exam performance. In the Netherlands, Switzerland and France (from age 15) too, pupils are assigned to different curricula based on their ability or attainment. In some cases these are different curricula within the same school, and in others, different schools; I return to this issue below. This was also the case in England and the Nordic countries until a wave of comprehensive reforms were adopted from the 1950s through the 1970s. In the US, elite 'exam schools' in New York or Boston are accessed by taking a competitive exam.

The two main questions that researchers have asked are: what is the impact on the system as a whole in terms of inequality and social mobility, and what is the benefit to the student of attending the elite schools. Again the empirical concerns are around dealing with the selection issues inherent in the problem and identifying a causal effect. A theoretical contribution from Brunello and Giannini (2004) sets out the trade-off in designing an optimal tracked system, differentiating vocational and general education in terms of required labour market skills.

There are two core distributional questions on tracking. Using cross-country evidence and a difference-in-difference methodology, Hanushek and Woessmann (2006) show that it raises inequality. Brunello and Checchi (2007) show that tracking from an early age across schools reinforces the impact of family background on attainment and labour market outcomes and so reduces social mobility. On the other hand, they report more nuanced results of tracking on the scope to access vocational training. The overall level of attainment is lower under tracking and it seems plausible that some families and pupils might reduce their investment in schooling if they know that they cannot go to higher education. Atkinson et. al. (2006) use NPD data to compare value-added attainment across selective and non-selective school districts in England. They use matched selective and non-selective districts and show that grammar-educated children in selective districts outperform similar children in non-selective districts on average, while non-grammar-educated children in selective districts underperform compared to similar children in non-selective districts. This fits well with the results of Burgess et al (2014) which show that earnings inequality among children growing up in selective areas is greater than that of similar children growing up in non-selective areas.

Major systemic school reforms took place in Sweden in the 1950s and Norway in the 1960s. These raised the school leaving age so that mandatory schooling was extended by two years and the system became comprehensive so that all students followed the same track. These have been studied by Meghir and Palme (2005) and Aakvik, Salvanes and Vaage (2010). Both studies found a weakening of the influence of family background, and Meghir and Palme show a causal impact of increased earnings among pupils from disadvantaged families. Pekkarinen, Uusitalo and Kerr (2009)



exploit a similar reform in Finland in the 1970s and show that the elimination of tracking reduced the intergenerational elasticity of income very substantially.

Students in the elite schools may benefit in many ways, as Brunello and Checchi (2007) describe: pupil peer effects, more effective teachers, and possibly greater resources (all these three discussed above). Estimating the gain to the marginal student of attending an elite school, Clark (2010) uses access data from a district in England to estimate the causal impact of attending a grammar school. He finds small effects of grammar schools on test scores at 16 but larger effects on longer-run outcomes such as taking more advanced courses and more academic courses. Clark and Del-Bono (2014) implement a regression discontinuity design to assess the impact of attending a grammar school for a cohort of young people born in Aberdeen in the 1950s. They find large effects on educational attainment, and for women there are long-run impacts labour market outcomes and reduced fertility. For men no long-term impacts were identified. In the US, Abdulkadiroglu, Angrist and Pathak (2011) and Dobbie and Fryer (2011) assessed the effect of attending elite exam schools in Boston and New York on attainment and test scores. Both studies found limited impacts on student achievements, though Dobbie and Fryer (2011) found positive effects on the rigour of the courses taken.

The top-level distinction is between comprehensive and tracked systems. As with the discussion of peer groups above, the key trade-off is between the unequalising effect of differential peer groups (and potentially lower overall investment) in tracking, against the potentially more efficient teaching possible from more homogenous classes that tracking brings. There is also an important distinction between within-school tracking and between-school tracking; the former meaning different curricula, different tracks, offered within the same school, and the latter meaning different schools. Here the issues are about the difficulty of rectifying incorrect assignment of children to schools, versus the cost and practicality of running many curricula within the same school. It seems that the slowly increasing age of tracking and the greater frequency of within-school tracking suggests that the latter is less of a problem.

#### **iv. What are the effects of neighbourhood schooling assignment rules?**

Neighbourhood schooling means that every child goes to her neighbourhood school. This is well illustrated by Fack and Grenet (2010) for Paris: "During the period under study (1997–2004), primary and middle school assignment was purely residence-based. It was also "strict" in the sense that each school catchment area contained one school only, which means that in principle parents had no control over the choice of their child's public school." (p.62) and "School catchment areas are released every school year in the form of booklets that indicate, for each street section, the assigned public middle school" (p. 63).

What are the implications of this? It does not make families into passive players in the school choice process; it simply turns the school choice into a choice of residence. Given parents' preferences for schools discussed above and a rule that you gain access to a certain school by living in a certain place, popular schools imply popular neighbourhoods. This affects housing demand and so house prices.

There is a substantial literature trying to estimate the true house price premium arising from a popular local school. An influential study is that of Black (1999), in which she adopted a regression discontinuity approach, comparing house prices either side but very close to a school assignment boundary to model the impact of school quality on house prices. She found that families paid 2% more on the value of the house for a 5% increase in academic quality measured by test scores. Gibbons and Machin (2003, 2006) carry out a similar analysis for primary schools in England, with similar results. The literature as a whole has been summarised by Black and Machin (2011): “parents are prepared to pay substantial amounts of money to get their children educated in better performing schools. ... A not unreasonable benchmark summary of the magnitude of the average causal impact is that a one standard deviation increase in test scores raises house prices by around 3%”

This has implications for schools and for neighbourhoods themselves. These are principally around sorting or segregation as discussed above; there are subtle externalities and dependencies at work. Nechyba (1999; 2000; 2003) and Epple and Romano (2003) among others have shown that neighbourhood schooling in a model with peer effects implies income and residential sorting in equilibrium. One of these models' key parameters is the valuation of school performance by parents: the higher is this parameter, the higher the level of sorting. It is important to stress that, far from producing an even mix of students (no sorting), neighbourhood schooling produces strong sorting of students by income and ability. This is because parents take steps to achieve their chosen school through other means – by choosing where they live. So the level of sorting in the absence of choice is potentially quite high.

This sorting produces very heterogeneous income-segregated neighbourhoods, which may matter for reasons beyond education. For schools, this will affect the distribution of attainment if peer groups are important in the education production function (see section 3.b.iv). But it also matters for inequality in access to the best schools. The high house prices exclude access to the highest-performing schools. Note that this also applies to choice-based schooling when the proximity criterion is used for rationing places under choice rules (see Burgess and Briggs, 2011, and Burgess et al, 2015, for estimates of this effect).

#### **v. Assigning by income: private schools and the state sector**

A fourth mechanism for assigning children to schools is by income and choice, to private, fee-paying schools. There is huge variation across the OECD in the fraction of pupils attending private schools, see OECD (2012), p. 21. This variation arises in part exogenously from different national laws and regulations, and in part endogenously from the attractiveness of free state schools. There are different dimensions of “private” schooling. This includes the degree of public/private funding and also whether the school is privately managed (again see OECD, pp. 19 -21). An alternative way of thinking about this is whether private schools fees’ are unregulated (for example in England) or are capped at the level of state funding (for example in the Netherlands and Sweden) in which case the ‘private’ aspect is in the operations and management of the school, and the system is more akin to a voucher scheme.

What are the implications of these schools for the national education system? Focussing first on the pupils themselves, since attending these schools is a pro-active choice, revealed preference suggests

that the parents are happy with the outcome relative to the marginal state school (for example, Green et al, 2012, track the changing earnings return to a private school education in England). As Brunello and Rocco (2008) note, this outcome may not always be the level of attainment, but may be the ability to cope with difficult-to-teach children.

The main question is about the systemic impact, about the interaction between state and private sectors. Epple et al (2004) set out a model of price-setting by private schools faced by state schools in the same market. Fack and Grenet (2010) discuss interaction in admissions – the impact of a local private school on admissions in an otherwise neighbourhood schooling scheme. OECD (2012) shows that socio-economic stratification across schools is not associated with the prevalence of privately managed schools in a country, but it is associated with the level of public funding to those schools. For example, in Sweden, Germany, Belgium, the Netherlands and Ireland, over 80% of school funding for privately managed schools comes from the government. By contrast, in the United Kingdom and Greece 1% or less of funding for privately managed schools comes from the state. In those countries where privately managed schools receive higher proportions of public funding, there is less stratification between publicly and privately managed schools. Green et al (2008) consider competition between state and private schools in the market for teachers. They show that private schools are increasingly recruiting teachers from the state sector. Teachers in the private sector report greater job satisfaction; while this may be causal, it may well be about selection into sector and a better worker-job match.

## **b) Accountability, autonomy and regulation**

Schools are given two very valuable resources by the government – a large amount of public money and, far more valuable, the task of optimising the future skills of the nation’s children. Schools should be accountable for how they deal with these resources. This accountability is enacted in different ways and to differing extents in countries around the world. The implications of this are discussed below.

Accountability makes most sense when those being held accountable can actually make a difference to the outcome – that is, have some autonomy in the running of their schools. Strong accountability mechanisms seem inefficient and unfair without autonomy. Evaluating school autonomy is a relatively recent topic of research generating interest in the US and the UK in particular, and this is reviewed below. Studies using international comparative tests suggest that market features enabling school accountability and autonomy are important for student performance (eg Woessmann, 2007).

### **i. School Accountability**

The essence of school accountability is the provision of rewards or sanctions attached to the test performance of pupils in the school. The sanctions or rewards can be explicit, such as the replacement of school leaders, or implicit, such as good performance raising applications to the school. The theoretical argument basis for the accountability system is a principal-agent model; the

publication of school performance data helps to monitor the output of the school. These tables might be scrutinised by parents who could react by avoiding low-performing schools and/or by the education authorities who may take action against such schools.

### *What effects does the accountability system have on pupil performance?*

Researchers face two main difficulties in trying to establish the impact of accountability systems on pupil performance. Figlio and Ladd (2008) note that typically a multifaceted performance management reform is introduced all at once, removing the possibility of evaluating an individual component; and that finding an adequate control group for the counter-factual is often hard.

Causal evidence on this comes from changes in accountability systems. Burgess et al (2013) are able to exploit a policy experiment that changed school accountability in Wales but not in England. Following a referendum, power over education policy was devolved to the Welsh Assembly Government, which immediately stopped publication of school performance tables. This event is useful for analysis as it sidesteps the two issues raised above. First, we have a ready-made control group in students in England as the education systems of the two countries were practically identical until that point. Second, there were no other major changes to the education system in Wales at the same time. Using a difference-in-difference analysis, Burgess et al (2013) find significant and robust evidence that this reform markedly reduced school effectiveness in Wales. The impact is sizeable, 0.23 of a (school-level) standard deviation, equivalent to 0.09 of a pupil-level standard deviation. In this study, the significant heterogeneity shows a much stronger effect on attainment of low-achieving pupils.

Two other recent studies have evaluated the introduction of school accountability in Portugal and the Netherlands. In Portugal, Reis et al (2015) show that the publication of school rankings make a significant difference to parents' choice of schools and to schools' enrolment. Koning and van der Weil (2012) show that once school quality scores are published (following campaigning by a newspaper), the lowest ranked schools raised performance substantially.

Much of the available evidence uses the introduction in the US of a mandatory school accountability system under the No Child Left Behind (NCLB) Act in 2002; this evidence is usefully summarised in Figlio and Loeb (2011). Dee and Jacob (2009) use the federal introduction of NCLB to estimate its effect on school performance, comparing states that had implemented a system of school accountability before NCLB as the control group. They found that NCLB had no impact on reading scores and a 0.15 pupil-level standard deviation impact on maths scores. Wong et al (2009) triangulate their evidence using different approaches, essentially by defining different control groups; they find a positive impact of the introduction of accountability in both approaches and both the fourth and eighth grades. Hanushek and Raymond (2005) actually use state-level accountability, pre-NCLB, and adopt a state-level fixed effects model to identify introduction of and find a positive effect of around 0.2 of a (state-level) standard deviation on test scores. Other studies exploit discontinuities in school accountability ratings and adopt a regression discontinuity approach. They show that schools receiving low ratings subsequently showed positive conditional impacts on pupil achievement gains, with strong and substantial effects (for example, Figlio and Rouse (2006), Rouse et al. (2013), Chiang (2007), Rockoff and Turner (2010)).

There is consensus that accountability measures raise performance, and typically more for low-performing pupils. There are fewer studies showing how this is achieved. Rouse et al (2013) show

that schools do change their teaching practices, for example spending more of the school day on instruction; Reback et al (2014) show that teachers work harder but also narrow the curriculum; and Craig et al (2015) show that school district administrators reinforce the effect of the ratings, rewarding high-performing schools by allocating them more funds.

To date there have been few studies of the long-run consequences of accountability; Deming et al (2013) find substantial positive long-run effects of accountability pressure on high-ability pupils, but find equally substantive negative effects for low-ability students.

#### *What about gaming, unintended consequences and cheating?*

Whilst one of the main issues in this literature is the impact of accountability on attainment, the other main focus is quantifying the strategies that schools undertake to game the system. These behaviour distortions can take many forms from a particular concentration of teacher time and effort, to outright cheating in the exams.

It has been generally established that schools will tend to focus their resources on whatever is tested: the subjects that are tested, the topics within subjects that are tested, the topics in which scores can be increased most easily, the school grades that are tested, and on the pupils who may be pivotal in reaching a threshold. Figlio and Loeb (2011) summarise all this evidence, and Rouse et al (2013) also review evidence on a range of responses by schools. Whether this focus on the things tested is a bad thing depends on the tests: this focus may in fact be what society wants and intends, and if the test is well-designed it may be wholly appropriate. Conversely, if the high-stakes tests are not well-designed, then the lack of broader knowledge and skills can be deleterious.

Boyd et al (2008) also show that high-stakes testing also altered the allocation of teachers to grades in New York. Relatedly, teachers face greater work pressure from accountability. Reback, Rockoff and Schwartz (2014) show that accountability pressure from NCLB lowers teachers' perceptions of job security and causes untenured teachers in high-stakes grades to work longer hours than their peers.

One way of gauging the degree of "teaching to the test" is to compare performance on high-stakes tests with that on low-stakes tests covering the same material. Jacob (2005) compared test score gains in maths in high stakes test to those on a comparable, but low-stakes, test; he showed that the gains for eighth graders were confirmed in the low-stakes tests, but that those for fourth grade pupils were not. Similarly, Figlio and Rouse (2006) find a smaller impact of accountability on low-stakes tests than on high-stakes tests.

Beyond focussing school resources on a subset of subjects, topics and pupils, researchers have documented other practices which, while not illegal, are certainly not what the accountability systems' designers would have had in mind as appropriate tactics. Figlio and Winicki (2005) show that schools change their lunch menus at the time of the tests, "substantially increasing calories in their menus on testing days"; Bokhari and Schneider (2009) show that pupils in schools under stronger accountability threat "are more likely to be diagnosed with Attention Deficit/Hyperactivity Disorder (ADHD) and consequently prescribed psychostimulant drugs"; and Anderson et al (2011) show that pupils in such schools have a higher chance of being obese, with one of the channels they cite being less exercise in school.

Finally, there is straightforward cheating on the test by teachers or school administrators. Jacob and Levitt (2003) show that the frequency of cheating appears to respond strongly to relatively minor changes in incentives, such as those implied by school accountability measures. Bertoni et al (2013) also implicitly detect cheating by noting that test scores in Italy are lower when external monitoring of tests takes place.

Of course, the existence of these inappropriate behaviours does not mean that accountability measures should be abandoned; the costs need to be weighed against the benefits. There are also implications for the design of the tests underlying the system and perhaps for the monitoring of the testing system.

#### *What is best content for accountability system?*

A subsidiary but important question is the nature of the performance data to be included in the accountability system. One key issue is whether it should be based on the level of student performance, or the per-pupil change in test score, also called value-added. The former is certainly of interest to policy makers and parents, but the latter is fairer to the schools in that it takes account of prior attainment. An excellent early analysis of the statistical issues involved for the implementation of NCLB is in Kane and Staiger (2001). Allen and Burgess (2011, 2013) use the long run of pupil-level data available in England to model different possibilities. There is a trade-off between functionality (whether the data actually helps parents to identify high-performing schools) and comprehensibility (whether the presentation is straightforward enough to make sense).

There is reasonable consensus that test-based school accountability measures raise pupil attainment, sometimes substantially. It is also clear that schools can be very sophisticated in designing strategies to game the metrics; some of these are arguably positive or benign (such as assigning strong teachers to the key classes) while others are strongly negative (risking pupil obesity, or cheating). Research priorities in this field include further exploration the long run impacts of schooling under strong accountability, and the impact of introducing accountability systems in a number of other countries, including Australia, Poland and Spain.

## **ii. School Autonomy**

There are two introductory questions to answer: why autonomy and autonomy from what? The basic concept behind the attraction of school autonomy is a simple one and a familiar one in economics: the people best placed to make 'production' decisions are those with most information, those closest to the process. This means teachers and school leaders; it follows that they should be able to implement those decisions, free from constraints from higher up the hierarchy. The argument is that school autonomy will therefore raise attainment, which is the empirical question I discuss below. The constraints placed on schools vary over time and countries. Typically, autonomy involves schools being able to determine all or some of: their own curriculum; hours and days of teaching; pedagogy and general approach; hiring and firing of teachers, and teachers' pay.

#### *Does school autonomy raise pupil attainment?*

The main cases of experimentation in school autonomy are Academies in England, Free Schools in Sweden (and more recently in England, too) and Charter schools in the US. Of these, the most secure evidence comes from recent quasi-experimental studies of the US case.

In England, there have been many new school ‘types’ tried over the past three decades, some introduced as offering more autonomy. One of these was Grant Maintained (GM) schools, studied by Clark (2009). These schools were able to opt out of the control of local government, and given control of teacher contracts and admissions. This reform is particularly susceptible to evaluation because the conversion to GM status required a vote of parents, and Clark is therefore able to do a like-for-like comparison between schools that just voted to convert and those that just voted not to (a regression discontinuity design). Attainment in the GM schools rose substantially, by about a quarter of a school-level standard deviation. Clark notes that GM schools were also more generously funded and he cannot rule out that this contributed to the rise in attainment (though see section 3.b.i). Looking at a more recent reform and schools with similar types of ‘freedoms’, Foundation schools, Allen (2013) repeats the regression discontinuity approach and finds little evidence for improved attainment having taken account of a rich set of pupil characteristics.

Academy schools are the latest type of school offered much greater freedoms in England. Machin and Veroit (2011) evaluate the impact on attainment of attending such schools, updated in Eyles and Machin (2015). This analysis provides the most robust evidence on Academies but was undertaken early in the Academy programme and relates to the schools set up under the Labour government before 2010, not those set up under later governments under very different criteria. There is no natural identification structure so the authors compare early converters to similar late converters. They find positive effects on attainment, of around 18% of a school-level standard deviation. The effect appears to be stronger the greater the increase in autonomy, either because of simply more time as an academy, or because of switching from a school with the lowest initial degree of autonomy.

The establishment of Free Schools followed a 1992 reform in Sweden, allowing schools with great operational and strategic autonomy to compete with state schools for pupils and funding. The evidence is mixed and studies vary in finding no, small or large effects. Because of the data they assemble, Bohlmark and Lindahl (2007, 2008) is possibly the most persuasive approach (see Allen, 2010, for a useful summary). They find a small positive impact of municipality-level free school growth on municipality-level academic performance for 16 year olds, though this dissipates by age 19. Again, the larger positive effect is on higher ability pupils. Using siblings data, Bohlmark and Lindahl show that this improved performance is due in part to the greater effectiveness of the free schools, but that competitive threat played a bigger role. Other studies include Ahlin (2003), Björklund et al. (2005), and Sandström and Bergström (2005).

However, it is not clear that these results can be attributed to increased school autonomy itself, as that reform was introduced alongside others (as documented in Björklund et al. (2005)) and the increased autonomy is confounded with increased parental choice.

Turning to the US, Charter schools have much more autonomy than regular state schools, and there are now many thousands of such schools since the first in 1992. Whether this improves the attainment of pupils is a controversial question. A comprehensive study of Charter schools across 16 states is published in CREDO (2009). The authors use matching techniques (creating a “virtual twin” for each Charter school pupil based on demographics and poverty status) to compare the outcomes for pupils in Charter schools and regular schools. They find that about half the Charter schools do no better for their pupils, 17% of Charter schools perform better and the remaining 37% perform worse



than the comparator regular school. Epple, Romano and Zimmer (2015) provide a wide ranging survey of the Charter movement as it approaches its 25<sup>th</sup> anniversary. They concur that the impact of Charters on pupil performance is very variable: some produce dramatically higher performance, but most are about the same, a bit worse or a bit higher.

More recently, an important set of studies has used an experimental approach to isolate the role played by Charter schools. The key is that some Charter schools that are over-subscribed use random chance, lotteries, to determine which of the applicants are given a place. This means that among the set of applicants to a school, Charter attendance is exogenous. While these studies are small-scale, this may be appropriate: charter schools are very heterogeneous, so charter school treatment effects are also likely to be heterogeneous. On the other hand, the very fact that the schools are over-subscribed suggests that they are likely to be at the higher end of the outcome distribution rather than representative. Hoxby and Rockoff (2004) is an early example, studying the Chicago Public School system. They found that pupils who win the lottery and enroll in lower elementary grades have higher subsequent attainment than students who are lotteried-out, but find no effect for pupils joining in the upper elementary grades. Using the same methodology, Hoxby and Murarka (2009) find positive and significant effects of charter school attendance in New York City charter schools, with the impact increasing for each additional year spent at a charter school. Sticking with New York, Dobbie and Fryer (2009) focus on the charter schools associated with the Harlem Children's Zone (HCZ). They too find significant increases in attainment in both maths and English, for pupils of all abilities. Likewise, in Boston, Abdulkadiroğlu et al (2009) also using assignment lotteries find large and significant gains in attainment for lottery winners in both middle school and high school. In a related paper, Angrist et al (2010) focus on a school belonging to the largest charter group, the Knowledge is Power Program (KIPP), a strong advocate of the 'No Excuses' approach to public education. This means they have a long school day and a long school year, highly selective teacher hiring, strict rules for behaviour and a focus on traditional learning. The lottery methodology shows huge gains in attainment: 0.35 standard deviations in maths and 0.12 standard deviations in reading for each year spent at KIPP Lynn.

The comparison of winners and losers within lotteries only provides a causal effect for charter school applicants, who might be very different to non-applicants. Abdulkadiroğlu et al (2014) study a case in which regular public schools are taken over by charter schools so the pupils are not as selected a group as lottery applicants. They confirm substantial test score gains for these pupils too, suggesting that there is something in the schools that substantially and significantly raises attainment. Discovering what that something is, is clearly a question of the first importance. Dobbie and Fryer (2013) and Angrist et al (2013) make a start on explaining this, and Fryer (2014) reports the impact of attempting to introduce those practices into regular public schools.

#### *What aspects of autonomy are crucial?*

Given that autonomy matters, what sorts of 'freedoms' matter for attainment? The main evidence on this comes from Abdulkadiroğlu et al (2009) who are able to compare regular charter schools with Boston Pilot schools: "These schools have some of the independence of charter schools, but operate within the school district, face little risk of closure, and are covered by many of same collective bargaining provisions as traditional public schools". The same lottery methodology that found large effects for the regular charter schools found small and insignificant effects for the Pilot schools.



Reviewing this evidence, the results from Sweden are mixed, but are somewhat difficult to evaluate as the increased school autonomy coincided with other major changes to the school system. In England, this was not the case and the increased autonomy was legislated within a settled system of parental choice. However, the way that academies were introduced means that the identification of a robust causal effect is difficult. The best evidence to date is from the lottery-based studies of US charter schools. There are two caveats here, however. First, the effect is identified among applicants to charter schools, so it is not clear how the findings will carry over to the wider population of pupils, although as discussed Abdulkadiroğlu et al (2014) do find similar non-experimental results. Second, as is widely noted, charter schools are very heterogeneous, and lottery-based results necessarily imply that these are very popular schools.

The stand-out results are for 'No Excuses' schools, both the HCZ schools and the KIPP schools, which bring very substantial causal impacts on attainment. Dobbie and Fryer (2011, 2013) make a start on understanding what it is about those schools that works, but this is surely a key endeavour for future research. We cannot necessarily expect similar results for all charter schools, and so from a systemic perspective, the rules on monitoring performance and awarding and removing autonomous status are likely to be very important.

### **iii. School leadership**

The nature of school leadership varies directly with the level of school autonomy. In a system of tight central control, school leaders are middle managers, line-managing teachers and implementing policies from the centre. The commitment to a centrally run system means that the values and beliefs of any one headteacher should not impact on the education outcomes for her pupils. In a decentralised model, the system needs good and great leaders, since there is much less central direction on how to run a school.

Policy-makers seem to set great store by the idea of transformational school leadership. The changing of headteachers or principals is taken very seriously as a school improvement policy. There are plenty of stories of how charismatic headteachers have turned around failing schools into beacons of achievement. Grissom and Loeb (2011) and Branch, Hanushek and Rivkin (2012) both document prior research, much of it qualitative, associating excellent school leadership with positive school outcomes.

But this is a hard arena in which to do quantitative research, and very hard to robustly identify causal effects. Changes in school leadership are rarely exogenous, and policy-makers are unlikely to be keen on randomising high- and low-quality principals across schools. Typical changes in school leadership may occur when a school is under-performing, for example, making it difficult to disentangle other compensatory responses as well as mean-reversion from the leadership change.

While there is now a small literature on what effective schools do (discussed below in section 4.b.ii, but for example Dobbie and Fryer, 2013; Angrist, Pathak and Walters, 2013), this has yet to be linked across to research on what effective or indeed transformative principals do.

Recent research on school leadership can be split roughly into papers attempting to measure the effectiveness of principals; papers looking at the career path or turnover history of principals and an

association with school effectiveness; and a set of papers on what principals do, or the management of schools.

### *The effectiveness of principals*

Grissom, Kalogrides and Loeb (2015) set the scene by highlighting some of the problematic issues involved in using test score data to estimate principal effectiveness, and setting out a series of models to capture different ways of thinking about what principals do. This useful foundational work shows that the choice of model matters as different approaches can yield very different results, ranging from 18% of a standard deviation to 5% using the same data. They also compare the results with other school outcomes including administrative evaluations, although this yields some puzzling results.

Going down the same track, Coelli and Green (2012) can identify principal effects using a dataset in which principals were rotated amongst schools by districts (using Rivkin et al's (2005) approach) and assuming constant effectiveness within-school. They find substantial effects, with one standard deviation of principal effectiveness implying 33% of a school-level standard deviation in graduation rates. Branch, Hanushek, and Rivkin (2012) also find large variation in principal effects; they also show greater variation for schools in disadvantaged neighbourhoods. They also note that schools with ineffective principals are estimated to have higher than average teacher turnover, and that this might be a mechanism through which low effectiveness affects school performance.

A different approach to estimating the importance of principals is taken by Lavy (2008). He exploits an experiment in Israel giving a very large pay rise (50%) to school principals. He finds significant but modest effects on attainment, probably insufficient to justify an expensive treatment.

### *Principals' careers and school effectiveness*

Béteille, Kalogrides and Loeb (2012) provide an overview of principals' career paths, and document substantial turnover rates: more than a 20% annual separation rate for principals. A typical path is to use a low-attaining, disadvantaged school as a stepping-stone to a more preferred school. They show that high principal turnover rates are generally associated with lower school performance: "The departure of a principal is associated with higher teacher turnover rates and lower student achievement gains" (p. 905) and that this negative relationship is stronger in more disadvantaged neighbourhoods. Because of the interlocking issues of principal's desired career paths, endogenous principal mobility and school performance, robustly attributing causality to this is likely to be difficult. All of this research paints a picture of schools in disadvantaged neighbourhoods underperforming and struggling to hire principals, struggling to keep them, and struggling to hire effective principals.

Clark, Martorelli and Rockoff (2009) try to characterise what effective principals look like in terms of observable characteristics. As with teachers, they find little evidence for a role for the principal's own academic record nor for their pre-principal experience. Again as with teachers, there is evidence of learning through experience, particularly steep in the early years. By contrast, Grissom and Loeb (2011) try to characterise what effective principals do, combining survey responses with administrative data. They isolated five skill categories – instruction management, organisation management, administration, internal relations and external relations. The results suggest that only organisation and management skills are consistently associated with school performance across different outcome measures and sources.

### *School management*

This stress on organisational management ties in well with the findings of Loeb, Horng and Klaskik (2010) who document principals' time use and relate that to school outcomes including student attainment but also teacher and parental assessments. They show that time spent on organisational management is associated with positive outcomes.

Bloom et al (2014) collect data on management practices in nearly 2000 schools (educating 15-year-olds) in eight countries. They show that higher management quality is strongly associated with better educational outcomes, and in particular that autonomous public schools have significantly higher management scores than regular government schools and private schools. They highlight the role of the principal, assigning a high fraction of the effect to differences in leadership and governance. Consistent with the evidence above on the connection between ineffective principals and high staff turnover, Bloom et al (2014) note that schools are generally weak in people management practices.

A big part of what principals need to do well is the selection of teachers. Jacob and Lefgren (2008) show that principals can generally identify teachers at the extremes of the distribution of effectiveness, but are much less able to distinguish teachers in the middle of the distribution. In a companion piece, Jacob (2010) shows that principals do weight measures of teacher effectiveness when firing probationary teachers, but only alongside demographic factors.

## **5. Data for education research**

Most research in the economics of education is empirical; and certainly most of the work reviewed here is. Empirical work needs data, and data can come from many different sources, including being generated by the researcher. This field is notable for using a variety of data types and sources, including emerging data sources that are giving rise to a lot of the leading research.

### **a) Survey Data**

But I start with more traditional data sources. Many of the classic studies in the economics of education were undertaken using survey data, small or large scale. These include for example Labour Force Surveys (LFS) typical in European countries, sampling the adult population and including information on employment and earnings, demographics and, crucially, educational background. In the US, the equivalent is the Current Population Survey (CPS). These are the basis for many of the early studies of the rate of return to schooling discussed above, in which (put simply) differences in earnings are compared to differences in education and work experience. Just to take one example, the fundamental work of Katz and others to establish the rising rate of return to education uses a long series of CPS datasets, each one a cross-sectional sample of the US adult population.

The advantages of this sort of datasets are that they are typically large scale and representative of the population; they are relatively easy for researchers to access and generally well documented and used. Furthermore, they have frequently been running for a long time on a very similar basis, allowing secular changes to be mapped.

## **b) Birth cohort data**

A second type of dataset widely used in the economics of education is a birth cohort. This defines a sample by a particular birth date and then tracks that sample over the following years. For example, the National Child Development Sample (NCDS) in the UK sampled everyone born in a particular week in 1958, and continues to track them today. These data typically offer rich characterisations of childhood, adolescence and adulthood. As such, they are ideal for studying education and the development of human capital in “early years”, through school and then university and through the transition into the labour market. The UK has a set of birth cohorts, covering much of the post-war period, reflecting the implicit and not-always-fulfilled ambition of starting a new one every decade or so (the 1948 cohort, the NCDS in 1958, the BCS in 1970, ALSPAC in the late 1980s, and the MCS in 2000. In the US, the National Longitudinal Study of Youth (NLSY) series performs a similar role; for example the NLSY79 continues to follow a cohort of young persons, aged between 14 and 21 in 1979.

The advantages of this type of dataset include the great detail on early childhood including items such as parenting styles and childcare. Also the very long-running longitudinal element means that later outcomes in life can be related back to childhood circumstances; this is clearly central to studying the role of education and human capital in social mobility. More recent birth cohorts have started to also collect biomedical data, which will increasingly allow for analysis of genetic influences on educational outcomes.

## **c) International comparative education data**

It is difficult to overstate the impact of the growing prominence of internationally comparative educational testing programmes. The most famous is the OECD’s Programme for International Student Assessment (PISA), but TIMSS (Trends in International Mathematics and Science Study) and PIRLS (Progress in International Reading Literacy Study) are also widely cited; a relatively new OECD survey relates to adult skills, PIAAC (Programme for the International Assessment of Adult Competencies). These apply standardised tests, questionnaires and measurements across countries, and again the repetition over time is also very useful. Data typically includes test score outcomes, pupil demographics, parental background measures and school characteristics.

The main advantage of this type of dataset for research is that they enable research into the role of national institutions and market regulations in determining educational outcomes. This is typically

very difficult using national datasets as institutions are the same for all and if they change, change for everyone at once. See for example the work summarised in Hanushek and Woessmann (2011).

It should be noted though that these data are important beyond research. PISA in particular has come to play a very significant role in national policy-making in education. The internationally comparative benchmarking of achievement has focussed attention on countries and regions that persistently do very well (Finland, Shanghai, others) and others that persistently perform below expectations. There is good evidence that PISA rankings have driven national education policy (for example, PISA results in the year 2000 led Germany to reform its educational system, see Grek, 2009).

#### **d) Administrative or register data**

Increasingly, researchers in this field are using datasets that have been collected for another purpose, known as administrative or register data. These might be the records of an individual school; more usually, they are national or regional data requirements on schools. For example, the National Pupil Database (NPD) in England has a record of every pupil in every state school in every school year. Each record contains pupil demographic data, all previous test score data in complete detail (there are over one hundred data items for the high-stakes age 16 tests alone), current school attended, home address and more. It is linkable through time using the pupil's ID creating a complete record of each pupil's school time. These data have been widely used. Other examples include linked register data from Norway on many aspects of up-bringing and schooling; very detailed longitudinal pupil data from North Carolina; teacher data from New York City schools and so on. Such data is very flexible and can be used to create datasets that are longitudinal for pupils – follow a particular pupil through successive school years – or longitudinal for schools – track a school's performance for successive cohorts of 16 year olds. One major research focus that has made enormous progress through administrative data is analysis of teacher effectiveness. Estimating teacher effectiveness requires a link between teacher and pupil (from schools' class lists), the pupils' test scores, and ideally pupils' prior attainment and demographics. Here the census nature of administrative data is important – having only a few pupils per teacher would produce very noisy measures of effectiveness.

The advantages of administrative data include that it covers everyone and is not just a sample, high data quality, and the capacity to produce a longitudinal perspective.

#### **e) Data from field experiments**

A great deal of empirical work in the economics of education is about policy evaluation, aiming to estimate a causal effect of some intervention. One way of identifying causality that is becoming more popular and feasible in this field is through a field experiment based on a Randomised Control Trial (RCT) methodology. Examples include studies of class size reduction, teacher incentives, and

teaching styles, see the survey in Bouguen and Gurgand (2012). Field experiments typically produce experimental data including baseline test scores, experimental outcomes and pupil and school characteristics. Field experiments in education have been popular in development economics, pioneered by Duflo, Glennerster, Kremer and others through JPAL; in the US, some leading recent examples include work by List and co-authors in Chicago and Fryer in Harvard.

The chief advantage of field experiment data is that it can identify a causal relationship, that is, properly evaluate an educational intervention, which is often difficult to do from observational data.

## **f) Other data**

These are the main types of data used in the economics of education but there are others. Researchers estimating the rate of return to education were very keen on utilising data on monozygotic twins in the hope of fully differencing out genetic factors from earnings. These datasets are necessarily few, but reasonable sample sizes were generated (Ashenfelter and Rouse, 1998). As economists attempt to take greater account of psychological approaches, datasets including measures of personality traits and non-cognitive skills are useful. For example, the NLSY79 includes measures of locus of control and self-worth (used by Heckman, Stixrud and Urzua, 2006); Mueller and Plug (2006) use the Wisconsin Longitudinal Study of over 10,000 Wisconsin students graduating in 1957 to study the impact of the “big 5” personality traits on earnings; see also Heckman and Kautz (2012). A very different type of dataset comes from combining standard measures of teacher effectiveness with data on what teachers actually do in class from the Cincinnati teacher evaluation system (Tyler et al 2010); they show that various classroom practices predicted pupil attainment.

## **6. Education Policy and Future Research**

This section summarises the review under two headings: the most promising areas for policy development in education, and the most important questions for future economics research in education. First, though I briefly discuss research taking different approaches to viewing education as a whole.

### **a) Putting it all together: Attainment, skills, inequality and sorting**

Earlier sections reviewed the evidence on the formation of human capital and the demand for education, the evidence on the supply of education and on the market structure and rules. Together these determine the equilibrium and the overall distributions of attainment, skills and earnings.

Researchers and policy-makers are interested in many different aspects of this overall outcome. Obviously the average level of skills is important, but also the types of skills gained. Sub-group analysis is also typically of great concern, for example differences in educational attainment by gender, by race and ethnicity, and by socio-economic status. At a more aggregated level, policy-makers rightly care about international differences in cognitive skills and attainment, as these are key components of national productivity and hence national income. All of these quantities are evolving over time, so another set of issues are, for example, changes in black-white attainment gaps, changes in the relative importance of maths skills versus language skills, changes in the return to schooling, and so on.

There are two facets to all of this: documenting these quantities, and understanding them. This survey is not a compendium of facts and figures on European education, and so will not discuss the former. A very full set of (consistent) figures on education in the OECD can be found here: <http://www.oecd.org/edu/Education-at-a-Glance-2014.pdf>; and a very helpful guide to European education institutions: <https://webgate.ec.europa.eu/fpfis/mwikis/eurydice/index.php/Countries> .

All the analysis reviewed here is about understanding the facts on education and human capital, typically focussing on individual components of the system. There are different approaches to trying to put together an analysis of the big picture. First, there is a literature looking at the general equilibrium of the school and housing market. Early papers included the work by Epple and Romano (2003) and Nechyba (2003), for example. More recently there have been a set of papers looking at the same topic using detailed micro data from California from Bayer et al (2007). They compute demand for housing by neighbourhood and the implied demand for schooling, deduce the degree of competitiveness facing schools and consider the feedback effect on school performance. This is a very useful framework for thinking about school demand in a school choice system.

A second approach to the big picture is to take the perspective of an individual pupil and track her progress from birth through all the stages of education and out into the labour market. This is particularly helpful in trying to understand at what point in life educational inequalities open up and evolve. Examples here include the set of papers by Fryer and Levitt exploring the genesis of the black-white test score gap in the US, Fryer and Levitt (2004, 2006, 2013). Similarly, Gregg and Goodman (2010) report on a project analysing the socio-economic attainment gap from birth to the end of compulsory schooling.

Thirdly, a cross-country approach is one way of comparing the summative outcome of all of a country's educational institutions and regulations together. Clearly, the outcome also depends on the financial resources of both people and government but this can be controlled for. The obvious example here is the substantial programme of work analysing the PISA, TIMSS and other datasets on cognitive skills across countries and across time. One major contributor to this has been Woessmann, see for example, Woessmann (2003), Hanushek and Woessmann (2011), and Hanushek, Link and Woessmann (2013).



## b) Policy

European countries differ in very many ways, including in their education systems (see <https://webgate.ec.europa.eu/fpfis/mwikis/eurydice/index.php/Countries>), both in terms of the overall system and in the detail. So this review cannot get into the ways in which specific policies might work in each country. And there is obviously already a lot of policy intervention in education in Europe, including in the fields noted below, so this should be read as a discussion of balance and emphasis, rather than identifying new fields.

As has been evident throughout this review, much of the research in the economics of education is about policy. This includes evaluation of specific interventions, cross-country analysis of systems and much else, so there is a lot of policy writing available. However, it should be clear that we are nowhere near, for example, an engineering level of precision in policy discussion. An engineer could say “if you want the bridge  $x\%$  longer it will need  $y\%$  more concrete and be subject to  $z\%$  more stress”. As education economists we cannot equivalently say: “if you reduce the schools budget by  $x\%$  by raising class sizes and put that money into  $y\%$  more child care, then end-of-schooling attainment will rise by  $z\%$ ”. This caution needs to be borne in mind.

In terms of direct policy proposals, here are four examples. Woessmann (2014) highlights a number of factors but puts first the importance of effective school systems: “the institutional setup of school systems is a crucial aspect for their efficiency” (p. 26); this recommendation derives from his programme of cross-country analysis. Under this heading, he includes effective governance, accountability and autonomy. Fryer (2014) focusses in on schools and argues for a set of five practices of effective schools. These practices were identified from close analysis of high-performing charter schools, and then introduced to regular public schools in a randomised field experiment; again this is the result of an (on-going) programme of work. The five practices are (p. 1355/6): “increased [teaching] time, better human capital [i.e. more effective teachers and administrators], more student-level differentiation [i.e. supplemental tutoring], frequent use of data to alter the scope and sequence of classroom instruction, and a culture of high expectations”. Cassen, McNally and Vignoles (2015) advance policy proposals around pre-school, parenting and specifically on numeracy and literacy among other things (on the basis of a wider evidence review than economics, though with a focus on the UK). The importance of high quality pre-school and parenting programmes is emphasised. A new and valuable resource is provided by the Education Endowment Foundation, which they describe as a ‘Teaching Learning Toolkit’ for schools teaching 5-16 year olds, <https://educationendowmentfoundation.org.uk/toolkit/>. It uses only evidence from randomised field experiments and grades each potential intervention in terms of effectiveness, cost and the robustness of the evidence.

On the basis of the evidence reviewed here, the areas I believe to be the most promising for policy development are as follows (in no particular order):

### i. Improving teacher effectiveness

We know that teacher effectiveness is extremely important for raising attainment, perhaps in fact the most important factor in schools. The effect size of improving teacher effectiveness is so



substantial that any policy that has even a minor impact on effectiveness is a great prize. We know that it matters and makes a substantial difference for pupils of all ages, well beyond the early years. The two main policy drivers are raising the effectiveness of individual teachers and changing the selection of teachers to increase the average. The first is about changing initial teacher training and about improving on-going development, continuing professional development. The second is about changing the contract structure for the teaching profession to place more of an emphasis on hiring, identifying and retaining, and rewarding the most effective teachers. Neither of these is easy, and for both there are a number of missing pieces of knowledge before we can be confident of success. And in both cases, the details of the systems and structures of individual countries will matter.

## **ii. Ensuring high quality 'early years' opportunities**

The early years are very important for the formation of children's human capital, encompassing both cognitive and non-cognitive skills. The child will spend part of her time with her family, and part in child care settings, the fractions in each depending on country and family. This means that both parenting styles and child care quality matter. While there is always more to learn, there is now some good evidence on what these look like and how they can be delivered. Unsurprisingly, small-scale intensive 'model' interventions are easier to design and are more effective than national-scale "as is" programmes, and some of the policy focus has to be on checking and certifying quality levels for child care provision. One of the more promising parenting interventions is the Nurse-Family Partnership programmes. There is obviously extensive child care provision already in place over Europe (see Del Boca, 2015, for an excellent survey of existing child care arrangements in Europe, varying between mostly private, a mix of public and private, and universal public child care). This also links across to parental leave policies (see Carneiro et al, 2015, for analysis of parental leave reform and child development). While the most visible measure of child care quality is staff/child ratio, this appears to have little relationship with outcomes or parental demand. It will be necessary to develop more content-based certificates of quality for providers.

## **iii. Designing the optimal school market structure**

This is the industrial organisation approach to schools, determining the market rules and the market incentives. There are a number of inter-locking factors that create an effective school system. Accountability matters for schools' performance, even relatively low-stakes accountability (Reis et al 2015). Accountability requires some common and consistent form of assessment, typically centralised external exit exams. Accountability also makes more sense if schools have autonomy in their operations. All of these factors have been robustly shown to raise school performance and pupil attainment. Such a policy also leads to a focus on two other things. First, if schools are held to account on a specific assessment basis, then this is undoubtedly what schools will focus on. So governments need to take care that the assessment is well designed, and that it does indeed test the skills that society wants pupils to have. Teaching to the test is detrimental if a test is badly designed. Second, publication of rankings showing schools are better performing can in principle increase socio-economic sorting of pupils, though the evidence on this is mixed. Whether it does so depends on the admissions process to schools and whether this is manipulable by parents.

On the basis of the evidence, the following areas are not proven, or are shown to have minimal impact.

#### **iv. Resources for schools and class-size**

The evidence here is mixed and there are a few robust studies showing positive effects of higher resources for schools and lower class sizes. Yet these results are not the norm, and there are more studies showing little or no effect of class-size on outcomes. A better research agenda is to determine what the criteria are which determine that some interventions raise attainment and others do not. The verdict here is that reducing class size does not have a widely proven effect on outcomes, whilst it is obviously expensive.

#### **v. Classroom IT**

Adding more IT hardware to classrooms seems to be a political favourite. But there is little strong evidence that this raises attainment. There are stories of what brilliant teachers can do with IT, but brilliant teachers raise skills without IT as well. This does not seem a cost-effective education strategy.

### **c) Future research priorities in the economics of education**

There are many highly promising and pressing issues for further research. Given the key importance of education for economic growth, social mobility and inequality, the potential rewards for research in the economics of education are huge. Some of these issues are listed below; of course, there will be other very promising ideas that currently exist only in the head of a researcher. This set of topics seems to me to be the next key steps in the areas surveyed in this review. No doubt, these are being thought about and worked on already, but there seems to be great scope for more. In particular, the single area that could most valuably be expanded is the use of RCTs in education (see Levitt and List, 2009, for a strong argument for this). While we have a huge amount of observational studies, robust evidence of causation is much harder to find, and there is great scope for harvesting low hanging fruit here.

It is also worth noting that spending on educational research is particularly low relative to its potential impact on national income and well-being; this is a classic public good problem as no-one other than governments will commit money to this. For example, in very approximate figures, the UK spends about £90bn per year on education and perhaps about £30m per year on education research. Given the EU's 2020 strategy target of 3% of output on R&D, this is far short of that target.

#### **i. Formation of human capital:**

- We need to have a better understanding of, and measurement system for, non-cognitive skills.

- We need to know much more about what aspects of parenting and pre-school raise cognitive and non-cognitive skills.
- We need to better understand the nature of the interactions between public and private investment, does family investment substitute for public investment or are they complementary?
- What are the sources of the different gender gaps in attainment and skills?
- There is also little causal evidence on how cognitive and non-cognitive skills are developed in school.
- The roles of conscientiousness in particular, and pupil motivation, effort and 'focus' need much more work

#### **ii. Demand for education:**

- We need greater development of models allowing for heterogeneity in the impact of schooling on earnings and other outcomes. This will involve developing estimation techniques and datasets.
- We need more studies using earnings data from complete actual cohorts throughout people's careers to estimate returns.
- We need a better understanding of how the cognitive and non-cognitive skills discussed above map into workplace skills (such as knowledge skills, problem-solving skills, critical thinking skills, and social skills) and how they are valued. Acemoglu and Autor (2012) launch a discussion of this analysis.
- We should further explore the role of education as a factor in inter-generational transmission of life-chances.

#### **iii. School resources:**

- While more studies are needed on class size changes, there needs to be a recognition that class size differences may not be simple and symmetric, and an exploration of much more asymmetric changes akin to creating resources for differentiated tutoring.
- We need to understand how schools decide how to spend their budgets.

#### **iv. Teachers and teaching:**

- It would be useful to know the distribution of teacher effectiveness in other contexts and more countries.
- We need to know more about the persistence of effectiveness across time, classes and contexts – is an effective teacher effective in all circumstances?
- The study of teacher impacts on other, non-test, outcomes and on non-cognitive skills is only just beginning, but is key as we have seen the importance of those attributes to later success.
- We don't really know what effective teachers do that makes them effective.
- Nor do we know how to structure, implement and incentivise take-up of effective initial teacher training and continuing professional development.
- We need to learn a lot about how to raise teacher effectiveness and now to (re)design teacher contracts to improve selection.

- Designing a pre-hire screen that is usefully selective on effectiveness would be very valuable.
- We are not yet close to reaching a consensus on when performance pay is effective and when it is not.

#### **v. School assignment mechanisms:**

- Until we can make all schools very effective and desirable, we need to understand more about the impact of different assignment mechanisms (elite/tracked; choice; neighbourhood) on the education we care about
- What is the best age to start tracking (if at all)? How can we quantify the trade-off that determines whether it is best done across schools or within-school?
- We need much more work estimating parental preferences for school characteristics, how these vary (if they do) by parental characteristics, and what the implications are for market structure, performance and equity.
- If we have a choice system, what is the impact of different mechanisms to ration places in over-subscribed schools?
- What are the best practical and fair ways to elicit true preferences in a choice system?

#### **vi. School accountability:**

- What are the long run impacts of schooling under strong accountability?
- Evaluations of the introduction of accountability systems into countries (such as Australia, Poland and Spain) will be fascinating and very informative.

#### **vii. School autonomy and the practices of effective schools:**

- What do autonomous schools do differently from non-autonomous ones? This links to the question above about how schools spend their budgets.
- What are the effects of introducing more widely the practices isolated and implemented by Fryer? Can the same or different practices be identified in other contexts and other countries?
- What are the systemic implications of a schooling system in which most or all schools are autonomous?

#### **viii. Universities:**

- We need a much deeper and more granular understanding of the heterogeneity in rates of return to different degrees, across subjects and institutions.
- We need to find new ways to estimate causal rates of return when neither random allocation nor score-based entry are available.
- We need to better understand pupils' decision whether or not to attend university. For those who would benefit, we need to evaluate a portfolio of methods to promote access for pupils from poorer backgrounds.
- There is a huge amount of very valuable research waiting to be done on the 'supply side' of higher education. In particular, we need research into the institutional responses to the higher education finance systems in different countries.

#### ix. Vocational education:

- This is a relatively neglected area by economists, and also one that merges into labour economics, so there are many questions needing more research.
- We need to evaluate the best ways to deliver vocational education: in tracked educational systems, alongside 'academic' subjects in regular comprehensive schools, in separate technical colleges, or through some new channel.
- Is it possible to design vocational qualifications that will be valued in the labour market? How should these relate to apprenticeships?
- How should we design incentives for schools, colleges and pupils to encourage the assignment of the "right" (most appropriate) qualifications to each pupil?

#### x. Data:

- Survey data will remain a very useful resource in economics of education research, and the longer the run of consistent data, the more valuable; these datasets need to be maintained.
- Data from international comparative tests will also continue to strongly inform research (and to drive policy).
- A lot of the exciting new progress in this field is coming from newer data types, and these need to be strongly supported and expanded:
  - data on characteristics such as psychological traits, 'soft' skills, and genetic traits;
  - data from administrative records and field experiments;
  - the combination of field experiments and administrative records is very powerful. It can deliver robust evidence of causation on 'hard' policy-relevant outcomes, can (cheaply) follow participants for long periods, whilst at the same time requiring minimal input from participants and schools, and essentially eliminating attrition from the follow-up.

## References

- Aakvik, A., Heckman, J. and Vytlacil, E. (2005) 'Estimating treatment effects for discrete outcomes when responses to treatment vary: An application to Norwegian vocational rehabilitation programs.' *Journal of Econometrics*, vol. 125 (1–2), pp. 15–51.
- Aakvik, A., Salvanes, K. and Vaage, K. (2010) 'Measuring heterogeneity in the returns to education using an education reform.' *European Economic Review*, vol. 54, pp. 483–500.
- Aaronson, D., Barrow, L. and Sander, W. (2007) 'Teachers and student achievement in the Chicago public high schools.' *Journal of Labor Economics*, vol. 25, pp. 95–136.
- Abdulkadiroglu A., J. Angrist, P. Hull and P. Pathak (2014) Charters Without Lotteries: Testing Takeovers in New Orleans and Boston, National Bureau of Economic Research Working Paper 20792.
- Abdulkadiroglu, A. and Sonmez, T. (2003) 'School choice: a mechanism design approach.' *American Economic Review*, vol. 93(3), pp. 729–47.
- Abdulkadiroglu, A., Angrist, J. and Pathak, P. (2012) 'The Elite Illusion: Achievement Effects at Boston and New York Exam Schools.' IZA DP no. 6790.
- Abdulkadiroglu, A., Angrist, J., Dynarski, S., Kane, T. and P. Pathak, P. (2009) 'Accountability and Flexibility in Public Schools: Evidence From Boston's Charters and Pilots.' National Bureau of Economic Research Working Paper 15549.
- Abdulkadiroglu, A., Angrist, J., Dynarski, S., Kane, T. and Pathak, P. (2011) 'Accountability and Flexibility in Public Schools: Evidence from Boston's Charters and Pilots.' *The Quarterly Journal of Economics*, vol. 126 (2), pp. 699–748.
- Abdulkadiroglu, A., Pathak, P.A. and Roth, A.E. (2009) 'Strategy-proofness vs. efficiency in matching with indifferences: redesigning the NYC high school match.' *American Economic Review*, vol. 99(5),
- Abdulkadiroglu, A., Pathak, P.A., Roth, A.E. and Sonmez, T. (2005) 'The Boston public school match.' *American Economic Review*, vol. 95(2), pp. 368–71.
- Abramovsky L., Harrison, R. and Simpson, H. (2007) 'University research and the location of business R&D.' *Economic Journal*, vol.117, pp. C114-C141.
- Acemoglu, D. and Autor, D. (2012) 'What Does Human Capital Do? A Review of Goldin and Katz's The Race between Education and Technology.' NBER Working Paper No. 17820.
- Aghion, P., Dewatripont, M., Hoxby, C., Mas-Colell, A. and Sapir, A. (2010) 'The Governance and Performance of Universities.' *Economic Policy*, vol. 25 , pp.7–59.
- Ahlin, Å. (2003) 'Does School Competition Matter? Effects of a Large-scale School Choice Reform on Student Performance.' Department of Economics, Uppsala University Working Paper No.2.
- Albrecht, J., van den Berg, G. and Vroman, S. (2004) 'The knowledge lift: The Swedish adult education program that aimed to eliminate low worker skill levels ' IFAU WORKING PAPER 2004:17
- Allen, J., Pianta, R., Gregory, A., Mikami, A. and Lun, J. (2011) 'An interaction-based approach to enhancing secondary school instruction and student achievement' *Science*. 333(6045):1034-37
- Allen, R. (2010) 'Replicating Swedish 'free school' reforms in England.' *Research in Public Policy*, vol. 10. <http://www.bristol.ac.uk/media-library/sites/cmpo/migrated/documents/allen10.pdf>
- Allen, R. (2013) 'Measuring foundation school effectiveness using English administrative data, survey data and a regression discontinuity design.' *Education Economics*, vol. 21(5), pp.431-446.
- Allen, R. and Burgess, S. (2011) 'Can school league tables help parents choose schools?' *Fiscal Studies*, vol.32(2), pp.245-261.
- Allen, R. and Burgess, S. (2013) 'Evaluating the provision of school performance information for school choice.' *Economics of Education Review*, vol.34, pp. 175-190.

- Allen, R., Burgess, S. and May, J. (2012) 'The teacher labour market, teacher turnover and disadvantaged schools: new evidence for England.' CMPO DP 12/294, CMPO, University of Bristol.
- Allen, R., Burgess, S., McKenna, L. and Rasul, I. (2012) 'Understanding school financial decisions.' Department for Education Research Report DFE-RR183.
- Almlund, M., Duckworth, A., Heckman, J. and Kautz, T. (2011) 'Personality psychology and economics.' In E. Hanushek, S. Machin, and L. Woessmann (eds.), *Handbook of the Economics of Education*, vol. 4. Amsterdam: North-Holland.
- Anderson, P., Butcher, K. and Schanzenbach, D.M. (2011) 'Adequate (or Adipose?) Yearly Progress: Assessing the Effect of "No Child Left Behind" on Children's Obesity.' NBER Working Paper No. 16873.
- Andersson, R., Quigley, J. and Wilhelmsson, M. (2004) 'University decentralization as regional policy: the Swedish experiment.' *Journal of Economic Geography*, vol. 4, pp.371-388.
- Andersson, R., Quigley, J. and Wilhelmsson, M. (2009) 'Urbanization, productivity and innovation: evidence from investment in higher education.' *Journal of Urban Economics*, vol. 66, pp.2-15.
- Angrist, J. and Guryan, J. (2008) 'Does Teacher Testing Raise Teacher Quality? Evidence from State Certification Requirements' *Economics of Education Review* vol. 27(5) pp. 483–503
- Angrist, J. and Lavy, V. (1999) 'Using Maimonides' Rule To Estimate the Effect of Class Size on Scholastic Achievement' *The Quarterly Journal of Economics*, 114 (2): 533-575
- Angrist, J., and Lavy, V. (2009) 'The Effect of High-Stakes High School Achievement Awards: Evidence from a Randomized Trial.' *American Economic Review*, vol. 99(4), pp.1384-1414.
- Angrist, J., Dynarski, S., Kane, T., Pathak, P. and Walters, C. (2010) 'Who Benefits from KIPP?' NBER Working Paper No. 15740.
- Angrist, J., Lang, D. and Oreopoulos, P. (2009) 'Incentives and Services for College Achievement: Evidence from a Randomized Trial.' *American Economic Journal: Applied Economics*, vol.1 (1), pp.136-163.
- Angrist, J., Pathak, P. and Walters, C. 2013. 'Explaining Charter School Effectiveness.' *American Economic Journal: Applied Economics*, vol. 5 (4), pp. 1–27.
- Arcidiacono, P., Bayer, P. and Hizmo, A. (2010) 'Beyond Signaling and Human Capital: Education and the Revelation of Ability.' *American Economic Journal: Applied Economics*, vol.2, pp.76–104.
- Ashenfelter, O. and Rouse, C. (1998) 'Income, Schooling and Ability: Evidence from a New Sample of Identical Twins.' *Quarterly Journal of Economics*, Feb. 1998, pp. 253-284.
- Atkinson A., Burgess S., Croxson B., Gregg P., Propper C., Slater H., and Wilson D. (2009) 'Evaluating the impact of performance-related pay for teachers in England.' *Labour Economics*, vol. 16(3), pp.251-61.
- Atkinson, A., Gregg, P. and McConnell, B. (2006) 'The Result of 11 Plus Selection: An Investigation into Equity and Efficiency of Outcomes for Pupils in Selective LEAs.' CMPO DP no. 06/150.
- Attanasio, O., Kugler, A. and Meghir, C. (2011) 'Subsidizing Vocational Training for Disadvantaged Youth in Colombia: Evidence from a Randomized Trial.' *American Economic Journal: Applied Economics*, vol.3 (3), pp.188-220
- Avery, C. and Pathak, P. (2015) 'The Distributional Consequences of Public School Choice' NBER Working Paper No. 21525
- Barber, M., Donnelly, K. and Rizvi, S. (2013) 'An avalanche is coming: Higher education and the revolution ahead.' IPPR Report [http://www.ippr.org/files/images/media/files/publication/2013/04/avalanche-is-coming\\_Mar2013\\_10432.pdf?noredirect=1](http://www.ippr.org/files/images/media/files/publication/2013/04/avalanche-is-coming_Mar2013_10432.pdf?noredirect=1)
- Barrera-Osorio, F., Bertrand, M., Linden, L. and Perez-Calle, F. (2011) 'Improving the Design of Conditional Transfer Programs: Evidence from Randomized Education Experiment in Colombia.' *American Economic Journal: Applied Economics*, vol.3, pp.167-195.
- Barrow, L., and Rouse, C.E. (2013) 'Financial Incentives and Educational Investment: the Impact of Performance-Based Scholarships on Student Time Use,' NBER Working Paper No. 19351.
- Barrow, L., Richburg-Hayes, L., Rouse, C. E. and Brock, T. (2014) 'Paying for Performance: The Education Impacts of a

Community College Scholarship Program for Low-income Adults.' *Journal of Labor Economics*, vol. 32 (3), 563-599

- Bayer, P. and McMillan R. (2010) 'Choice and competition in local education markets.' ERID Working Paper Number 48.
- Bayer, P., Ferreira, F. and McMillan, R. (2007) 'A unified framework for measuring preferences for schools.' *Journal of Political Economy*, vol. 115 (4), pp588-638.
- Becker, G. (1964) 'Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education.' First Edition. NBER New York
- Behrman J., Parker S., Todd P., and Wolpin K. (2012) 'Aligning learning incentives of students and teachers: results from a social experiment in Mexican high schools.' PIER Working Paper No. 13-004.
- Behrman, J., Sengupta, P. and Todd, P. 2005 'Progressing through PROGRESA: An Impact Assessment of a School Subsidy Experiment in Rural Mexico.' *Economic Development and Cultural Change*, vol.54(1), pp.237-275.
- Belzil, C. and Hansen, J. (2002) 'Unobserved Ability and the Return to Schooling', *Econometrica*, vol. 70 (5), pp. 2075-2091.
- Belzil, C. and Hansen, J. (2007) 'A structural analysis of the correlated random coefficient wage regression model.' *Journal of Econometrics*, vol. 140, pp.827-848.
- Benjamin, D., Cesarini, D., Chabris, C., Glaeser, E., Laibson, D., Age, Gene/Environment Susceptibility-Reykjavik Study: Vilmundur Guðnason, Tamara B. Harris, Lenore J. Launer, Shaun Purcell, and Albert Vernon Smith. Swedish Twin Registry: Magnus Johannesson and Patrik K.E. Magnusson. Framingham Heart Study: Jonathan P. Beauchamp and Nicholas A. Christakis. Wisconsin Longitudinal Study: Craig S. Atwood, Benjamin Hebert, Jeremy Freese, Robert M. Hauser, and Taissa S. Hauser . Swedish Large Schizophrenia Study: Alexander Grankvist, Christina M. Hultman, and Paul Lichtenstein (2012) 'The Promises and Pitfalls of Genoeconomics' Annual Review of Economics vol. 4, pp. 627–62
- Berger, L., Paxson, C. and Waldfogel, J. (2009) 'Income and child development. *Children and Youth Services Review*, vol.31 (9), pp.978–989.
- Bertoni, M., Brunello, G., and Rocco, L. (2013) When the Cat is Near the Mice Won't Play: the Effect of External Examiners in Italian Schools *Journal of Public Economics*, vol. 104, pp. 65-77
- Béteille, T., Kalogrides, D. and Loeb,S. (2012) 'Stepping stones: principal career paths and school outcomes.' *Social Science Research*, vol. 41 (4), pp. 904–19.
- Bettinger, E. (2005) 'The effect of charter schools on charter students and public schools.' *Economics of Education Review*, vol. 24 (2), pp.133–147.
- Bettinger, E. (2010) 'Paying to Learn: The Effect of Financial Incentives on Elementary School Test Scores.' NBER Working Paper 16333.
- Bettinger, E. and Baker, R. (2014) 'The Effects of Student Coaching: An Evaluation of a Randomized Experiment in Student Advising.' *Educational Evaluation and Policy Analysis*, vol. 36(1), pp.3-19 .
- Bettinger, E., Long, B., Oreopoulos, P. and Sanbonmatsu, L. (2012) 'The Role of Application Assistance and Information in College Decisions: Results from the H&R Block Fafsa Experiment.' *Quarterly Journal of Economics*, vol. 127(3) pp.1205-1242.
- Betts, J. (2011) 'The Economics of Tracking in Education.' Handbook of Economics of Education vol. 3 (eds) Hanushek, E., Machin, S. and Woessmann, L. Amsterdam: North-Holland
- Bhuller, M., Mogstad, M. and Salvanes, K. (2014) 'Life cycle earnings, education premiums and internal rates of return.' *National Bureau of Economic Research*, WP no. 20250.
- Bifulco, R. and Ladd, H. (2006) 'The Impacts of Charter Schools on Student Achievement: Evidence from North Carolina.' *Education Finance and Policy*, vol. 1 (1), pp.50-90
- Bishop, J. (2006) 'Drinking from the Fountain of Knowledge: Student Incentive to Study and Learn – Externalities, Information Problems and Peer Pressure.' *Handbook of the Economics of Education*, vol. 2, edited by Eric A. Hanushek and Finis Welch



- Björklund, A., Lindahl, M. and Plug, E. (2006) 'The Origins of Intergenerational Associations: Lessons from Swedish Adoption Data' *Quarterly Journal of Economics* 121(3) , 999-1028
- Björklund, A. and Salvanes, K. (2011) 'Education and Family Background: Mechanisms and Policies.' *Handbook of the Economics of Education*, vol. 3, edited by E. Hanushek, S. Machin and L. Woessmann.
- Björklund, A., Clark, M., Edin, P.-A., Fredriksson, P. and Krueger, A. (2005) *The market comes to education in Sweden: An evaluation of Sweden's surprising school reforms*. Russell Sage Foundation: New York.
- Black, S. (1999) 'Do better schools matter? Parental valuation of elementary education.' *Quarterly Journal of Economics*, vol. 114(2), pp. 578–99.
- Black, S. and Devereux, P. (2011) 'Recent Developments in Intergenerational Mobility.' In *Handbook of Labor Economics*, vol. 4B, Elsevier B.V., pp. 1487– 1541.
- Black, S. and Machin, S. (2011) 'Housing Valuations of School Performance' in Eric A. Hanushek, Stephen Machin and Ludger Woessmann (eds.) *Handbook of the Economics of Education*, vol. 3.
- Black, S., Devereux, P. and Salvanes, K. (2005) 'Why the Apple Doesn't Fall Far: Understanding Intergenerational Transmission of Human Capital.' *American Economic Review*, vol.95, pp. 437-449.
- Black, S., Devereux, P. and Salvanes, K. (2008) 'Staying in the Classroom and out of the Maternity Ward? The Effect of Compulsory Schooling Laws on Teenage Births.' *Economic Journal*, vol. 118, pp. 1025–54.
- Black, S., Devereux, P., and Salvanes, K. (2005b) 'Why the apple doesn't fall far: Understanding intergenerational transmission of human capital.' *American Economic Review*, vol. 95 (1), pp.437–449.
- Blakemore, S.-J. (2010) 'The Developing Social Brain: Implications for Education.' *Neuron*, vol.65, pp.744- 747.
- Blatchford, P., Bassett, P., Brown, P., Martin, C., Russell, A. and Webster, R.(2009) 'Deployment and Impact of Support Staff Project.' DCSF Research Brief DCSF-RB148. [http://www.ioe.ac.uk/diss\\_research\\_summary.pdf](http://www.ioe.ac.uk/diss_research_summary.pdf)
- Blau, D. and Currie, J. (2006) 'Pre-School, Day Care, and After-School Care: Who's Minding the Kids?' *Handbook of the Economics of Education*, edited by E. Hanushek and F. Welch.
- Blazar, D. and Kraft, M. (2015) Teacher and Teaching Effects on Cognitive versus Non-cognitive Outcomes. Mimeo
- Bloom, N., Lemos, R., Sadun, R. and Van Reenen, J. (2014) 'Does management matter in schools?' NBER Working Paper No. 20667.
- Böhlmark, A. and Lindahl, M. (2008) 'Does School Privatization Improve Educational Achievement? Evidence from Sweden's Voucher Reform.' IZA Discussion Paper No. 3691.
- Böhlmark, A. and Lindahl, M. (2012) 'Independent schools and long-run educational outcomes – evidence from Sweden's large scale voucher reform.' IFAU Working Paper 2012:19.
- Böhlmark, A. and M Lindahl (2007) 'The Impact of School Choice on Pupil Achievement, Segregation and Costs: Swedish Evidence.' IZA Discussion Paper No.2786.
- Borghans, L., Duckworth, A., Heckman, J. and ter Weel, B. (2008) 'The Economics and Psychology of Personality Traits.' *Journal of Human Resources*, vol. 43 (4) pp.972– 1059.
- Borghans, L., Golsteyn, B., Heckman, J. and Humphries, J. (2011) 'Identification Problems in Personality Psychology.' *Personality and Individual Differences*, vol. 51 (3), pp. 315–20.
- Borghans, L., Golsteyn, H., Heckman, J. and Humphries, J. (2011) 'IQ, achievement, and personality.' Unpublished manuscript, University of Maastricht and University of Chicago
- Borghans, L., Meijers, H. and ter Weel, B. (2013) 'The importance of intrinsic and extrinsic motivation for measuring IQ' *Economics of Education Review*, vol. 34(C), pp. 17-28.
- Bouguen, A. and Gurgand, M. (2012) 'Randomized Controlled Experiments in Education' EENEE Analytical Report No. 11. <http://www.eenee.de/eeneeHome/EENEE/Analytical-Reports.html>
- Bowles, S., Gintis, H., and Osborne, M., (2001) 'Incentive-Enhancing Preferences: Personality, Behavior, and Earnings' *American Economic Review* vol. 91 (2), 155–158.

- Boyd, D., Grossman, P., Hammerness, K., Lankford, H., Loeb, S., McDonald, M., Reininger, M., Ronfeldt, M., and Wyckoff, J. (2008) 'Surveying the landscape of teacher education in New York city: Constrained variation and the challenge of innovation.' *Education Evaluation and Policy Analysis*, vol. 30(4), pp. 319-343.
- Boyd, D., Grossman, P., Lankford, H., Loeb, S., Ronfeldt, M., and Wyckoff, J. (2012) 'Recruiting effective math teachers: Evidence from New York city.' *American Education Research Journal*, vol. 49(6), pp.1008-1047.
- Boyd, D., Lankford, H., Loeb, S. and Wyckoff, J. (2008) 'The impact of assessment and accountability on teacher recruitment and retention: Are there unintended consequences?' *Public Finance Review* vol. 36, pp. 88–111.
- Boyd, D., Lankford, H., Loeb, S. and Wyckoff, J. (2011) 'Teacher Layoffs: An Empirical Illustration of Seniority versus Measures of Effectiveness'. *Education Finance and Policy*, vol. 6(3), pp. 439-454.
- Boyd, D., Lankford, H., Rockoff, J., Loeb, S. and Wyckoff, J. (2008). 'The narrowing gap in New York City teacher qualifications and its implications for student achievement in high-poverty schools.' *Journal of Policy Analysis and Management*, vol. 27 (4), pp. 793–818.
- Bramoullé, Y., Djebbari, H. and Fortin, B. (2009) 'Identification of peer effects through social networks.' *Journal of Econometrics*, vol. 150, pp.41–55.
- Branch, G., Hanushek, E. and Rivkin, S. (2012) 'Estimating the effect of leaders on public sector productivity: the case of school principals.' Working Paper 17803. National Bureau of Economic Research.
- Branigan, A.R., McCallum, K. and Freese, J. (2013) "Variation in the Heritability of Educational Attainment: An International Meta-Analysis" *Social Forces* vol. 92(1) pp. 109–140.
- Brunello, G. and Cappellari, L. (2008) 'The labour market effects of *Alma Mater*: Evidence from Italy' *Economics of Education Review*, vol 27, no. 5, pp. 564-574
- Brunello, G. and Checchi, D. (2007) 'Does school tracking affect equality of opportunity? New international evidence.' *Economic Policy*, vol.22 (52), pp.781–861.
- Brunello, G., Fort, M. and Weber, G. (2009) 'Changes in compulsory schooling, education and the distribution of wages in Europe.' *Economic Journal*, vol. 119, pp. 516-539.
- Brunello, G. and L. Rocco (2008) 'Educational Standards in Private and Public Schools' *Economic Journal* vol. 118, pp. 1866–1887.
- Brunello, G. and L. Rocco (2015), "The effects of vocational education on adult skills and wages: What can we learn from PIAAC?" OECD Social, Employment and Migration Working Papers, No. 168, OECD Publishing, Paris. <http://dx.doi.org/10.1787/5jrxfmjvw9bt-en>
- Brunello, G. and Schlotter, M. (2011) 'Non Cognitive Skills and Personality Traits: Labour Market Relevance and their Development in E&T Systems' EENEE Analytical Report No. 8. <http://www.eenee.de/eeneeHome/EENEE/Analytical-Reports.html>
- Burgess, S. and Briggs, A. (2009) 'School assignment, school choice and social mobility.' *Economics of Education Review*, vol. 29(4), pp. 639–49.
- Burgess, S. and Ratto, M. (2003) 'The role of incentives in the public sector: issues and evidence.' *Oxford Review of Economic Policy*, vol. 19(2,) pp. 285-300.
- Burgess, S. and Slater, H. (2006) 'Using Boundary Changes to Estimate the Impact of School Competition on Test Scores.' CMPO DP 06/158. CMPO.
- Burgess, S., Dickson, M. and Macmillan, L. (2014) 'Selective schooling systems increase inequality.' CMPO Discussion Paper 14/323, CMPO, University of Bristol.
- Burgess, S., Greaves, E., Vignoles, A. and Wilson, D. (2014) 'What Parents Want: School Preferences and School Choice.' *The Economic Journal*, doi: 10.1111/econ.12153.
- Burgess, S., List, J., Metcalfe, R. and Sadoff, S. (2015) 'Using behaviour incentives to improve performance on high stakes tests: Evidence from a field experiment.' mimeo, CMPO.
- Burgess, S., McConnell, B., Propper, C. and Wilson, D. (2007) 'The impact of school choice on sorting by ability and socioeconomic factors in English secondary schools.' In Woessmann, L. and Peterson, P. (eds.) *Schools and the Equal Opportunity Problem*, MIT Cambridge.

- Burgess, S., Sanderson, E. and Umaña-Aponte, M. (2011) 'School ties: An analysis of homophily in an adolescent friendship network.' CMPO DP 11/267, CMPO.
- Burgess, S., Wilson, D. and Worth, J. (2013) 'A natural experiment in school accountability: The impact of school performance information on pupil progress.' *Journal of Public Economics*, vol.106 (C), pp.57-67.
- Buscha, F. and Dickson, M. (2015) 'Heterogeneity over the Life-Cycle: Re-Examining the Wage Returns to Education in Britain.' mimeo, University of Bath.
- Calsamiglia, C., Haeringer, G. and Klijn, F. (2010) 'Constrained school choice: an experimental study.' *American Economic Review*, vol. 100(4), pp. 1860–74.
- Card, D. (1999) 'The causal effects of education on earnings.' In: Ashenfelter, O., Card, D. (eds.), *Handbook of Labor Economics*, vol. 5. North-Holland, New York, pp. 1801–1863.
- Card, D. (2001) 'Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems.' *Econometrica*, vol. 69, pp. 1127-1160.
- Card, D. and Krueger, A. (1992) 'Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States' *Journal of Political Economy* vol. 100(1), pp. 1-40.
- Card, D. and Lemieux, T. (2001) 'Can falling supply explain the rising return to college for younger men? A cohort-based analysis', *Quarterly Journal of Economics*, vol. 116 (2), pp. 705–746.
- Card, D., Ibarra-rá'n, P., Regalia, F., Rosas-Shady, D. and Soares, Y. (2011) 'The Labor Market Impacts of Youth Training in the Dominican Republic' *Journal of Labor Economics*, vol. 29, (2), pp. 267 – 300.
- Carlsson, M., Dahl, G., Öckert, B. and Rooth, D. (2015) 'The Effect of Schooling on Cognitive Skills. *Review of Economics and Statistics*, vol. 97 (3), pp. 533-547.
- Carneiro, P. and Heckman, J. (2002) 'The Evidence on Credit Constraints in Post-Secondary Schooling.' *Economic Journal*, vol. 112, pp. 705–734.
- Carneiro, P. and Heckman, J. (2002) 'The evidence on credit constraints in post-secondary schooling.' *The Economic Journal*, vol.112, pp.705–730.
- Carneiro, P., Hansen, K. and Heckman, J. (2001) 'Removing the veil of ignorance in assessing the distributional impacts of social policies', *Swedish Economic Policy Review*, vol. 8, no. 2, pp. 273–301.
- Carneiro, P., Hansen, K. and Heckman, J. (2003) 'Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice', *International Economic Review*, vol. 44, no. 2, pp. 361–422.
- Carneiro, P., Heckman, J. and Vytlacil, E. (2011) 'Estimating Marginal Returns to Education.' *American Economic Review*, vol. 101 (6), pp. 2754-81.
- Carneiro, P., Løken, K. and Salvanes, K. (2015) 'A Flying Start? Maternity Leave Benefits and Long Run Outcomes of Children' *Journal of Political Economy*, 123 (2) 365-412.
- Carneiro, P., Meghir, C. and Parey, M. (2013) 'Maternal Education, Home Environments, And The Development Of Children And Adolescents' *Journal of the European Economic Association*, vol. 11, pages 123-160
- Cassen, R., McNally, S. and Vignoles, A. (2015) *Making a Difference in Education* Routledge, London.
- Chapman, B. (2006) 'Income Contingent Loans for Higher Education: International Reforms.' *Handbook of the Economics of Education*, vol. 2, ed. E. Hanushek and F. Welch.
- Chetty, R., Friedman, J. and Rockoff, J. (2014a) 'Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates.' *American Economic Review*, September, pp. 2593-2632.
- Chetty, R. Friedman J., and Rockoff J. (2014b) 'Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood.' *American Economic Review*, September, pp. 2633-2679.
- Chetty, R., Friedman, J. and Rockoff, J. (2015) 'Measuring the Impacts of Teachers: Response to Rothstein.' (2014) mimeo Harvard.

- Chetty, R., Friedman, J., Hilger, N., Saez, E., Schanzenbach, D. and Yagan, D. (2011) 'How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR.' *The Quarterly Journal of Economics*, vol. 126 (4), pp. 1593-1660.
- Chetty, R., Hendren, N., Kline, P. and Saez, E. (2014a) 'Where is the land of Opportunity? The Geography of Intergenerational Mobility in the United States.' *The Quarterly Journal of Economics*, vol.129, pp.1553–1623.
- Chevalier, A. (2004) 'Parental education and child's education: A natural experiment.' IZA Discussion Paper no. 1153.
- Chevalier, A., Harmon, C., Walker, I. and Zhu, Y. (2004) 'Does education raise productivity, or just reflect it?' *The Economic Journal*, vol.114, F499–F517.
- Clark, D. (2009) 'The Performance and Competitive Effects of School Autonomy.' *Journal of Political Economy*, vol. 117(4), pp. 745-783.
- Clark, D. (2010) 'Selective Schools and Academic Achievement.' *B.E. Journal of Economic Analysis and Policy*, vol.10 (1), pp.1935-1682.
- Clark, D. and Del Bono, E. (2014) 'The Long-Run Effects of Attending an Elite School: Evidence from the UK.' ISER Working Paper No. 2014-05.
- Clark, D., Martorell, P. and Rockoff, J. (2009) 'School principals and school performance.' CALDER Working Papers. Center for Analysis of Longitudinal Data in Education Research.
- Clotfelter, C. Glennie, E., Ladd, H. and Vigdor, J. (2006) 'Would higher salaries keep teachers in high-poverty schools? Evidence from a policy intervention in North Carolina' *Journal of Public Economics* vol. 92(5-6) pp. 1352-1370
- Clotfelter, C., Ladd, H. and Vigdor, J. (2006) 'Teacher-student matching and the assessment of teacher effectiveness.' *Journal of Human Resources*, vol. 41, pp. 778–820.
- Clotfelter, C., Ladd, H. and Vigdor, J. (2007) 'Teacher credentials and student achievement: longitudinal analysis with student fixed effects', *Economics of Education Review*, vol. 26, pp. 673–682.
- Coelli, M. and Green, D. (2012) 'Leadership effects: school principals and student outcomes.' *Economics of Education Review*, vol. 31 (1), pp. 92–109.
- Coleman, J., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfeld, F. and York, R. (1966) *Equality of Educational Opportunity*. Washington, D.C.: U. S. Government Printing Office.
- Craig, S., Imberman, S. and Perdue, A. (2015) 'Do administrators respond to their accountability ratings? The response of school budgets to accountability grades.' *Economics of Education Review*, (forthcoming).
- CREDO (2009) 'Multiple Choice: Charter Performance in Sixteen States.' Center for Research on Education Outcomes, Stanford University. [http://credo.stanford.edu/reports/MULTIPLE\\_CHOICE\\_CREDO.pdf](http://credo.stanford.edu/reports/MULTIPLE_CHOICE_CREDO.pdf)
- Cullen, J., Jacob, B. and Levitt, S. (2006) 'The Effect of School Choice on Participants: Evidence from Randomized Lotteries.' *Econometrica*, vol. 74(5,) pp.1191-1230.
- Cunha, F. and Heckman, J. (2007) 'Identifying and estimating the distributions of *Ex Post* and *Ex Ante* returns to schooling: A survey of recent developments.' *Labour Economics*, vol. 14 (6), pp. 870-893.
- Cunha, F. and Heckman, J. (2007). 'The technology of skill formation.' *American Economic Review*, vol. 97(2), pp.31–47.
- Cunha, F., Heckman, J. and Navarro, S. (2005) 'Separating uncertainty from heterogeneity in life cycle earnings, The 2004 Hicks lecture', *Oxford Economic Papers*, vol. 57, no. 2, pp. 191–261.
- Cunha, F., Heckman, J. and Navarro, S. (2006) 'Counterfactual analysis of inequality and social mobility', in: Morgan, S., Grusky, D., Fields, G. (eds.), *Mobility and Inequality: Frontiers of Research in Sociology and Economics*, Stanford University Press, Palo Alto. Chapter 4.
- Cunha, F., Heckman, J. and Schennach, S. (2010) 'Estimating the Technology of Cognitive and Noncognitive Skill Formation.' *Econometrica*, vol. 78 (3), pp. 883–931.
- Cunha, F., Heckman, J., Lochner, L. and Masterov, D. (2006) 'Interpreting the Evidence on Life Cycle Skill Formation.' *Handbook of the Economics of Education*, edited by E. Hanushek and F. Welch. Amsterdam: North-Holland

- Currie, J. and Almond, D. (2011) 'Human capital development before age five.' Handbook of Labor Economics, vol 4 (B), edited by D. Card and O. Ashenfelter. Amsterdam: North-Holland
- Currie, J. and Thomas, D. (2000) 'School Quality and the Longer-Term Effects of Head Start.' *Journal of Human Resources*, vol.35(4), pp.755–74.
- Currie, J. and Hyson, R. (1999) 'Is the impact of shocks cushioned by socioeconomic status? The case of low birth weight.' *American Economic Review*, vol. 89 (2), pp.245–250.
- De Fraja, G., and Iossa, E. (2002) 'Competition among Universities and the Emergence of the Elite Institution.' *Bulletin of Economic Research* vol. 54(3), pp.275-293.
- de Haan, M., Leuven, E. and Oosterbeek, H. (2011) 'Scale economies can offset the benefits of competition: Evidence from a school consolidation reform in a universal voucher system.' CEPR Discussion Papers 8272.
- de Haan, M. (2015) 'The effect of additional funds for low-ability pupils: A nonparametric bounds analysis.' *Economic Journal* forthcoming.
- Dearden, L., McIntosh, S., Myck, M. and Vignoles, A. (2002) 'The returns to academic, vocational and basic skills in Britain.' *Bulletin of Economic Research*, vol. 54, pp. 249–274.
- Dee, T. (2004) 'Are There Civic Returns to Education?' *Journal of Public Economics*, vol. 88 (9–10), pp. 1697–1720.
- Dee, T. and Jacob, B. (2009) 'The impact of No Child Left Behind on student achievement.' NBER Working Paper No 15531, NBER: Cambridge MA.
- Dee, T. and Wyckoff, J. (2013) 'Incentives, selection, and teacher performance: Evidence from IMPACT' NBER Working Paper No. 19529
- Del Boca, D. (2015) 'Child Care Arrangements and Labor Supply.' IDB Working Paper Series IDB-WP-569.
- Del Boca, D., Flinn, C. and Wiswall, M. (2014) 'Household Choices and Child Development.' *Review of Economic Studies*, vol. 81, pp. 137-185.
- Del Bono, E., Francesconi, M., Kelly, Y and Sacker, A. (2015) 'Early Maternal Time Investment and Early Child Outcomes.' *Economic Journal*, forthcoming.
- Del Bono, E., Francesconi, M., Kelly, Y. and Sacker, A. (2015) 'Early Maternal Time Investment and Early Child Outcomes.' *Economic Journal*, forthcoming.
- Del Rey, E. (2001) 'Teaching versus Research: A Model of State University Competition.' *Journal of Urban Economics*, vol. 49, pp. 356-373.
- Dickson, M., Gregg, P. and Robinson, H. (2015) 'Early, late or never? When does parental education impact child outcomes?' *Economic Journal*, forthcoming.
- Dickson, M. and Harmon, C. (2011) 'Economic Returns to Education: What We Know, What We Don't Know, and Where We Are Going – Some Brief Pointers' *Economics of Education Review*, Vol. 30 (6), pp. 1118-1122
- Dijkgraaf, E., Gradus, R. and de Jong, J.M. (2013) 'Competition and Educational Quality: Evidence from the Netherlands.' *Empirica*, vol. 40, pp. 607 – 634.
- Dixit, A. (2002) 'Incentives and organisations in the public sector: an interpretive review.' *Journal of Human Resources*, vol. 37(4), pp. 696-727.
- Dizon-Ross, R. (2014) 'Parents' Perceptions and Children's Education: Experimental Evidence from Malawi. mimeo, MIT, <http://www.mit.edu/rdr/perceptions.pdf>
- Dobbie, W. (2011) 'Teacher Characteristics and Student Achievement: Evidence from Teach for America.' Harvard University Working Paper.
- Dobbie, W. and Fryer, R. (2011) 'Are High Quality Schools Enough to Close the Achievement Gap? Evidence from a Social Experiment in Harlem.' *American Economic Journal: Applied Economics*, vol. 3(3)
- Dobbie, W. and Fryer, R. (2011) 'Exam High Schools and Academic Achievement: Evidence from New York City.' NBER WP no. 17286.

- Dobbie, W., and Fryer, R. (2013) 'Getting beneath the Veil of Effective Schools: Evidence from New York City.' *American Economic Journal: Applied Economics*, vol.5(4), pp.28-60.
- Dolton, P. (2006) Teacher Supply in Handbook of Economics of Education vol. 2 (eds) Hanushek, E, and Welch, F. Amsterdam: North-Holland
- Doyle, O., Harmon, C., Heckman, J. and Tremblay, R. (2009) 'Investing in early human development: Timing and economic efficiency.' *Economics and Human Biology* vol.7, pp. 1–6.
- Duckworth, A., Peterson, C., Matthews, M. and Kelly, D. (2007) 'Grit: Perseverance and passion for long-term goals.' *Journal of Personality and Social Psychology*, vol.92 (6), p. 1087-1095
- Duflo, E., Dupas, P. and Kremer, M. (2011) 'Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya.' *American Economic Review*, vol. 101(5) pp. 1739-74.
- Dynarski, S. (2000) 'Hope for Whom? Financial Aid for the Middle Class and Its Impact on College Attendance.' *National Tax Journal*, vol. 53 (3), pp. 629-662.
- Epple, D., Figlio, D. and Rockoff, J. (2004) 'Competition between Private and Public Schools: Testing Stratification and Pricing Predictions.' *Journal of Public Economics*, vol. 88, pp. 1215 – 1245.
- Epple, D. and Romano, R. (2003) 'Neighbourhood schools, choice, and the distribution of educational benefits.' in Hoxby, C. (ed.) *The Economics of School Choice*, Chicago: University of Chicago Press.
- Epple, D. and Romano, R. (2011) 'Peer Effects in Education: A Survey of the Theory and Evidence.' In Benhabib, J., Bisin, A. and Jackson, M. (eds) *Handbook of Social Economics 1A, 1B*, Amsterdam: North-Holland.
- Epple, D. and Romano, R. (2012) 'Economic Modeling and Analysis of Educational Vouchers.' *Annual Review of Economics*, vol. 4, pp. 159-184.
- Epple, D., Romano, R., Sarpça, S. and Sieg, H. (2013) 'The U.S. Market for Higher Education: A General Equilibrium Analysis of State and Private Colleges and Public Funding Policies.' NBER Working Paper 19298.
- Epple, D., Romano, R. and Urquiola, M. (2015) 'School Vouchers: A Survey of the Economics Literature' NBER Working Paper No. 21523
- Epple, D., Romano, R. and Zimmer, R. (2015) 'Charter Schools: A Survey of Research on Their Characteristics and Effectiveness' NBER Working Paper No. 21256
- European Commission (2015) "An in-depth analysis of adult learning policies and their effectiveness in Europe" European Commission, Brussels.
- Eyles, A. and Machin, S. (2015) 'The Introduction of Academy Schools to England's Education' CEP Discussion Paper No 1368
- Faber, B., Sanchis-Guarner, R. and Weinhardt, F. (2015) 'ICT and Education: Evidence from Student Home Addresses' NBER Working Paper No. 21306
- Fack, G. and Grenet, J. (2010) 'When do better schools raise housing prices? Evidence from Paris public and private schools.' *Journal of Public Economics*, vol. 94, pp.59–77
- Falch, T. (2011) 'Teacher mobility responses to wage changes: Evidence from a quasi-natural experiment' *American Economic Review* vol. 101(3) pp. 460-465
- Falch, T. and Rønning, M. (2007) 'The influence of student achievement on teacher turnover' *Education Economics* vol. 15(2), pp. 177-202
- Falch, T. and Strøm, B. (2005) 'Teacher Turnover and Non-Pecuniary Factors' *Economics of Education Review* vol. 24(6), pp. 611-631
- Farasat, A. and Bokhari, H. S. (2011) 'School accountability laws and the consumption of psychostimulants' *Journal of Health Economics*, vol. 30 (2), pp. 355 – 372
- Figlio, D. (2005) 'Boys named Sue: Disruptive children and their peers.' NBER Working Paper No. 11277.
- Figlio, D. and Ladd, H. (2008) 'School accountability and student achievement.' In Ladd, H. and Fiske, E. (eds.) *Handbook of Research in Education Finance and Policy*, London: Routledge.

- Figlio, D. and Loeb, S. (2011) 'School Accountability' in Handbook of Economics of Education vol. 3 (eds) Hanushek, E., Machin, S. and Woessmann, L. Amsterdam: North-Holland
- Figlio, D. and Page, M.E. (2002) 'School choice and the distributional effects of ability tracking: Does separation increase inequality?' *J. Urban Econ*, vol.51 (3), pp.497–514.
- Fiorini, M. and Keane, M. (2014) 'How the allocation of children's time affects cognitive and noncognitive development.' *Journal of Labor Economics*, vol. 32(4), pp 787 – 836.
- Fredriksson, P., Öckert, B. and Oosterbeek, H. (2013) Long-term effects of class size *Quarterly Journal of Economics* vol. 128, pp. 249-285
- Fredriksson, P., Öckert, B. and Oosterbeek, H. (2015) 'Parental responses to public investments in children: Evidence from a maximum class size rule' *Journal of Human Resources*. (Forthcoming)
- Fryer R, and Dobbie W. (2011) 'Are High-Quality Schools Enough to Increase Achievement Among the Poor? Evidence from the Harlem Children's Zone.' *American Economic Journal: Applied Economics*, vol.3 (3), pp. 158 – 187.
- Fryer R, and Katz L. (2013) 'Achieving Escape Velocity: Neighborhood and School Interventions to Reduce Persistent Inequality.' *American Economic Review (Papers and Proceedings)*, vol.103 (3) pp.232-237.
- Fryer R., Levitt S., List J., and Sadoff S. (2012) 'Enhancing the efficacy of teacher incentives through loss aversion: a field experiment.' NBER Working Paper 18237.
- Fryer, R. (2011) 'Financial Incentives and Student Achievement: Evidence from Randomized Trials.' *The Quarterly Journal of Economics*, vol. 126(4), pp.1755-1798.
- Fryer, R. (2012) 'Aligning Student, Parent and Teacher Incentives: Evidence from Houston Public Schools.' NBER Working Paper 17752.
- Fryer, R. (2013) 'Teacher incentives and student achievement: evidence from New York City public schools.' *Journal of Labor Economics*, vol. 31(2), pp.373-407.
- Fryer, R. (2014) 'Injecting Charter School Best Practices into Traditional Public Schools: Evidence from Field Experiments.' *Quarterly Journal of Economics*, 129(3):1355-1407
- Fryer, R. and Dobbie, W. (2015) 'The Medium-Term Impacts of High-Achieving Charter Schools.' *Journal of Political Economy*, forthcoming.
- Fryer, R. and Levitt, S. (2004) 'Understanding the Black-White Test Score Gap in the First Two Years of School.' *Review of Economics and Statistics*, vol.86 (2), pp. 447–64.
- Fryer, R. and Levitt, S. (2013) 'Testing for Racial Differences in the Mental Ability of Young Children.' *American Economic Review*, vol.103 (2), pp.981-1005.
- Fryer, R., and Levitt, S. (2006) 'The Black-White Test Score Gap through Third Grade.' *American Law and Economics Review*, vol. 8 (2), pp.249–81.
- Fryer, R., Levitt, S. and List, J. (2015) 'Parental Incentives and Early Childhood Achievement: A Field Experiment in Chicago Heights' NBER Working Paper No. 21477
- Fryer, R. and Torelli, P. (2010) 'An Empirical Analysis of Acting White.' *Journal of Public Economics*, vol. 94, pp. 380 – 396.
- Garibaldi, P., Giavazzi, F., Ichino, A. and Rettore, E. (2012) 'College Cost and Time to Complete a Degree: Evidence from Tuition Discontinuities.' *Review of Economics and Statistics*, vol. 94(3) pp.699-711.
- Gelber, A. and Isen, A. (2013) 'Children's schooling and parents' behavior: Evidence from the Head Start Impact Study' *Journal of Public Economics*, vol. 101(C), pp. 25-38
- Gibbons, S. and Machin, S. (2003) 'Valuing English primary schools' *Journal of Urban Economics* vol. 53 pp. 197 – 219
- Gibbons, S. and Machin, S. (2006) 'Paying for Primary Schools: Admission Constraints, School Popularity or Congestion.' *Economic Journal*, vol.116, pp. 77-92.
- Gibbons, S. and McNally, S. (2013) 'The Effects of Resources Across School Phases: A Summary of Recent Evidence.' CEP Discussion Paper No 1226.



- Gibbons, S., Machin, S. and Silva, O. (2010) 'Choice, competition, and pupil achievement.' *Journal of the European Economic Association*, vol. 6 (4), pp. 912–947
- Gibbons, S., McNally, S. and Viarengo, M. (2011) 'Does Additional Spending Help Urban Schools? An Evaluation Using Boundary Discontinuities.' SERC Discussion Paper No. 90.
- Glaeser, E., Ponzetto, G. and Shleifer, A. (2005) 'Why Does Democracy Need Education?' *Journal of Economic Growth*, vol. 12 (2), pp. 77–99.
- Glazerman, S., Isenberg, E., Dolfin, S., Bleeker, M., Johnson, A., Grider, M. and Jacobus, M. (2010) Impacts of Comprehensive Teacher Induction: Final Results from a Randomized Controlled Study. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
- Goldhaber, D. and Chaplin D. (2012) 'Assessing the `Rothstein Falsification Test': Does It Really Show Teacher Value-Added Models Are Biased?' CEDR Working Paper 2011-5.
- Goldhaber, D., Gross, B. and Player, D. (2011) 'Teacher career paths teacher quality and persistence in the classroom: Are public schools keeping their best?' *Journal of Policy Analysis and Management* vol. 30(1) pp. 57-87
- Goldhaber, D., Grout, C., and Huntington-Klein, N. (2014) 'Screen Twice, Cut Once: Assessing the Predictive Validity of Teacher Selection Tools' CEDR WP 2014-9, Center for Education Data and Research, Seattle.
- Goldhaber, D. and Hansen, M. (2010) 'Using performance on the job to inform teacher tenure decisions.' *American Economic Review*, vol. 100, pp. 250–255.
- Goldhaber, D. and Hansen, M. (2013) 'Is it Just a Bad Class? Assessing the Long-term Stability of Estimated Teacher Performance.' *Economica*, vol. 80(319), pp 589–612.
- Goldhaber, D. and Liddle, S. (2011) 'The Gateway to the Profession. Assessing Teacher Preparation Programs Based on Student Achievement.' CEDR Working Paper 2011-2. University of Washington, Seattle, WA.
- Goldin, C. and Katz, L. (2008) *The race between education and technology*. Cambridge, MA: Harvard University Press.
- Goodman, A. and Gregg, P. (2010) 'Poorer children's educational attainment: how important are attitudes and behaviour?' JRF report <http://www.irf.org.uk/publications/educational-attainment-poor-children>
- Gordon, R., Kane, T. and Staiger, D. (2006) 'Identifying Effective Teachers Using Performance on the Job.' The Hamilton Project White Paper, 2006-01, Washington, DC.
- Green, D. and W. Craig Riddell (2009) 'Understanding Educational Impacts: The Role of Cognitive Skills' UBC mimeo.
- Green, F., Machin, S., Murphy, R. and Zhu, Y. (2012) 'The changing economic advantage from private schools' *Economica* vol. 79, pp. 658-79
- Green, F., Machin, S., Murphy, R. and Zhu, Y. (2008) 'Competition for Private and State School Teachers', *Journal of Education and Work*, 21, 385- 404
- Gregg, P., Jonsson, J., Macmillan, L. and Mood, C., (2013) 'Understanding income mobility: the role of education for intergenerational income persistence in the US, UK and Sweden' DoQSS DP 13-12, UCL Institute of Education
- Grek, S. (2009) 'Governing by numbers: the PISA 'effect' in Europe.' *Journal of Education Policy*, vol.24 (1), pp. 23–37.
- Grissom, J., and Loeb, S. (2011) 'Triangulating principal effectiveness: how perspectives of parents, teachers, and assistant principals identify the central importance of managerial skills.' *American Educational Research Journal*, vol. 48 (5), pp. 1091– 1123.
- Grissom, J., Kalogrides, D. and Loeb, S. (2015) 'Using student test scores to measure principal performance.' *Education Evaluation and Policy Analysis*, vol.37 (1), pp. 3-28.
- Grossman, M. (2006) 'Education and Nonmarket Outcomes.' *Handbook of the Economics of Education*, vol. 1, Elsevier, pp. 578-633.
- Grossman, P., and Loeb, S. (2010) 'Learning from multiple routes: The variation in teacher preparation pathways can propel our understanding of how best to prepare teachers.' *Educational Leadership*, vol. 67(8), pp.22-27.
- Haeringer, G. and Klijn, F. (2009) 'Constrained school choice.' *Journal of Economic Theory*, vol. 144(5), pp. 1921–47.



- Hammermesh, D. (1999) 'Changing Inequality in Markets for Workplace Amenities.' *Quarterly Journal of Economics*, vol. 114 (4), pp. 1085–1123.
- Hanushek E. (1971) 'Teacher characteristics and gains in student achievement: estimation using micro data.' *American Economic Review*, vol. 61(2), pp.280-88.
- Hanushek, E. (2003) 'The Failure of Input-based Schooling Policies.' *Economic Journal*, vol. 113, pp. F64-98.
- Hanushek, E. (2006) 'School Resources' in Handbook of Economics of Education vol. 2 (eds) Hanushek, E, and Welch, F. Amsterdam: North-Holland
- Hanushek E. (2011) 'The economic value of higher teacher quality.' *Economics of Education Review*, vol. 30, pp.466-479.
- Hanushek, E., Kain, J. and Rivkin, S. (2004) 'Why Public Schools Lose Teachers' *Journal of Human Resources* vol. 39(2) pp. 326-354 .
- Hanushek, E., Kain, J. and Rivkin, S. (2009) 'New evidence about Brown v. Board of Education: The complex effects of school racial composition on achievement.' *Journal of Labor Economics*, vol.27 (3), pp.349–383.
- Hanushek, E., Link, S. and Woessmann, L. (2013) 'Does School Autonomy Make Sense Everywhere? Panel Estimates from PISA.' *Journal of Development Economics*, vol. 104, pp. 212-232.
- Hanushek, E. and Raymond, M. (2005) 'Does school accountability lead to improved student performance?' *Journal of Policy Analysis and Management*, vol.24 (2), pp.297-327.
- Hanushek, E. and Rivkin, S. (2006) 'Teacher quality.' In E. Hanushek and F. Welch (eds.), *Handbook of the economics of education*, vol. 2, pp. 1051–1078. Amsterdam: North Holland.
- Hanushek, E. and Rivkin, S. (2010) 'Generalizations about using value added measures of teacher quality.' *American Economic Review*, vol.100 (2), pp.267–271.
- Hanushek, E. and Woessmann, L. (2006) 'Does Educational Tracking Affect Performance and Inequality? Differences-in-Differences Evidence across Countries' *Economic Journal*, vol.116 (510), pp. C63-C76.
- Hanushek, E. and Woessmann, L. (2010) The High Cost of Low Educational Performance: the long-run economic impact of improving PISA outcomes. OECD
- Hanushek, E. and Woessmann, L. (2011) 'The Economics of International Differences in Educational Achievement.' *Handbook of the Economics of Education* vol 3, Edited by Eric A. Hanushek, Stephen Machin and Ludger Woessmann.
- Hanushek, E. and Woessmann, L. (2012) 'Do Better Schools Lead to More Growth? Cognitive Skills, Economic Outcomes, and Causation.' *Journal of Economic Growth*, vol.17(4), pp. 267-321.
- Hanushek, E. and Woessmann, L. (2015) *The Knowledge Capital of Nations: Education and the Economics of Growth*. Cambridge, MA: MIT Press.
- Harmon, C. and Walker, I. (1995) 'Estimates of the Economic Return to Schooling for the UK.' *American Economic Review*, vol.85, pp.1278-86.
- Harmon, C. and Walker, I. (1999) 'The marginal and average returns to schooling in the UK.' *European Economic Review*, vol.43, pp.879-887.
- Harmon, C., Hogan, V. and Walker, I. (2003) 'Dispersion in the economic return to schooling.' *Labour Economics*, vol. 10, pp. 205-214.
- Harris, D. and Sass, T. (2011) 'Teacher Training, Teacher Quality and Student Achievement'. *Journal of Public Economics* vol. 95(7-8), pp. 798-812
- Hastings, J. and Weinstein, J. (2008) 'Information, School Choice, and Academic Achievement: Evidence from two Experiments.' *Quarterly Journal of Economics*, vol. 123, pp. 1373 – 1414.
- Hastings, J., Kane, T. and Staiger, D. (2006, 2008) 'Heterogeneous Preferences and the Efficacy of Public School Choice.' NBER Working Paper No. 12145.

- Hastings, J., Kane, T. and Staiger, D. (2008) 'Heterogeneous Preferences and the Efficacy of Public School Choice.' Combines and replaces National Bureau of Economic Research Working Papers No. 12145 and 11805. [http://aida.econ.yale.edu/~jh529/papers/HKS\\_Combined\\_200806.pdf](http://aida.econ.yale.edu/~jh529/papers/HKS_Combined_200806.pdf)
- Hastings, J., Neilson, C. and Zimmerman, S. (2012) 'The Effect of School Choice on Intrinsic Motivation and Academic Outcomes.' NBER Working Paper No. 18324.
- Heckman, J. (1990) 'Varieties of selection bias', *American Economic Review*, vol. 80 (2), pp. 313–318.
- Heckman, J. (2004) 'Lessons from the Technology of Skill Formation.' *Annals of the New York Academy of Sciences*, vol. 1038, Issue: Understanding and Optimizing Human Development: From Cells to Patients to Populations, pp. 179–200.
- Heckman, J. (2008) 'Schools, skills and synapses.' *Economic Inquiry*, vol. 46(3), 289–324.
- Heckman, J. and Mosso, S. (2014) 'The Economics of Human Development and Social Mobility.' NBER Working Paper No. 19925.
- Heckman, J. and Navarro, S. (2007) 'Dynamic discrete choice and dynamic treatment effects.' *Journal of Econometrics*, vol. 136 (2), pp. 341-396.
- Heckman, J. and Smith, J. (1998) 'Evaluating the welfare state.' In: Strom, S. (ed.), *Econometrics and Economic Theory in the Twentieth Century: The Ragnar Frisch Centennial Symposium*, Cambridge University Press, New York, pp. 241–318.
- Heckman, J. and T. Kautz (2012) 'Hard evidence on soft skills.' *Labour Economics*, vol.19 (4), pp.451–464.
- Heckman, J. and Vytlacil, E. (2005) 'Structural equations, treatment effects and econometric policy evaluation.' *Econometrica*, vol. 73 (3), pp. 669–738.
- Heckman, J., LaLonde, R. and Smith, J. (1999) 'The economics and econometrics of active labor market programs.' in O. Ashenfelter and D. Card, (eds.) *Handbook of Labor Economics*, vol. 3, North-Holland, Amsterdam.
- Heckman, J., Lochner, L. and Todd, P. (2006) 'Earnings functions, rates of return and treatment effects: The Mincer equation and beyond.' In *Handbook of the Economics of Education*, vol. 1. Elsevier, pp. 307-458.
- Heckman, J., Lochner, L. and Todd, P. (2008) 'Earnings Functions and Rates of Return.' *National Bureau of Economic Research*, WP no. 13780.
- Heckman, J., Lochner, L. and Todd, P. (2003). 'Fifty Years of Mincer Earnings Regressions.' NBER Working Paper No. 9732.
- Heckman, J., Pinto, R. and Savelyev, P. (2013) 'Understanding the Mechanisms through Which an Influential Early Childhood Program Boosted Adult Outcomes.' *American Economic Review*, vol. 103(6), pp.2052-86.
- Heckman, J., Stixrud, J. and Urzua, S. (2006) 'The Effects of Cognitive and Noncognitive Abilities on Labor market Outcomes and Social Behaviour.' NBER Working Paper No. 12006
- Heckman, J., Tobias, J. and Vytlacil, E. (2001) 'Four parameters of interest in the evaluation of social programs.' *Southern Economic Journal*, vol. 68 (2) pp. 210-223.
- Heckman, J., Urzua, S. and Vytlacil, E. (2006) 'Understanding instrumental variables in models with essential heterogeneity.' *Review of Economics and Statistics*, vol. 88 (3), pp. 389-432.
- Hendricks, M. (2014) 'Does it pay to pay teachers more? Evidence from Texas' *Journal of Public Economics* vol. 109 pp. 50-63
- Hill, A. (2015) 'The Girl Next Door: The Effect of Opposite Gender Friends on High School Achievement.' *American Economic Journal: Applied Economics*, vol.7 (3), pp.147-77.
- Hill, H., Charalambous, C. and Chin, M. (2015) 'Teacher characteristics and student learning: a comprehensive assessment.' Mimeo. Harvard Graduate School of Education
- Holmlund, H., Lindahl, M., and Plug, E. (2015) 'The causal effect of parents' schooling on children's schooling: a comparison of estimation methods.' *Journal of Economic Literature* (forthcoming).
- Holmlund, H., McNally, S. and Viarengo, M. (2010) 'Does Money Matter for Schools?' *Economics of Education Review*, vol. 29 (6), pp. 1154-64.

- Howard-Jones, P. and Washbrook, L. (2011) 'Educational investment: Interrelating neuroscientific, educational and economic perspectives.' CUBEC Policy report 11/02, University of Bristol. <http://www.bristol.ac.uk/media-library/sites/cubec/migrated/documents/pr2.pdf>
- Hoxby, C. (2000a) 'Does Competition among Public Schools Benefit Students and Taxpayers?' *American Economic Review*, Vol. 90, No. 5, pp. 1209-1238
- Hoxby, C. (2000b) 'The Effects of Class Size on Student Achievement: New Evidence from Population Variation' *Quarterly Journal of Economics*, vol. 115 (4), pp. 1239-1285.
- Hoxby, C. (2001) 'If families matter most, where do schools come in?' In Moe, T. (ed) *A Primer on America's Schools*. Hoover Institution Press, Stanford.
- Hoxby, C. (2003a) 'School choice and school competition: evidence from the United States.' *Swedish Economic Policy Review*, vol. 10, pp. 11-67.
- Hoxby, C. (2003b) Introduction, in Hoxby, C. (ed.), *The Economics of School Choice*, Chicago: University of Chicago Press.
- Hoxby, C. (2003c) 'School choice and school productivity: could school choice be a tide that lifts all boats?' In Hoxby C (ed.), *The Economics of School Choice*, Chicago: University of Chicago Press.
- Hoxby, C. and Avery, C. (2012) 'The Missing "One-Offs": The Hidden Supply of High-Achieving, Low Income Students' NBER Working Paper 18586. Forthcoming Brookings Papers on Economic Activity
- Hoxby, C. and Murarka, S. (2009) 'Charter Schools in New York City: Who Enrolls and How They Affect Student Achievement.' NBER Working Paper 14852.
- Hoxby, C. and Rockoff, J. (2004) 'The Impact of Charter Schools on Student Achievement.' Mimeo, Harvard.
- Hoxby, C. and Turner, S. (2014) 'Expanding College Opportunities for High-Achieving, Low Income Students' SIEPR Discussion Paper No. 12-014, Stanford.
- Hoxby, C. and Weingarth, G. (2005) 'Taking race out of the equation: School reassignment and the structure of peer effects.' Mimeo.
- Ichino, A., Karabarbounis, L. and Moretti, E. (2009) 'The political economy of intergenerational income mobility.' Discussion paper no. 4767, Institute for the Study of Labor (IZA), Bonn.
- Jackson, C.K. (2010) 'A little now for a lot later: an evaluation of a Texas advanced placement incentive program.' *Journal of Human Resources* vol. 45(3), pp.591-639.
- Jackson, C.K. (2013) 'Non-Cognitive Ability, Test Scores, and Teacher Quality: Evidence from 9th Grade Teachers in North Carolina.' NBER Working Paper No. 18624.
- Jackson, C.K. and Bruegmann, E. (2009) 'Teaching Students and Teaching Each Other: The Importance of Peer Learning for Teachers' *American Economic Journal: Applied Economics* vol 1:4, 85–108
- Jackson, C.K., Johnson, R. and Persico, C. (2014) 'The Effect of School Finance Reforms on the Distribution of Spending, Academic Achievement and Adult Outcomes' NBER WP 20118
- Jackson, C.K., Rockoff, J. and Staiger, D. (2014) 'Teacher Effects and Teacher-Related Policies.' *Annual Review of Economics*, August 2014, pp. 801-825.
- Jacob B. (2005) 'Accountability, incentives and behavior: the Impact of high-stakes testing in the Chicago public schools.' *Journal of Public Economics*, vol. 89(6), pp.761-796.
- Jacob, B. (2010) 'Do principals fire the worst teachers?' NBER Working Paper No. 15715.
- Jacob, B. and Levitt, S. (2003) 'Rotten apples: an investigation of the prevalence and predictors of teacher cheating.' *The Quarterly Journal of Economics*, vol.118(3), pp.843-877.
- Jacob, B. and Lefgren, L. (2008) 'Can principals identify effective teachers? Evidence on subjective performance evaluation in education.' *Journal of Labor Economics*, vol. 25 pp. 101–136
- Jacob, B., Lindy, B., Rockoff, J., Rosen, R. and Taylor, E. (2015) Teacher Applicant Hiring and Teacher Performance: Evidence from DC Public Schools. Mimeo.

- Jäntti, M., Bratsberg, B., Røed, K., Raaum, O., Naylor, R., Eva, O., Björklund, A. and Eriksson, T. (2006) 'American exceptionalism in a new Light: a comparison of intergenerational earnings mobility in the Nordic countries, the United Kingdom and the United States.' Discussion paper no. 1938, Institute for the Study of Labor (IZA), Bonn.
- Jensen, R. (2010) 'The (Perceived) Returns to Education and the Demand for Schooling' *Quarterly Journal of Economics* vol. 125 (2): 515-548.
- Kane, T. (2006) 'Public Intervention in Post-Secondary Education.' *Handbook of the Economics of Education*, vol. 2, ed. E. Hanushek and F. Welch.
- Kane, T. and Staiger D. (2012) 'Gathering Feedback for Teaching: Combining High-Quality Observations with Student Surveys and Achievement Gains.' MET Project Research Paper, Bill & Melinda Gates Foundation, Seattle WA.
- Kane, T. and Staiger, D. (2008) 'Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation,' NBER Working Paper No. 14607.
- Kane, T., McCaffrey D., Miller, T. and Staiger D. (2013) 'Have We Identified Effective Teachers? Validating Measures of Effective Teaching Using Random Assignment'. Seattle, WA: Bill & Melinda Gates Foundation.
- Kane, T., Rockoff, J. and Staiger, D. (2008) 'What does certification tell us about teacher effectiveness? Evidence from New York City.' *Economics of Education Review*, vol.27 (6), pp. 615–631
- Kane, T. and Staiger, D. (2001) 'Improving School Accountability Measures.' NBER Working Paper No. 8156.
- Katz, L. and Autor, D. (1999) 'Changes in the wage structure and earnings inequality.' In O. Ashenfelter and D. Card (eds.) *Handbook of Labour Economics*, vol. 3A, North-Holland:Amsterdam.
- Kane, T., Taylor, E., Tyler, J. and Wooten, A. (2010) 'Identifying effective classroom practices using student achievement data' *Journal of Human Resources*, 46(3), pp. 587-613. 2010.
- Kautz, T., Heckman, J., Diris, R., ter Weel, B. and Borghans, L. (2014) 'Fostering and Measuring Skills: Improving Cognitive and Non-cognitive Skills to Promote Lifetime Success.' OECD Education Working Papers no. 110.
- Kerr, S. (1975) 'On the folly of rewarding A, while hoping for B.' *Academy of Management Journal*, vol. 18, pp. 769 – 783.
- Ketel, N., Leuven, E., Oosterbeek, H., and Van der Klaauw, B. (2015) 'The returns to medical school: Evidence from admission lotteries.' *American Economic Journal: Applied Economics* . (Forthcoming)
- Kluve, J., Card, D., Fertig, M., Gora, M., Jacobi, L., Jensen, P., Leetmaa, R., Nima, L., Patacchini, E., Schaffner, S., Schmidt, C., van der Klaauw, B. and Weber, A. (2007) *Active labor market policies in Europe: Performance and perspectives*. Berlin: Springer.
- Koedel, C. and Betts, J. (2011) 'Does Student Sorting Invalidate Value-Added Models of Teacher Effectiveness? An Extended Analysis of the Rothstein Critique.' *Education Finance and Policy*, vol. 6(1), pp.18-42.
- Koedel, C., Mihaly, K. and Rockoff, J. (2015) 'Value-Added Modeling: A Review.' *Economics of Education Review*, vol.47, pp.180-195.
- Koning, P. and van der Wiel (2012) 'School responsiveness to quality rankings: An empirical analysis of secondary education in the Netherlands.' *De Economist*, vol. 160(4), pp. 339 – 355.
- Koop, G. and Tobias, J. (2004) 'Learning about heterogeneity in returns to schooling.' *Journal of Applied Econometrics*, vol. 19, pp. 827-849.
- Kraft, M. and Blazar, D. (2014) 'Improving teacher practice: Experimental evidence on individualized teacher coaching.' Working Paper, Harvard
- Kraft, M. and Papay, J. (2014) 'Can Professional Environments in Schools Promote Teacher Development? Explaining Heterogeneity in Returns to Teaching Experience.' *Educational Effectiveness and Policy Analysis*. vol.36(4) pp.476-500.
- Kraft, M. 'Teacher layoffs, teacher quality and student achievement: Evidence from a discretionary layoff policy.' *Education Finance and Policy*. Forthcoming.

- Kremer, M., Miguel, E. and Thornton, R. (2009) 'Incentives to Learn.' *Review of Economics and Statistics*, vol. 91 (3), pp.437-456.
- Krueger, A. (2003) 'Economic Considerations and Class Size.' *Economic Journal*, vol. 113, pp. F34-63.
- Krueger, A., (1999) "Experimental Estimates of Education Production Functions.' *Quarterly Journal of Economics*, vol. 114 , pp. 497–532.
- Ladd, H. and Sorensen, L. (2015) 'Do Masters Degrees Matter: Advanced degrees, career paths and the effectiveness of teachers.' Mimeo, Duke.
- LaLonde, R. (1986) 'Evaluating the econometric evaluations of training programs with experimental data.' *American Economic Review*, vol. 76(4), pp. 604 – 620.
- Lavy, V. (2002) 'Evaluating the effect of teachers' group performance incentives on pupil achievement.' *Journal of Political Economy*, vol. 110(6), pp.1286-1317.
- Lavy, V. (2008) 'Does raising the principal's wage improve the school's outcomes? Quasi-experimental evidence from an unusual policy experiment in Israel.' *Scandinavian Journal of Economics*, vol. 110(4), pp. 639-662.
- Lavy, V. (2009) 'Performance pay and teachers' effort, productivity, and grading ethics,' *American Economic Review*, vol.99 (5), pp. 1979–2011.
- Lavy, V. (2010) 'Effects of free choice among public schools.' *Review of Economic Studies*, vol. 77(3), pp. 1164 – 1191.
- Lavy, V., Ebenstein, A. and Roth, S. (2014) 'The Long Run Human Capital and Economic Consequences of High-Stakes Examinations.' NBER Working Papers 20647, National Bureau of Economic Research, Inc.
- Lazear, E. (2001) 'Educational production.' *Quarterly Journal of Economics*, vol.116, pp. 777–803.
- Lazear, E.P. (2000) 'Performance pay and productivity.' *The American Economic Review*, vol. 90 (5), pp. 1346-1361.
- Lemieux, T. (2006) 'The 'Mincer equation' Thirty Years after Schooling, Experience, and Earnings.' In *Jacob Mincer: A Pioneer of Modern Labor Economics*, Grossbard, S. (ed.), Springer: New York, pp. 127–145.
- Leuven, E. and Oosterbeek, H. (2007) 'The effectiveness of human capital policies for disadvantaged groups in the Netherlands.' In Woessmann, L. and Peterson, P. (eds.) *Schools and the Equal Opportunity Problem*, MIT Cambridge.
- Leuven, E., Lindahl, M., Oosterbeek, H. and Webbink, D. (2007) 'The effect of extra funding for disadvantaged students on achievement' *Review of Economics and Statistics* 89(4) , 721-736
- Leuven, E., Oosterbeek, H. and van der Klaauw, B. (2010) 'The Effect of Financial Rewards on Students' Achievements: Evidence from a Randomized Experiment.' *Journal of the European Economic Association*, vol.8 (6), pp.1243-1265.
- Levitt, S. and List, J. (2009) 'Field experiments in economics: The past, the present, and the future' *European Economic Review* vol. 53, pp. 1–18.
- Levitt, S., List, J. and Sadoff, S. (2010) 'The Effect of Performance-Based Incentives on Educational Achievement: Evidence from a Randomized Experiment.' Working Paper.
- Levitt, S., List, J., Neckermann, S. and Sally Sadoff (2012) 'The Behavioralist Goes to School: Leveraging Behavioral Economics to Improve Educational Performance.' NBER Working Paper 18165.
- Lindo, J., Swensen, I. and Waddell, G. (2012) 'Are Big-Time Sports a Threat to Student Achievement?' *American Economic Journal: Applied Economics*, vol.4 (4), pp.254-274.
- Lochner, L. and Moretti, E. (2004) 'The Effect of Education on Criminal Activity: Evidence from Prison Inmates, Arrests, and Self-Reports.' *American Economic Review*, vol. 94 (1), pp. 155–89.
- Loeb, S., Horng, E. and Klasik, D. (2010) 'Principal's time use and school effectiveness.' *American Journal of Education*, vol. 116(4), pp. 491-523.
- Loeb, S., Miller, L. and Wyckoff, J. (2015) 'Performance Screens for School Improvement: The Case of Teacher Tenure Reform in New York City.' *Educational Researcher*, forthcoming.

- Machin, S. and Veroit, J. (2011) 'Changing School Autonomy: Academy Schools and Their Introduction to England's Education.' CEE Discussion Paper 123.
- Machin, S., Marie, O. and Vujic, S. (2011) 'The crime reducing effect of education.' *Economic Journal*, vol. 121, pp. 463-484.
- Machin, S., McNally, S. and Silva, O. (2007) 'New Technology in Schools: Is there a Payoff?' *Economic Journal*, vol. 117, pp. 1145-67.
- Meghir, C. and Rivkin, S. (2011) 'Econometric Methods for Research in Education' in Handbook of the Economics of Education vol. 3 (eds) Hanushek, E., Machin, S. and Woessmann, L. North-Holland Amsterdam
- MaCurdy, T. and Mroz, T. (1995) 'Estimating macro effects from repeated cross-sections', mimeo, Stanford University.
- Manski, C. (1993) 'Identification of endogenous social effects: The reflection problem.' *Review of Economic Studies*, vol.60, pp. 531-542.
- Master, B., Loeb, S. and Wyckoff, J. (2014) 'Learning that Lasts: Understanding Variation in Teachers' Effects on Students' Long-Term Knowledge.' CEPA Working Paper Stanford
- Maurin, E., and McNally, S. (2008) 'Vive la révolution! Long-term educational returns of 1968 to angry students.' *Journal of Labor Economics*, vol. 26 (1), pp.1-34.
- Mayer, S. and Lopoo, L. (2008) 'Government spending and intergenerational mobility.' *Journal of Public Economics*, vol. 92, pp.139-158.
- Mcintosh, S. (2006) 'Further Analysis of the Returns to Academic and Vocational Qualifications.' *Oxford Bulletin of Economics and Statistics*, vol.68 (2), pp. 305-49
- McPherson, M. and Schapiro, M. (2006) 'US Higher Education Finance.' *Handbook of the Economics of Education*, vol. 2, ed. E. Hanushek and F. Welch.
- Meghir, C. and Palme, M. (2005) 'Educational Reform, Ability, and Family Background' *American Economic Review*, vol. 95 (1), pp. 414 - 424.
- Metcalfe, R., Burgess, S., and Proud, S. (2014) 'Pupils' effort and educational attainment: Using the timing of the World Cup to vary the value of leisure.' mimeo, CMPO.
- Mihaly, K., McCaffrey, D., Staiger, D. and Lockwood, J. (2013) 'A Composite Estimator of Effective Teaching.' MET Project Research Paper, Bill & Melinda Gates Foundation, Seattle, WA.
- Milligan, K., Moretti, E. and Oreopoulos, P. (2004) 'Does education improve citizenship? Evidence from the United States and the United Kingdom.' *Journal of Public Economics*, vol. 88 (9-10), pp. 1667-1695.
- Mincer, J. (1974) *Schooling, Experience and Earnings*. Columbia University Press for National Bureau of Economic Research, New York.
- Moretti, E. (2011) 'Local Labor Markets.' In: Card, D., Ashenfelter, O. (eds.), *Handbook of Labor Economics*, vol. 4B. Elsevier, Amsterdam.
- Mueller, G. and Plug, E. (2006) 'Estimating the effect of personality on male and female earnings.' *Industrial and Labor Relations Review*, vol. 60 (1), pp. 3-22.
- Muralidharan, K. and Sundararaman, V. (2011) 'Teacher performance pay: experimental evidence from India.' *Journal of Political Economy*, vol.119 (1), pp.39-77.
- Naylor, R., Smith, J. and Telhaj, S. (2015) 'Graduate returns, degree class premia and higher education expansion in the UK' mimeo, University of Warwick
- Neal, D. (2011) 'The design of performance pay in education.' In Hanushek, E., Machin, S. and Wossmann, L. (eds.) *Handbook of the Economics of Education*, vol.4. Amsterdam: North-Holland.
- Neal, D. and Schanzenbach, D. (2007) 'Left behind by design: proficiency counts and testbased accountability.' NBER Working Paper No. 13293, NBER: Cambridge MA.
- Nechyba, T. (2003a) 'Introducing school choice into multi-district public school systems.' in Hoxby, C. (ed.), *The Economics of School Choice*, Chicago: University of Chicago Press.

- Nechyba, T. (2003b) 'School finance, spatial income segregation and the nature of communities.' *Journal of Urban Economics*, vol.54 (1), pp. 61-88.
- Nechyba, T. (2006) 'Income and peer quality sorting in public and private schools.' In Hanushek, E. and Welch F.,(eds.) *Handbook of Economics of Education*, Amsterdam: North Holland.
- Neilson, C. and Zimmerman, S. (2014) 'The Effect of School Construction on Test Scores, School Enrollment, and Home Prices.' *Journal of Public Economics*, vol. 120, pp. 18 – 31
- Neumark, D. and Simpson, H. (2015) 'Place-Based Policies' in *The Handbook of Regional and Urban Economics*, vol 5, eds. Gilles Duranton, J. Vernon Henderson and William C. Strange.
- OECD (2012) 'Does money buy strong performance in PISA?' PISA in Focus, 13.  
<http://www.oecd.org/pisa/pisaproducts/pisainfocus/49685503.pdf>
- OECD (2012) 'Public and Private Schools: How Management and Funding Relate to their Socio-economic Profile' OECD Publishing <http://dx.doi.org/10.1787/9789264175006-en>
- OECD (2013), OECD Skills Outlook 2013: First Results from the Survey of Adult Skills, OECD Publishing.  
<http://dx.doi.org/10.1787/9789264204256-en>
- Olds, D., Kitzman, H., Hanks, C., Cole, R., Anson, E., Sidora-Arcoleo, K., Luckey, D., Henderson Jr., C., Holmberg, J., Tutt, R., Stevenson, A., and Bondy, J. (2007) 'Effects of nurse home visiting on maternal and child functioning: age-9 follow-up of a randomized trial.' *Pediatrics*, vol.120 (4), pp.832–845.
- Oosterbeek, H. (2013) 'The financing of adult learning' EENEE Analytical Report No. 15.  
<http://www.eenee.de/eeneeHome/EENEE/Analytical-Reports.html>
- Oosterbeek, H. and Webbink, D. (2007) 'Wage effects of an extra year of basic vocational education.' *Economics of Education Review*, vol. 26, pp. 408–419.
- Oreopoulos, P. (2007) 'Do Dropouts Drop Out Too Soon? Wealth, Health, and Happiness from Compulsory Schooling.' *Journal of Public Economics*, vol. 91 (11–12), pp. 2213–2229.
- Oreopoulos, P. and Petronijevic, U. (2013) 'Making College Worth It: A Review of Research on the Returns to Higher Education.' NBER Working Paper No. 19053
- Oreopoulos, P. and Salvanes, K. (2009) 'How large are returns to schooling? Hint: Money isn't everything.' NBER Working Paper No. 15339.
- Oreopoulos, P. and Salvanes, K. (2011) 'Priceless: The Nonpecuniary Benefits of Schooling.' *Journal of Economic Perspectives*, vol. 25 (1), pp. 159–184.
- Oreopoulos, P., Page, M. and Stevens, A. (2006) 'Does human capital transfer from parent to child? The intergenerational effects of compulsory schooling.' *Journal of Labor Economics*, vol.24 (4), pp.729–760.
- Palacios-Huerta, I. (2003) 'An empirical analysis of the risk properties of human capital returns.' *American Economic Review*, vol. 93 (3), pp. 948–964.
- Papay, J. and Kraft M. (2015) 'Productivity returns to experience in the teacher labor market: Methodological challenges and new evidence on long-term career improvement.' *Journal of Public Economics*, forthcoming.
- Papay, J., Taylor, E., Tyler, J. and Laski, M. (2015) 'Learning Job Skills from Colleagues at Work: Evidence from a Field Experiment Using Teacher Performance Data' mimeo.
- Pathak, P.A. and Sonmez, T. (2008) 'Levelling the playing field: sincere and sophisticated players in the Boston mechanism.' *American Economic Review*, vol. 98(4), pp. 1636–52.
- Pekkarinen, T., Uusitalo, R. and Kerr, S. (2009) 'School tracking and intergenerational income mobility: Evidence from the Finnish comprehensive school reform' *Journal of Public Economics*, vol. 93(7-8), pp. 965-973.
- Piketty, T. (2004) 'L'impact de la taille des classes et de la ségrégation sociale sur la réussite scolaire dans les écoles françaises : une estimation à partir du panel primaire 1997.' mimeo.
- Psacharopoulos, G. (2006) 'The Value of Investment in Education: Theory, Evidence, and Policy.' *Journal of Education Finance*, vol. 32 (2), pp. 113-136.

- Reback, R. (2008) 'Teaching to the rating: school accountability and the distribution of student achievement.' *Journal of Public Economics*, vol.92(5-6), pp. 1394-1415.
- Reback, R., Rockoff, J. and Schwartz, H. (2014) 'Under Pressure: Job Security, Resource Allocation, and Productivity in Schools Under NCLB' *American Economic Journal: Economic Policy* vol. 6(3): 207-41.
- Reis, A., Nunes, L. and Seabra, C. (2015) 'The publication of school rankings: a step toward increased accountability?' *Economics of Education Review*, (forthcoming). doi:10.1016/j.econedurev.2015.07.008
- Rimfeld, K., Kovas, Y., Dale, P. and Plomin, R. (2015) 'Pleiotropy across academic subjects at the end of compulsory education.' *Nature, Scientific Reports*, vol. 5 Article number: 11713.
- Rivkin, S., Hanushek, E. and Kain, J. (2005) 'Teachers, schools, and academic achievement.' *Econometrica*, vol. 73, pp. 417-458.
- Rockoff, J. (2004) 'The impact of individual teachers on student achievement: Evidence from panel data.' *American Economic Review*, vol. 94, pp. 247-252.
- Rockoff, J. (2009) 'Field Experiments in Class Size from the Early Twentieth Century.' *Journal of Economic Perspectives*, vol. 23(4), pp. 211-230.
- Rockoff, J. and Turner, L. (2010) 'Short Run Impacts of Accountability on School Quality.' *American Economic Journal: Economic Policy*, vol. 2(4) pp.119-147.
- Rockoff, J., Jacob, B., Kane, T. and Staiger, D. (2011) 'Can You Recognize an Effective Teacher When You Recruit One?' *Education Finance and Policy*, vol. 6(1), pp.43-74.
- Rockoff, J., Staiger D., Kane T., and Taylor E. (2012) 'Information and Employee Evaluation: Evidence from a Randomized Intervention in Public Schools.' *American Economic Review*, vol.102 (7), pp.3184-3213.
- Ronfeldt, M., Lankford, H., Loeb, S. and Wyckoff, J. (2011) 'How Teacher Turnover Harms Student Achievement' *American Educational Research Journal* vol. 50(1) pp. 4-36
- Rothstein, J. (2006) "Good Principals or Good Peers? Parental Valuation of School Characteristics, Tiebout Equilibrium, and the Incentive Effects of Competition among Jurisdictions." *American Economic Review*, 96(4): 1333-1350.
- Rothstein, J. (2009) 'Student Sorting and Bias in Value-Added Estimation: Selection on Observables and Unobservables.' *Education Finance and Policy*, vol. 4(4), pp.537-71.
- Rothstein, J. (2010) 'Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement.' *Quarterly Journal of Economics* vol. 125(1), pp. 175-214 52.
- Rothstein, J. (2014) 'Revisiting the Impacts of Teachers.' UC-Berkeley Working Paper.
- Rothstein, J. (2015) 'Teacher Quality Policy When Supply Matters.' *American Economic Review*, vol. 105(1), pp. 100-130.
- Rothstein, J. and Mathis, W. (2013) 'Review of Two Culminating Reports from the MET Project.' Published online in Think Tank Review, National Education Policy Center, Boulder Colorado, January.
- Rouse, C., Hannaway, J., Goldhaber, D. and Figlio, D. (2013) 'Feeling the Florida Heat? How Low-Performing Schools Respond to Voucher and Accountability Pressure.' *American Economic Journal: Economic Policy* vol. 5(2) pp. 251-81.
- Sacerdote, B. (2011) 'Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?' in *Handbook of Economics of Education* vol. 3 (eds) Hanushek, E., Machin, S. and Woessmann, L. Amsterdam: North-Holland
- Sacerdote, B. (2002) 'The nature and nurture of economic outcomes.' *American Economic Review*, vol. 92, pp. 344-348.
- Sacerdote, B., (2001) 'Peer effects with random assignment: Results for Dartmouth roommates.' *Quarterly Journal of Economics*, vol.116, pp. 681-704.
- Sandström, F. and Bergström, F. (2005) 'School Vouchers in Practice: Competition Will Not Hurt You.' *Journal of Public Economics*, vol. 89, pp. 351-80.



- Schneider, M. and Buckley, J. (2002) 'What Do Parents Want from Schools? Evidence from the Internet' *Educational Evaluation and Policy Analysis* Vol. 24, No. 2, pp. 133-144
- Slater, H., Davies, N, and Burgess, S. (2012) 'Do Teachers Matter? Measuring the Variation in Teacher Effectiveness in England' *Oxford Bulletin of Economics and Statistics* Vol. 74 (5) pp. 629–645
- Soderstrom, M. and Uusitalo, R. (2010) 'School Choice and Segregation: Evidence from an Admission Reform.' *Scandinavian Journal of Economics* vol. 112 (1), pp.55–76.
- Solon, G. (2004) 'A model of intergenerational mobility variation over time and place.' In Corak, M. (ed.), *Generational income mobility in North America and Europe*. Cambridge: Cambridge University Press, pp. 38–47.
- Steinberg, M. and Sartain, L. (2015). 'Does teacher evaluation improve school performance? Experimental evidence from Chicago's Excellence in Teaching Project.' *Education Finance and Policy*. Forthcoming
- Steinberg, M. and Sartain, L. (2014) 'Does better observation make better teachers?' *Education Next*, vol. 15(1), pp.70-76.
- Steiner, V. and Wrohlich, K. (2012) 'Financial Student Aid and Enrollment in Higher Education: New Evidence from Germany.' *Scandinavian Journal of Economics*, 114 (1) pp.124-147 .
- Taylor, E. and Tyler J. (2012) 'The effect of evaluation on teacher performance.' *American Economic Review*, vol. 102(7): 3628–3651
- Todd, P. and Wolpin, K. (2003) 'On the specification and estimation of the production function for cognitive achievement' *The Economic Journal*, 113, F3–F33
- Tyler, J., Taylor, E., Kane, T. and Wooten, A. (2010) 'Using student performance data to identify effective classroom practices.' *American Economic Review, Papers & Proceedings*, vol. 100 (2), pp. 256–260.
- Veugelers, R. and Del Rey, E. (2014) 'The contribution of universities to innovation, (regional) growth and employment.' EENEE Analytical Report 18.  
[http://www.eenee.de/dms/EENEE/Analytical\\_Reports/EENEE\\_AR18.pdf](http://www.eenee.de/dms/EENEE/Analytical_Reports/EENEE_AR18.pdf)
- Vigdor, J. and Nechyba, T. (2007) 'Peer effects in North Carolina public schools.' In Woessmann, L. and Peterson, P. (eds.) *Schools and the Equal Opportunity Problem*, The MIT Press, pp. 73–101.
- Vilsa E. and Fryer, R. (2014) 'Estimating the Returns to Urban Boarding Schools: Evidence from SEED.' *Journal of Labor Economics*, vol. 32 (1), pp. 65-93.
- Walker, I. and Zhu, Y. (2008) 'The College Wage Premium and the Expansion of Higher Education in the UK.' *Scandinavian Journal of Economics*, vol. 110, pp. 695-709.
- Walker, I. and Zhu, Y (2011) 'Differences by Degree: Evidence of the net financial rates of return to undergraduate study for England and Wales.' *Economics of Education Review*, vol.30, pp.1177-1186.
- Walker, J. and Brooks, C. (2014) 'Better Understanding through Data: Completion, Motivation, and Learning in Minnesota MOOCs.' SEI Case Study. <https://net.educause.edu/ir/library/pdf/SEI1402.pdf>
- Wiswall, M. (2013) 'The dynamics of teacher quality.' *Journal of Public Economics*, vol. 100, pp. 61-78.
- Woessman, L. (2007) 'School Accountability, Autonomy, Choice, and the Level of Student Achievement: International Evidence from PISA 2003.' OECD Education Working Papers, No. 13, OECD Publishing.
- Woessman, L. (2010) 'Cross-country evidence on teacher performance pay.' IZA Discussion Paper No. 510.
- Woessmann, L. (2003) 'Schooling Resources, Educational Institutions, and Student Performance: The International Evidence.' *Oxford Bulletin of Economics and Statistics* vol.65 (2) pp. 117-170.
- Woessmann, L. (2014) 'The Economic Case for Education.' EENEE Analytical Report No. 20, EENEE.  
<http://www.eenee.de/eeneeHome/EENEE/Analytical-Reports.html>
- Woessmann, L., Hanushek, E. and Zhang, L. (2011) 'General Education, Vocational Education, and Labor-Market Outcomes over the Life-Cycle.' NBER Working Paper 17504.
- Woessmann, L. (2007) 'International Evidence on School Competition, Autonomy, and Accountability: A Review.' *Peabody Journal of Education*, vol.82 (2-3), pp. 473-497.

- Wolter, S. and Ryan, P. (2011) 'Apprenticeship.' *Handbook of the Economics of Education*, vol. 3, edited by Eric A. Hanushek, Stephen Machin and Ludger Woessmann.
- Wong, M., Cook, T., and Steiner, P. (2009) 'No Child Left Behind: An interim evaluation of its effects on learning using two interrupted time series each with its own non-equivalent comparison series.' Northwestern University Institute for Policy Research, Northwestern University, Working Paper Series WP-09-11.
- Zimmerman, S. (2014) 'The Returns to College Admission for Academically Marginal Students.' *Journal of Labor Economics*, vol. 32 (4), pp. 711 – 754.