CENTRE FOR MARKET AND PUBLIC ORGANISATION

Research in Public Policy

ISSUE 15
WINTER 2012/13

IMPROVING EDUCATIONAL ATTAINMENT

EVIDENCE FROM AFRICA, THE US AND THE UK

Gender differences in maths achievement in sub-Saharan Africa

The case for Conditional Cash Transfers in Colombia

Community monitoring improves performance

bristol.ac.uk/cmpo

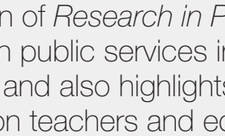
WINTER 2012/3

This edition of *Research in Public Policy* focuses on public services in developing countries, and also highlights recent research on teachers and educational outcomes in the UK and US.

In all too many cases, basic services such as education and health are poorly delivered in developing countries, but they are crucial for increasing living standards. In April 2012, a two-day conference held at CMPO brought together researchers working in this area. The conference served to highlight the similarity between the debate about what to do about public services in developing countries and the academic and policy debate in the UK, which has featured so heavily in CMPO research in the past. Among the issues considered were the role of financial incentives, conditionality in transfer programmes, evaluation of past reforms, the importance of political incentives and institutions.

There is a common focus on education in developing and developed countries. The first three articles from the conference explore education and schools. They look at gender differences in maths performance across 19 African countries, the impact of an extra year in high school on student achievement in Ghana, and community monitoring of local schools.

Our second set of articles cover research on teachers and teacher labour markets in the UK and US. It is widely acknowledged that teacher effectiveness is the single most important factor for raising educational achievement. These articles discuss new findings on teacher turnover and teacher pay and their effects on pupil performance. CMPO researchers also examine evidence on teacher effectiveness and argue that major changes to current teacher training are required to raise student attainment in the UK.



Helen Simpson and Sarah Smith

Editor:
Saskia Walcott

Produced by:
Centre for Market and Public Organisation
Bristol Institute of Public Affairs (BIPA)
University of Bristol
2 Priory Road, Bristol BS8 1TX

Tel: +44 (0)117 331 0799
Fax: +44(0)117 331 0705
Email: cmpo-admin@bristol.ac.uk

bristol.ac.uk/cmpo

CONTENTS

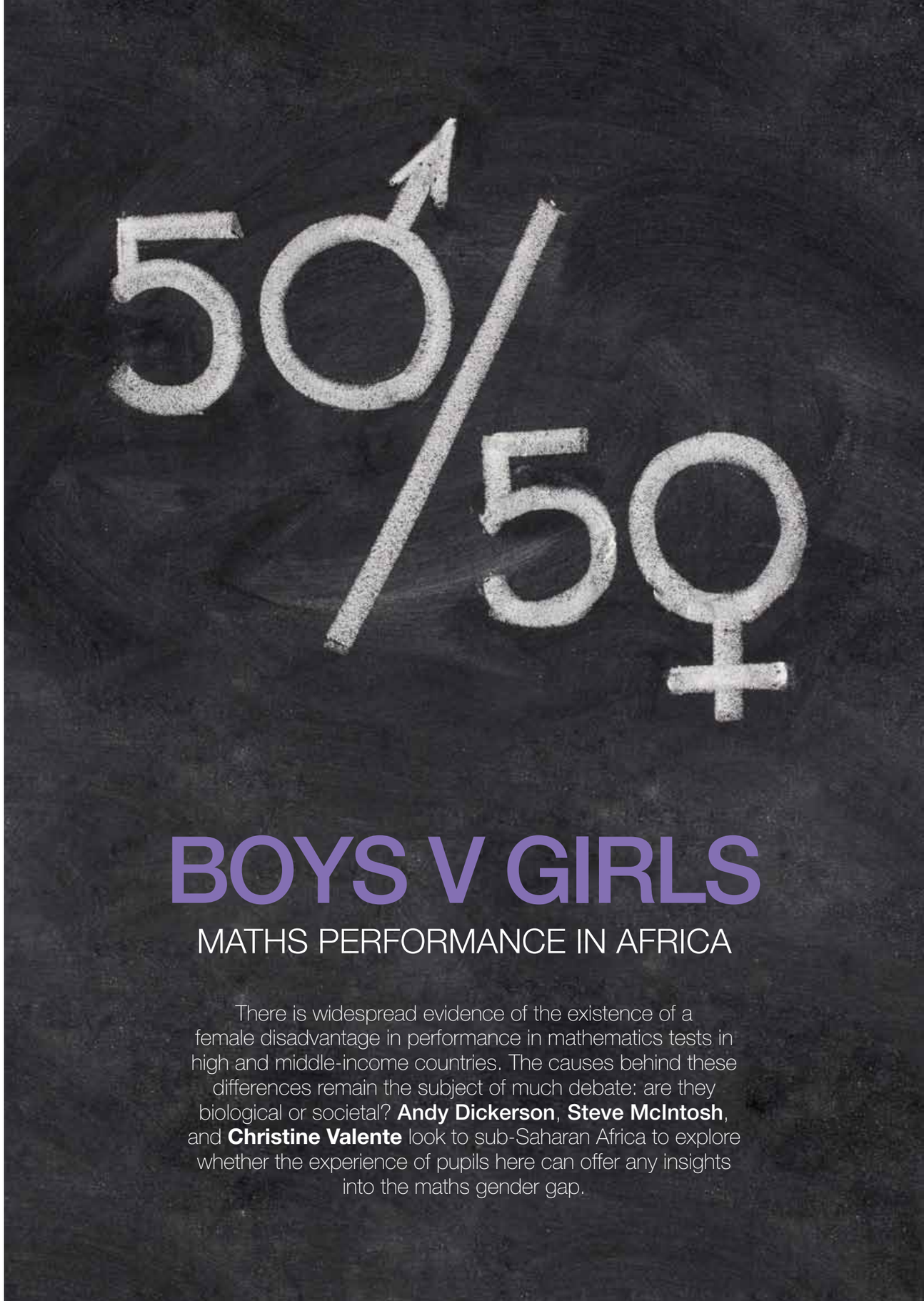
Keep updated

Visit the CMPO blog 'Viewpoint': **bristol.ac.uk/cmpo/blog**

Watch us on YouTube: **youtube.com/user/CMPObristol**

Follow us on Twitter: **@cmpobristol**

Listen to our podcasts: **bristol.ac.uk/cmpo/audio**



BOYS V GIRLS

MATHS PERFORMANCE IN AFRICA

There is widespread evidence of the existence of a female disadvantage in performance in mathematics tests in high and middle-income countries. The causes behind these differences remain the subject of much debate: are they biological or societal? **Andy Dickerson**, **Steve McIntosh**, and **Christine Valente** look to sub-Saharan Africa to explore whether the experience of pupils here can offer any insights into the maths gender gap.

In the literature focussing on developing countries, an increasingly important theme has been the study of female discrimination in access to education. Lack of data has largely precluded the study of gender differences in test scores and instead the focus has been on school enrolment and, to a lesser extent, grade completion.

However, it is also important to analyse gender differences in the acquisition of cognitive skills in schools, and especially so for mathematics. Indeed, skills learnt matter more for individual income and economic growth than mere attendance (Hanushek and Woessman 2008). Research also shows that *mathematic* cognitive skills are more predictive of income than literacy skills (Glewwe 1996 and Jolliffe 1998 for Ghana, Moll 1998 for South Africa), and, in developed countries where more research on the subject has been carried out, the gender gap in mathematics has been found to be a significant contributor to gender wage differentials (Paglin and Rufolo 1990).

Nature or nurture?

The debate surrounding the causes of the gender gap in mathematics has essentially opposed two types of arguments, one biological and one societal. The first potential explanation for a female disadvantage in maths tests is a biological one. The two most prominent genetic arguments are: gender-specific profiles in spatial and numerical abilities leading to a greater male aptitude for mathematics, and higher dispersion of male than female performance in quantitative and spatial ability, so that larger numbers of men have unusually high scores.

The second type of argument focuses on cultural or societal explanations for the gender gap. It could be that factors that influence mathematics tests scores vary systematically across the genders, with boys having better characteristics for producing high test scores.

In the context of sub-Saharan Africa, it could, for instance, be the case that parents send their sons to better schools than their daughters, for the same reason as the enrolment statistics show that they are more likely to send their sons to school in the first place. If the gender gap is due to girls having worse characteristics (e.g. they are enrolled in worse schools) than boys, then after accounting for differences in such characteristics, the gender gap should disappear.

The second possibility is that family, school and societal influences affect boys and girls differently. If this were the case, then girls would only perform worse than boys under unfavourable circumstances.

A closer look at the data shows that the overall gap hides a lot of variation between groups.

Overall, girls do worse than boys

In order to measure and analyse gender differences in maths performance, we use data from two pan-African surveys of educational quality, the Southern and Eastern African Consortium for Monitoring Educational Quality (SACMEQ), and the Program for the Analysis of Education Systems (PASEC) which, taken together, cover 19 western, southern and eastern African nations. In total, our sample covers more than 40,000 Grade 5 (PASEC) and Grade 6 (SACMEQ) pupils spread between more than 3,000 schools surveyed between 2000 and 2006. We merge these microdata with national-level information on gender equality and development from various sources, as well as with characteristics of the subnational region, which we calculate based on Demographic and Health Surveys microdata.

Overall, girls obtain scores that are below those of boys by 1.4 percentage points in SACMEQ and 0.6 percentage points in PASEC countries, when we do not control for differences in the characteristics of the boys and girls. After accounting for these differences, the gap is about 1.7 percentage points in SACMEQ countries and 2.1 percentage points in PASEC countries, i.e. 4.2 and 5.4 percent of the boys' mean in each sample, respectively.

The data allow us to investigate the gender gap between pupils in the same school, thus ruling out systematic differences in school quality between boys and girls, as well as to control for differences in individual access to schooling inputs within the classroom.

In fact, the gap *increases* when differences in household and school characteristics are controlled for because of the lower enrolment of girls relative to boys in sub-Saharan Africa. As a consequence of this gender difference in enrolment, the girls we observe in schools are more likely to have wealthier, better educated parents than their male classmates.

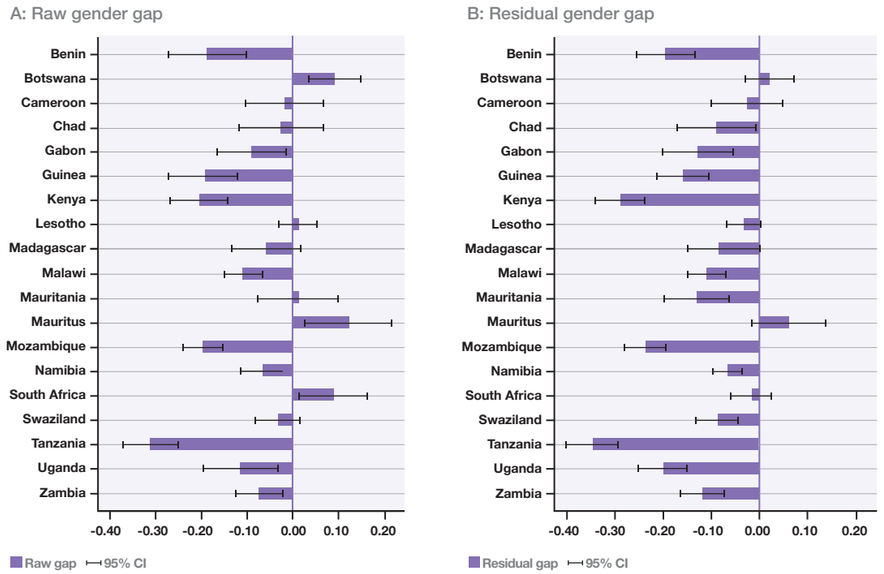
Significant variations between groups

At first glance, these findings may seem to suggest that girls indeed have some natural disadvantage in maths, since the gap is observed in Africa across both groups of countries, and is relatively stable both between the two groups of countries and compared to estimates for developed countries. However, a closer look at the data shows that the overall gap hides a lot of variation between groups.

Figure 1 plots the size of the gender gap in maths in each country, both in the raw data (left-hand side) and after accounting comprehensively for differences in family background, the schooling environment at home and at school (right-hand side). When comparing boys and girls with similar characteristics, girls obtain lower scores in 17 out of 19 countries, and the gender difference is statistically significant (i.e., meaningful) in 14 out of 19 countries.

When comparing boys and girls with similar characteristics, girls obtain lower scores in 17 out of 19 countries.

Figure 1
Raw and residual gender gap



Source: Authors' calculations using data from Southern and Eastern African Consortium for Monitoring Educational Quality (Botswana, Kenya, Lesotho, Malawi, Mauritius, Mozambique, Namibia, South Africa, Swaziland, Tanzania (Mainland), Uganda, Zambia) and Program for the Analysis of Education Systems (Benin, Cameroon, Chad, Gabon, Guinea, Madagascar, Mauritania). The gap is expressed in standard deviations (s.d.) of the test score distribution for each group of countries. SACMEQ: 1 s.d. corresponds to 15.4 percentage points. PASEC: 1 s.d. corresponds to 18.5 percentage points.

A closer inspection of Figure 1 shows that the magnitude of the gap varies widely:

- In Mauritius, girls obtain on average 1 percentage point more than boys but in Tanzania, they lag behind boys by as much as 5.3 percentage points.
- Characteristics of parents, schools and region have an effect on the size of the gender gap. The gap is twice as large among children of uneducated mothers in SACMEQ countries, and only exists among children of illiterate parents in PASEC countries.
- Among children taught maths by a male teacher in SACMEQ countries the gap between boys and girls in twice as large, yet completely disappears when the teacher is female in PASEC countries.
- The gender gap tends to decrease with GDP per capita, although once differences in fertility rates are accounted for, differences in GDP become irrelevant

Taken together, our findings show that, although girls do not have characteristics that are less conducive to performing well in maths tests than boys, family, school and societal influences affect boys and girls differently in sub-Saharan Africa. They also suggest that there is no simple genetic explanation to the gender gap in maths. More research is needed in order to test the effect of specific interventions aimed at improving female performance in maths, but our results indicate that there is room for effective policy interventions since the observed gap is not due to some immutable female disadvantage.

There is room for effective policy interventions.

This article is a summary of *'Do the Maths: An analysis of the gender gap in mathematics in Africa'*, by Andy Dickerson, Steve McIntosh and Christine Valente (2012) Mimeograph, University of Bristol and University of Sheffield.

Christine Valente is a Lecturer in Economics at the University of Bristol. Andy Dickerson is Professor of Economics and Steve McIntosh a Reader in Economics at the University of Sheffield.

Further reading

Fryer, R. and Levitt, S. (2010). *'An empirical analysis of the gender gap in mathematics'*, American Economic Journal: Applied Economics, 2(2), 210–40.

Glewwe, P. (1996). *'The Relevance of Standard Estimates of Rates of Return to Schooling for Education Policy: A Critical Assessment'*, Journal of Development Economics 51(2), 267–90.

Guiso, L., Monte, F., Sapienza, P., and Zingales, L. (2008). *'Culture, Gender and Math'*, Science, 320, pp.1164-1165.

Hanushek, E. and Woessman, L. (2008). *'The role of cognitive skills in economic development'*, Journal of Economic Literature, 46, 607-668.

Hedges, L. and Nowell, A. (1995). *'Sex Differences in Mental Test Scores, Variability, and Numbers of High-Scoring Individuals'*, Science, 269, 41-45.

Jolliffe, D. (1998). *'Skills, Schooling, and Household Income in Ghana'*, World Bank Economic Review, 12(1), 81-104.

Lewis, M. and Lockheed, M. (2008). *'Social exclusion and the gender gap in education'*, World Bank Policy Research Working Paper No. 4562.

Moll, P. (1998). *'Primary Schooling, Cognitive Skills and Wages in South Africa'*, Economica, 65, 263-284.

Paglin, Morton and Rufolo, Anthony M. (1990) *'Heterogenous Human Capital, Occupational Choice, and Male-Female Earnings Differences'* The Journal of Labor Economics, 8(1), 123-144.

CAN AN EXTRA YEAR AT HIGH SCHOOL IMPROVE ACHIEVEMENT?

For countries in the developing world improving the education outcomes of their citizens is a central and ongoing challenge. **Christopher Ksoll** and **Kim Lehrer** use the recent experience of Ghana to consider whether an additional year of schooling has a causal impact on pupil's achievement in their high school exams.

Estimating impacts of country-wide education reforms on achievement is often complicated by a lack of comparable control-groups or tests. Educational reforms typically affect the whole population so that observed changes in test scores might reflect changes in grading or test difficulty rather than learning. In this paper we analyse the impact of an additional year of high school in Ghana on high school leaving test scores where we observe a credible control group, Nigeria.

The high school leaving exams in both Ghana and Nigeria – the West African Senior School Certificate Examination – are identical for a subset of subjects in any given year. Ghana implemented a reform for the 2011 graduating cohort that extended senior high school from three to four years. The target knowledge assessed by the exam remained the same.

Compared with previous years, the cohort receiving an additional year of education performed much better on the tests, with approximately 70 percent passing the test in 2011 compared with just below 50 percent in 2008 and 2009. Receiving a credit pass in six (specific) subjects – the requirement for admittance to university – doubled from 15 percent to 30 percent. These are very large gains, statistically and economically significant.

We then turn to studying heterogeneous impacts. We find that there was no difference between achievement gains of girls versus boys. The extra year benefitted students from weaker schools most. At the individual level, the non-parametric evidence on the whole distribution of grades suggests that students in the middle of the distribution of test scores benefitted most. Both results suggest that

15% increase in passing rates.

The extra year benefitted students from weaker schools most.

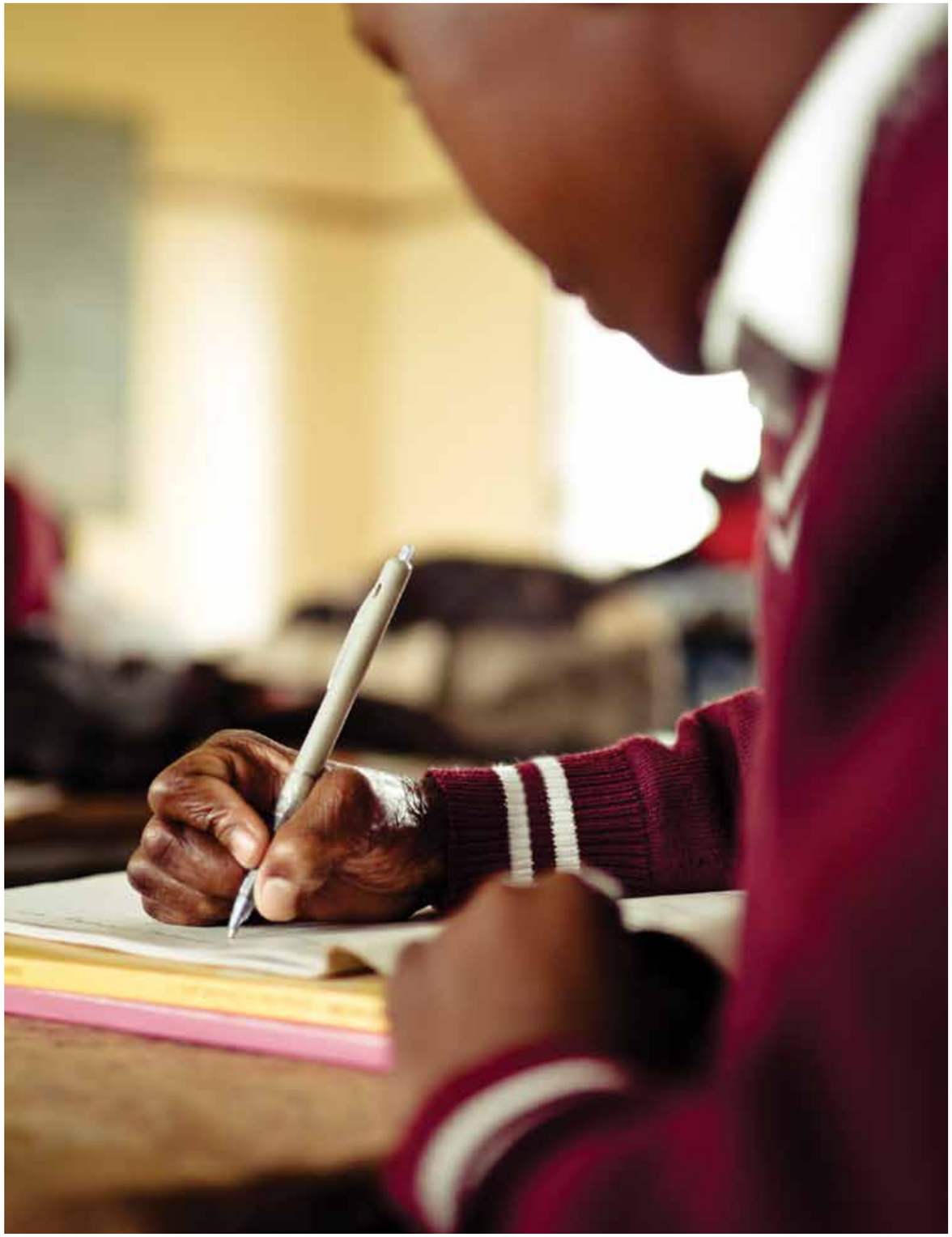
the better students in poorer schools might have made particular advances.

We then look at our control group, Nigeria. We find that test scores in Nigeria also rose during this period, though by much less. Taking into account that these changes might have been due to changes in the test difficulty reduces our estimate of the impact of the extra year somewhat. Even with this reduction, on average, the additional year that Ghana implemented for the 2011 graduating cohort lead to approximately 15 percent increase in passing rates in Ghana (when we use Nigeria's definition of passing) relative to previous graduation cohorts, and relative to Nigeria over this period. These achievement gains are significant and large.

Our research contributes to the literature on returns to education by focussing on the intermediate outcome, namely achievement. Achievement is rarely used to measure outcomes of large-scale policy reforms, and our analysis reveals both the power and pitfalls of this measure. On the one hand, it can be used to evaluate policies almost immediately. On the other hand, it requires a credible control group, as our results would have been upward biased if we had not been able to account for changes in test difficulty that also affected Nigeria.

Lastly, our research has substantial policy implications. The quality of students is decried by tertiary institutions as being quite low. We show that extending high school by one year can lead to students who are better prepared. This is a particularly relevant topic in developing countries such as Ghana, where, previous to the reform, half of the students did not pass the school leaving exams and over 60 percent did not achieve grades that permitted them to attend university.

Christopher Ksoll is Assistant Professor in Economics, School for International Development and Global Studies, University of Ottawa; and an associate member, Nuffield College and the Centre for the Study of African Economies (CSAE), University of Oxford. Kim Lehrer is a Research Officer at CSAE.



DOES COMMUNITY MONITORING IMPROVE PERFORMANCE?

Most countries have in place systems to monitor the quality of education. Despite heavy investment over the last two decades, Uganda continues to suffer low quality primary education. New research carried out in Uganda by the Economic Policy Research Centre (EPRC) in partnership with the Centre for the Study of African Economies (CSAE), suggests that giving local committees control over the criteria by which their school's performance is judged offers significant improvements to educational quality at little cost.

The EPRC and the CSAE implemented two different programmes, involving community monitoring, to improve schooling in Uganda. In both programmes, School Management Committees (SMCs), the organisations that allow parents, teachers and other community members to express their opinion about school performance, were trained how to use scorecards to help them monitor schools. In one programme these scorecards were designed by central organisations including NGOs and education authorities, whilst in the other they were designed by SMC participants themselves.

A hundred schools across Uganda were assigned randomly either to one of these two community monitoring programmes, or to a control group where no additional monitoring programme was implemented.

The criteria emphasised by SMCs for inclusion in scorecards were substantially different from those decided upon centrally. In particular, the SMC-designed scorecards paid little explicit attention to teacher absence, although underlying issues such as staff housing were frequently monitored. Also, the importance of parent contributions to learning appeared high on the list of criteria in the SMC-designed scorecards, further reflecting the need for parent-teacher participation and co-ordination.

We found that monitoring using community-designed scorecards made students and teachers significantly less likely to be absent from the classroom – by nine percent and 13 percent respectively – at the time of surprise visits by survey teams. There was also a significant increase in children's literacy and numeracy test scores. Pupil scores on National Assessment of Progress in Education exams, administered by Uganda National Examinations Board officials, improved by 19 percent of a standard deviation – enough to move the median student from the 50th to the 57.5th percentile.

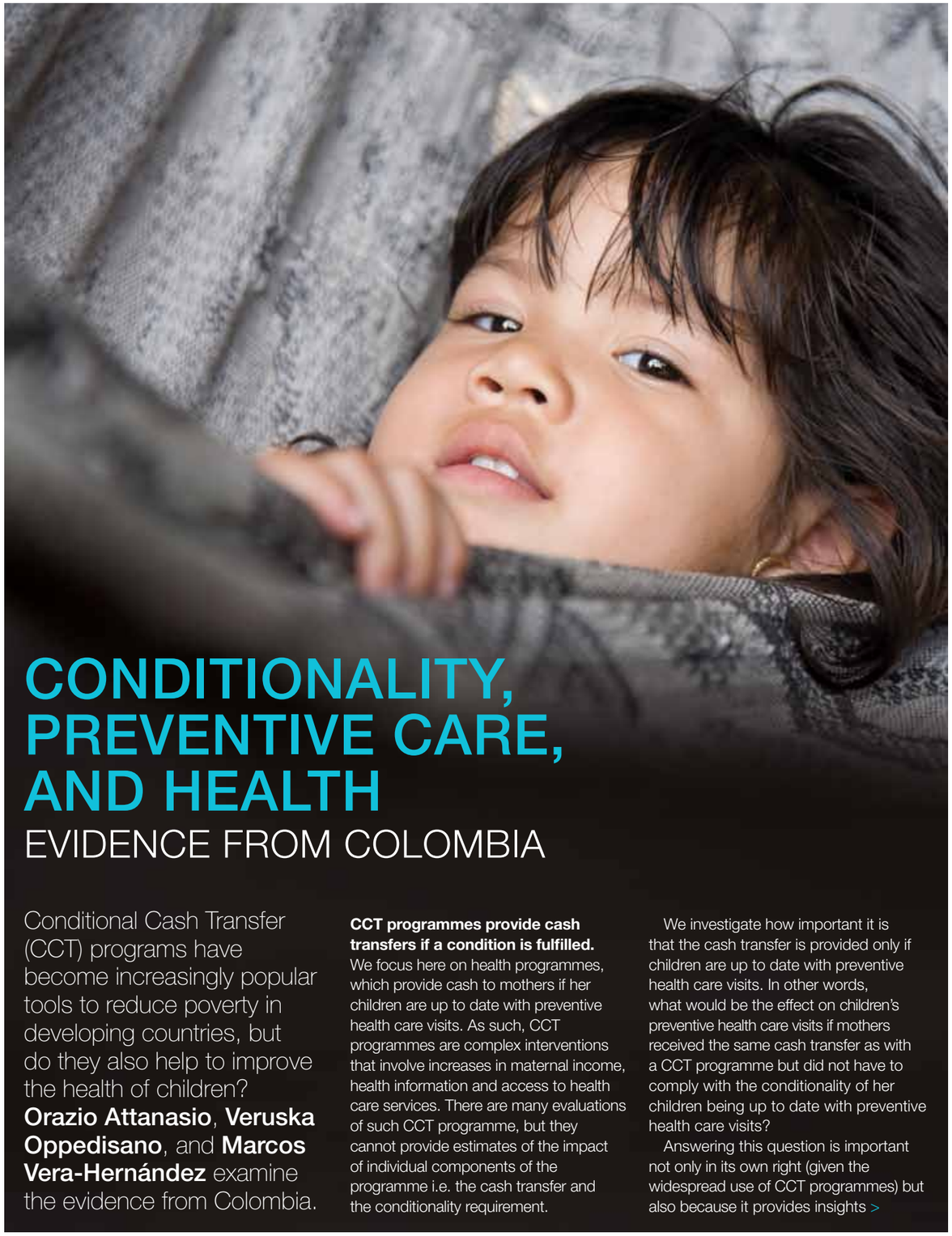
These improvements were achieved at relatively low cost and were not detected in schools that used centrally-designed scorecards. These results highlight the importance of participatory approaches when formulating community monitoring schemes and suggest that governments could achieve significant results by allowing school priorities to be chosen locally by consensus.

This article is based on a briefing paper for Improving Institutions for Pro-Poor Growth, University of Oxford. For further information please contact: iig.enquiries@economics.ox.ac.uk

Further reading

Andrew Zeitlin, Lawrence Bategeka, Madina Guloba, Ibrahim Kasirye and Frederick Mugisha (2011). *'Management and motivation in Ugandan primary schools: Impact evaluation final report'*. http://www.iig.ox.ac.uk/

Abigail Barr and Andrew Zeitlin (2011). *'Conflict of interest as a barrier to local accountability'*. CSAE Working Paper Series, No. 2011-13



CONDITIONALITY, PREVENTIVE CARE, AND HEALTH

EVIDENCE FROM COLOMBIA

Conditional Cash Transfer (CCT) programs have become increasingly popular tools to reduce poverty in developing countries, but do they also help to improve the health of children? **Orazio Attanasio**, **Veruska Oppedisano**, and **Marcos Vera-Hernández** examine the evidence from Colombia.

CCT programmes provide cash transfers if a condition is fulfilled. We focus here on health programmes, which provide cash to mothers if her children are up to date with preventive health care visits. As such, CCT programmes are complex interventions that involve increases in maternal income, health information and access to health care services. There are many evaluations of such CCT programme, but they cannot provide estimates of the impact of individual components of the programme i.e. the cash transfer and the conditionality requirement.

We investigate how important it is that the cash transfer is provided only if children are up to date with preventive health care visits. In other words, what would be the effect on children's preventive health care visits if mothers received the same cash transfer as with a CCT programme but did not have to comply with the conditionality of her children being up to date with preventive health care visits?

Answering this question is important not only in its own right (given the widespread use of CCT programmes) but also because it provides insights >

Conditionality is important for preventive care visits.

of why some mothers do not take their children to preventive health care visits. If the conditionality is important it implies that low perceived benefits of preventive care must be a reason why some children do not receive preventive care in developing countries. Hence, analysing the role of conditionality provides interesting insights on the barriers for preventive health care use in developing countries.

Secondly, we estimate the effect of preventive health care visits on children's health. Again, this is not only important in its own right, but it is also crucial to provide an interpretation to the effect of conditionality. If preventive care is ineffective then the conditionality would be a complete waste of time and an unconditional cash transfer would be preferable.

There are other reasons why CCT programs might be preferred to Unconditional Cash Transfer programs (UCT). As well as improving children's health, CCTs may have wider benefits within the community for example, by reducing infectious diseases. CCTs may also receive more support from tax-paying households in developing countries. However, this conditionality may also come at some cost. It might increase the cost of running the programme compared to UCTs, and it might also dissuade certain type of households who might benefit from participating in the programme.

We use data on children aged 0-3 years old from *Familias en Acción*, the CCT programme implemented by the Colombian government since 2002. We find that conditionality is important for preventive care visits: lack of conditionality reduces by just over one (1.16) the number of preventive care visits for young children.

In attempting to understand the mechanisms that drive preventive health care use, we find that the number of small children in the household is an important barrier for mothers to take their youngest children to preventive health care. In other words, the effect of the conditionality is much more important for mothers with more children than for mothers with fewer ones.

Another logical result is that the effect of the conditionality is smaller if the child had a sibling subject to the conditionality, as mothers take younger children to preventive care when they take their older ones. Childcare costs, broadly understood, emerge as an important cost of paying preventive care visits. Other factors such as availability of health services, travelling time and income do not significantly interact with the effect of conditionality.

Our findings show that preventive care has a causal effect on a child's health status. An additional check-up reduces by 4.7 percentage points a child's probability of suffering from acute diarrhoea, and increases by 14.6 percent of a standard deviation his height. These results provide evidence in favour of conditioning cash transfers with a health component in developing countries.

Orazio Attanasio is Professor of Economics; Veruska Oppedisano, is a Research Fellow and Marcos Vera-Hernández is a Senior Lecturer, at UCL.

PUBLIC WORKS PUSH UP WAGES

Most of the world's extremely poor live in rural areas, and at the bottom of the pyramid are landless agricultural workers subsisting on casual wage labour. Policies that can put upward pressure on agricultural wages are likely to be some of the most effective ways of improving the welfare of the poorest people on the planet. **Erlend Berg** analyses the effect of a large scale public works programme on wages in India.

In principle upward pressure on wages for the poorest in society could be achieved by introducing minimum wages, but enforcing a minimum wage rate is unrealistic in most developing countries. On the other hand, public works, which typically employ large numbers of unskilled workers to improve public infrastructure, may push wages up. If so, the welfare effects of public works programmes would reach well beyond the people who are directly employed by them.

In recent research, we look at a large-scale public works programme: the Indian government's National Rural Employment Guarantee (NREG). Using a decade's worth of wage data for >

250 Indian districts, we find that for each person-day of employment generated by NREG per capita in a district, agricultural wage rates increase by 1.6 percent. Since NREG generates approximately 3.3 person-days of employment per person per year on average, the implication is that the programme boosted real daily wages in India by 5.3 per cent (1.6×3.3 =5.3).

In principle, there are two ways in which a large-scale public employment programme can influence market wages. The first is by shifting up the demand curve for labour, thereby increasing its equilibrium price. The second is that the public goods generated by the programme may increase labour productivity and thus wages. We are not able to separate these two effects econometrically, but we concur with a recent World Bank report stating that 'the objective of asset creation runs a very distant second to the primary objective of employment generation.' Our impression is that in reality, the infrastructure built under NREG is often of low quality and unlikely to raise local productivity by much.

Real daily wages were boosted by 5.3%.

We are also able to show that it takes six to 11 months for an NREG intensity shock to pass through to higher wages. This supports anecdotal evidence suggesting that it takes time for workers and employers to build trust in the safety net provided by continued presence of a certain level of employment provision, and also for workers to be able to use this alternative source of employment as a bargaining chip in bidding up their private-sector wages.

NREG also has an important gender perspective. Men and women have equal rights to employment under the programme, and they must be paid the same daily wage rates. Since women are paid significantly less than men in the private market for agricultural labour, one might expect that the upward pressure on wages caused by the equal wages provided under the programme would be greater for women. However, we find that effects on men's and women's wages are about the same. That is, NREG does not contribute to closing the gender wage gap, but nor does it exacerbate it.

Our focus is on wages for unskilled labourers, but we also look at the impact on wage rates for *skilled* labour. We would expect the impact to be less since NREG specifically provides employment for unskilled labourers. To the extent that these are separate labour markets, NREG could have an effect on wage rates in one but not the other. The results confirm that NREG does not have an effect on skilled labour wages. The exception is a small positive effect on blacksmiths' wages, which seems to make sense given that this type of labour is probably more directly linked to standard agricultural field labour than the other two skilled labour categories for which we have data (carpenters and cobblers).

NREG was implemented across India's rural district in three phases. The poorest districts received the programme first, and the better-off districts, last. We find that the effect of the programme on wages is strongest in phase I and II districts, and not significant in phase III districts. This may be because wage rates in phase I and II districts were generally lower than phase III districts before the programme was introduced. Since the statutory NREG wage rates are equal across all districts in a state, they are likely to exert the most upward pressure in the poorest districts.

NREG is an enormous programme by any standards: in the financial year 2010–11, it generated 2.57 *billion* person-days of employment. It is therefore of considerable interest in its own right. The size implies that the scheme, notwithstanding the many problems in its implementation, is eminently scalable. Evaluations of small pilot schemes are often criticised on the basis that the observed effects may not be reproduced at larger scales. That critique does not apply here, and the lessons learned should be of broad interest.

We find that the effect of the programme on wages is strongest in the poorest districts.

Public works programmes provide governments with an additional mechanism with which to influence wage rates in the rural unskilled labour market. Since the link between agricultural wages and poverty rates are well established, if public works can influence agricultural wages then they constitute an attractive policy instrument to reduce poverty.

This article is based on a working paper by Erlend Berg, Sambit Bhattacharyya, Rajasekhar Durgam and Manjula Ramachandra, 2012. *'Can rural public works affect agricultural wages? Evidence from India.'* University of Oxford, CSAE Working Paper Series WPS/2012-05.

Public works programmes can provide governments with an additional mechanism with which to influence wages rates in the rural unskilled labour market.

REFORMING TEACHER TRAINING TO FIND THE BEST TEACHERS

Everyone remembers a brilliant teacher; and most people remember an ineffective one too. Teachers are more important to pupil achievement than small class sizes, resources, or any large-scale policy intervention. **Rebecca Allen** and **Simon Burgess** argue that current teacher selection and training policies do not make sense given what is known about teacher effectiveness.

So, what do we know about teacher effectiveness?

- **Teacher effectiveness matters enormously.** A pupil being taught for eight GCSEs by all effective teachers (those at the 75th percentile of the teacher effectiveness distribution) will achieve an overall GCSE score four grades higher than the same pupil being taught for eight GCSEs by all ineffective teachers (at the 25th percentile). A range of studies have consistently shown a very high impact of teacher effectiveness on pupil progress.

- **Measures of teacher effectiveness are 'noisy'.** Numerous factors affect exam scores, from good or bad luck on exam day, the pupil's ability, motivation and background through to a school's resources. Research shows that it is possible to measure a teacher's contribution to this, but it is an estimate with less-than-perfect precision. For example, a teacher's score in any one year may be affected by being assigned a particularly difficult class, or a particularly well-motivated class in a way not accounted for in the analysis.

Selection into teacher training should be very broad, with relatively low academic requirement.

- **Experience doesn't help beyond three years.** Research shows that on average teachers do become more effective in their first two or three years. Thereafter, there is no evidence of systematic gains as their experience increases: a teacher is as effective after three years as s/he will be after 13 years and 30 years.

- **Good teachers are hard to spot ex ante.** One of the more surprising findings has been that the characteristics that one might have thought would be associated with better teachers simply aren't. Experience, a masters degree, and a good academic record in general are not correlated with greater effectiveness in the classroom. These results have been found in both the US and England.

  We need to be careful what we are claiming here. The research shows that easily observable, objective characteristics variables typically available to researchers are no use in predicting teacher effectiveness. This is not to say that no-one can identify an effective teacher. No doubt many headteachers are adept at spotting teaching talent. But there are enough who aren't to mean that there are ineffective teachers working in classrooms.

- **Finally, very few teachers are dismissed from the profession in England.**

The key judgement for passing probation should be a minimum threshold of progress made by the pupils.

What does this imply for the best way to structure the entry into the profession?

Initial teaching training (ITT) encompasses both the initial training and the probationary year. We assume that the point of ITT is to produce effective teachers who will have the greatest possible impact on pupil progress. ITT therefore plays two roles for the profession – training and selection. The emphasis has typically been on the former, but both are important and neither should be neglected.

We need to consider the process of training teachers as a whole, thinking of the PGCE (Postgraduate Certificate of Education) and the probationary period as both being critical parts of ITT. Based on the facts set out above, we suggest that the following changes to the current ITT could raise the average effectiveness of teachers in England.

- **Broader selection:** The difficulty in identifying people likely to be good teachers suggests a high degree of agnosticism would be appropriate when faced with applicants: it is very hard to tell who will be a good teacher. This is certainly true for selection based on objective criteria from the applicants' academic records. We know that these are unrelated to teaching ability, and so should be irrelevant in selection into ITT.

  We argue that selection into ITT should be very broad, with a relatively low academic entry requirement. This of course is not the situation now, nor the direction of travel of current policy. We argue that the tightening of academic entry requirements into teaching is not helpful: it will restrict the quantity of recruits and have no impact at all on average teaching effectiveness.

- **Graduation from ITT should be tough.** Given that much of an ITT course is now school-based, time spent in the classroom will form an important part of the assessment. Arguably the classroom experience is the key part of the course. However, in such a short space of time it will not generate sufficient data for a robust and objective view of the trainee's effectiveness. It will nevertheless allow the trainee to discover whether teaching is for them.

- **Longer probation:** Once in a job in a school, the key decision on final certification should be made after a longer probation period, for example three years, rather than the current one. The appropriate length of the probation would need to be analysed properly and, depending on the statistical reliability of any pre-hire indicators, school-based performance data, and the cost of being wrong (Staiger and Rockoff (2010) discuss this issue in depth). However, there should be an expectation that not all probationers will make it through to final certification, and indeed only the most effective should be retained.

  The key judgement for passing probation should be a minimum threshold of progress made by the pupils. Obviously, the measurement of that progress and the parameters of the threshold require a great deal of careful work. Like any statistical data, estimates of teacher effectiveness will never be perfect, and a good deal of evidence over a number of years will be necessary to reach a decision, but this is clearly necessary to raise the average effectiveness of the teaching profession in England.

The best way to reduce the problem of low-performing teachers is to make it very difficult for ineffective teachers to get into the profession in the first place.

Our view is that the evidence shows that the selection aspect of ITT is completely the wrong way round. Selection is tight to get into ITT, but once in, progression to full certification is normal and expected. We argue it should be the other way around: we need to be more open to likely teaching ability in the first place and allow a much broader group of people to try teaching and we need a much tougher probation regime before trainees are given final certification. It seems to be very hard to fire ineffective teachers. While the regulations on this have recently changed it may be that the best way to reduce the problem of low-performing teachers is to make it very difficult for ineffective teachers to get into the profession in the first place.

This article is based on a CMPO working paper 09/212 *'Do teachers matter? Measuring the variation in teacher effectiveness in England'*. http://www.bristol.ac.uk/cmpo/publications/papers/2009/abstract212.html

Rebecca Allen is a CMPO Research Associate and a Reader at the Institute of Education, University of London. Simon Burgess is a Professor of Economics at the University of Bristol and Director of CMPO.

Further reading

Staiger, D. and Rockoff, J. (2010) *'Searching for Effective Teachers with Imperfect Information.'* Journal of Economic Perspectives vol. 24 no. 3, pp. 97-118.

Chetty, R., Friedman, J., and Rockoff, J. (2011) *'The long-term impacts of teachers: teacher value-added and student outcomes in adulthood.'* NBER WP 17699. www.nber.org/papers/w17699

Slater, H., Davies, N. and Burgess, S. (2011) *'Do teachers matter? Measuring the variation in teacher effectiveness in England Forthcoming'*, Oxford Bulletin of Economics and Statistics

CONTRACTS AND SERVICE DELIVERY IN FRAGILE STATES

Clare Leaver presents initial findings from a report being compiled by the University of Oxford to assess different approaches to delivering basic services in developing countries. Using case studies from the World Bank and the Department for International Development (DfiD) the Oxford team will attempt to draw out general lessons to feed into future policy decisions.

Ensuring that services such as education, health, and water and sanitation are effectively delivered to citizens is an issue facing every government. In fragile environments, state actors may lack the motivation and, or, institutional capability to rise to this challenge, thereby creating a potential role for donor-funded contributions from non-state providers. Donors taking on such a role typically express a desire to incorporate contributions from non-state providers while enabling the state to retain and strengthen its stewardship and legitimacy. This twin objective has resulted in a wide range of approaches. Understandably, there is a growing appetite among donors and partner governments to establish which, if any, of these approaches work.

During 2011, researchers from the World Bank and DFID prepared case studies of approaches to service delivery in fragile environments using a common framework developed by a team at Oxford. The Oxford team is now drafting a report that draws out the main empirical regularities evident in these studies, together with implications for future policy.

Each case study follows the same format: describing the environment prior to the intervention; summarising the details of the intervention; and assessing the efficacy of the intervention.

The case studies were chosen to reflect a range of environments and donor approaches:

- Six cases focus on Africa, five on the Middle East and Asia, and one on Central America.
- Eleven cases focus on a donor-funded intervention that changed the organisational arrangement in either the health sector, the education sector, or in multiple sectors.
- One case, a study of Eritrea, focuses on a government funded intervention that introduced innovative practices within the publicly financed and run health system.

Success is assessed in terms of the scale of improvements in coverage and other delivery outcomes, the sustainability of any such improvements, and the extent of any problems in process. The report restricts attention to cases where independent evaluative material is available.

An important initial observation is that the quality of the evaluative material is poor. At the time of writing, no independent study is available for four of the twelve cases. Even where independent studies are available, the ability to judge success is limited by the fact that typically the interventions were not designed with program evaluation in mind (e.g. failure to pilot with control areas, failure to collect baseline data).

The small number of useable case studies, and the scarcity of causal estimates within these studies, inevitably makes it hard to draw firm conclusions.

With this caveat in mind, the main observations run as follows (see opposite page for definitions of environment types):

- **In the State-Direct and NGO-Direct environments, the direct form of contracting was associated with fast and sustainable improvements in coverage and other delivery outcomes.** Despite early concerns, governments proved capable of managing the contracting process.
- **In the Mixed environment, the direct form of contracting had a positive (causal) impact on outcomes but was politically unpopular and ultimately withdrawn.** Moreover, performance incentives for frontline staff were needed to overcome problems of motivation.
- In all three environments, allowing non-state providers freedom to innovate appeared to be beneficial.
- **In both the State-Direct and NGO-Direct environments, the delegated form of contracting was associated with fast and sustainable improvements in coverage and other delivery outcomes.** There also appear to have been advantages, relative to direct contracting, in terms of efficiency, accountability, and securing additional funding. However, tensions arose among line ministries, largely due to a lack of engagement in the choice of projects for funding.
- **The unsupported form of decentralisation** (one case in the State-Direct environment) **was associated with a mixed impact on coverage and other delivery outcomes.** Sustainability was also questionable, both in terms of capacity at local level and political support from the centre.
- **The supported form of decentralisation** (one case in the NGO-Direct environment) **was associated with a fast improvement in coverage and other delivery outcomes.** Questions were raised over sustainability, both of the local government structures and contracting in of support.

The Oxford team, together with a wider steering group, is now considering how these observations (drawn from a small but consistent set of case studies) might be used to formulate common lessons that can inform the design of future policy towards service delivery in fragile environments.

This article is based on a *'Work in Progress'*, presentation at CMPO's conference in April 2012. The full report entitled *'Innovative Approaches to Service Delivery in Fragile Situations: An Interpretative Review of the Evidence'* will be available in 2013 from the OxiGED website www.oxiged.ox.ac.uk/index.php/service-delivery

Clare Leaver is a Lecturer in Economics at the University of Oxford.

Fragile Environment Categories

Broadly speaking, the case studies fall into three categories of environment prior to intervention:

1. **State Direct:** recent conflict, the state plays the predominant role in delivery, gaps in coverage particularly in rural areas, main shortcomings relate to low state capacity (six cases);
2. **NGO-Charity Direct:** recent conflict, significant involvement from non-state providers in delivery, gaps in coverage, main shortcomings relate to low state capacity and free-standing NGO programs (five cases);
3. **Mixed:** history of conflict, mixed arrangement for delivery with significant involvement of the private sector, main shortcomings relate to low motivation of state actors (one case);

Donor-funded intervention

Two types of donor-funded interventions were identified:

A. **Contracting:** the central government contracts non-state providers, either directly or through a managing agent/independent agency, to deliver a package of basic services; to deliver infrastructure projects; and/or provide management support to state providers (eight cases);

B. **Decentralisation:** the central government decentralises responsibility for service delivery to lower tiers of government, which may or may not receive support from non-state providers (three cases).

What has been tried where?

- The contracting intervention was more common than the decentralisation intervention in all three environments. The ratio of contracting to decentralisation was largest in the NGO-Direct and Mixed environments and smallest in the State-Direct environment.
- Six of the eight contracting interventions were single sector (four in health and two in education). All three of the decentralisation interventions were multi-sectoral.
- The direct form of contracting was more common than the delegated form in the NGO-Direct and Mixed environments, and for single sector interventions. In contrast, the delegated form was more common than the direct form in the State-Direct environment, and for multi-sectoral interventions.
- The supported form of decentralisation was more common than the unsupported form in the NGO-Direct environment. The reverse was true in the State-Direct environment.

FLAT REGIONAL PAY LEADS TO UNEVEN PERFORMANCE IN THE CLASSROOM

Teacher wages are negotiated through national wage bargaining in England, and consequently exhibit little regional variation. **Jack Britton** and **Carol Propper** investigate whether this flat wage structure across the country has adverse consequences for school performance.

National wage bargaining is a prominent feature of public service provision in England and frequently results in wages that are flat across regions. This is in stark contrast to the private sector, where wages vary to compensate for differences in the cost of living, quality of location and amenities across regions. The lack of variation in wages in the public sector means public sector employees' relative spending power will differ across the country. If wages are important for their performance, this may affect the quality of service provision.

The subject has received considerable attention from the coalition government, with Chancellor George Osborne writing to the pay review bodies in December 2011 arguing that pay for teachers, nurses, police, prison officers, doctors, dentists and member of the armed forces should be more reflective of local labour markets. Osborne claims "there is substantial evidence that the differential between public and private sector wages varies considerably between local labour markets. This has the potential to... lead to unfair variations in public sector service quality."

While it is true that the public-private sector wage differential does vary considerably – for example the difference between teacher wages in the North and London is around nine percent while the equivalent difference in the private sector is around 30 percent – evidence that this contributes towards differential quality in service provision in limited.

10% increase in local labour market wages is associated with the loss of one GCSE point per pupil.

One study to date has shown that centralised pay regulation of nurses resulted in patients being more likely to die from heart attacks after admission to hospital in areas where nurse wages were relatively worse (Propper and Van Reenen 2010). Our work aims to contribute to the debate on national wage bargaining for public sector employees by investigating whether there is an adverse effect on pupil performance from pay regulation for teachers.

We examine the relationship between what a teacher would be paid in the private sector (the outside local labour market wage) and the performance of schools. We focus on school performance in the key GCSE exams taken at age 16. We allow for variable quality of intake by looking at the 'value added' to a pupil by the school, which compares how pupils have done relative to other pupils who achieved similar Key Stage 2 scores across the country. We also allow for the very different composition of students in school across the country.

Overall we find a significant negative relation between local labour market wages and school performance, meaning schools perform relatively worse in areas where local labour market wages are higher. We find that a 10 per cent increase in local labour market wages is associated with the loss of one GCSE point per pupil – this is equivalent to every pupil losing one GCSE grade in one subject. This effect is roughly equivalent to the effect of increasing school class sizes from 30 to 34.

The relationship between local labour market wages and school performance could work through increased vacancies, through lower quality hires, or through reduced teacher effort (we reject the hypotheses that it is working either through the pupil or through the parent).

Our calculations suggest that around one third of the effect might operate through difficulties with recruiting, while the remainder works through a lack of quality teachers and through reduced effort. A reason why this might matter is that a large proportion of work teachers do is discretionary; time spent lesson planning, engaging with pupils or organising extra-curricular activities may be lower under the disaffected teacher.

The implications are that wages are too low in areas of the country where living costs are high. One persistent objection to this policy conclusion is that raising wages for teachers in these areas would redirect resources away from already impoverished areas. However, a high outside wage area is not necessarily one in which there are more wealthy students in state schools, since the use of private schools is also higher in these areas.

Furthermore, we find 'non-linearity' in the relationship between outside wages and school performance. The largest losses in pupil performance occur in areas which have had high private sector wages over a long period and these losses are not offset by gains in low outside wage areas.

Overall performance of schools could be improved with a simple restructuring of the current wage budget.

The implications of this are that if teacher wages were lowered in the low outside wage areas and increased in high outside wage areas, performance in schools in low outside wage areas would not change that much, whilst performance in schools in high outside wage areas would increase.

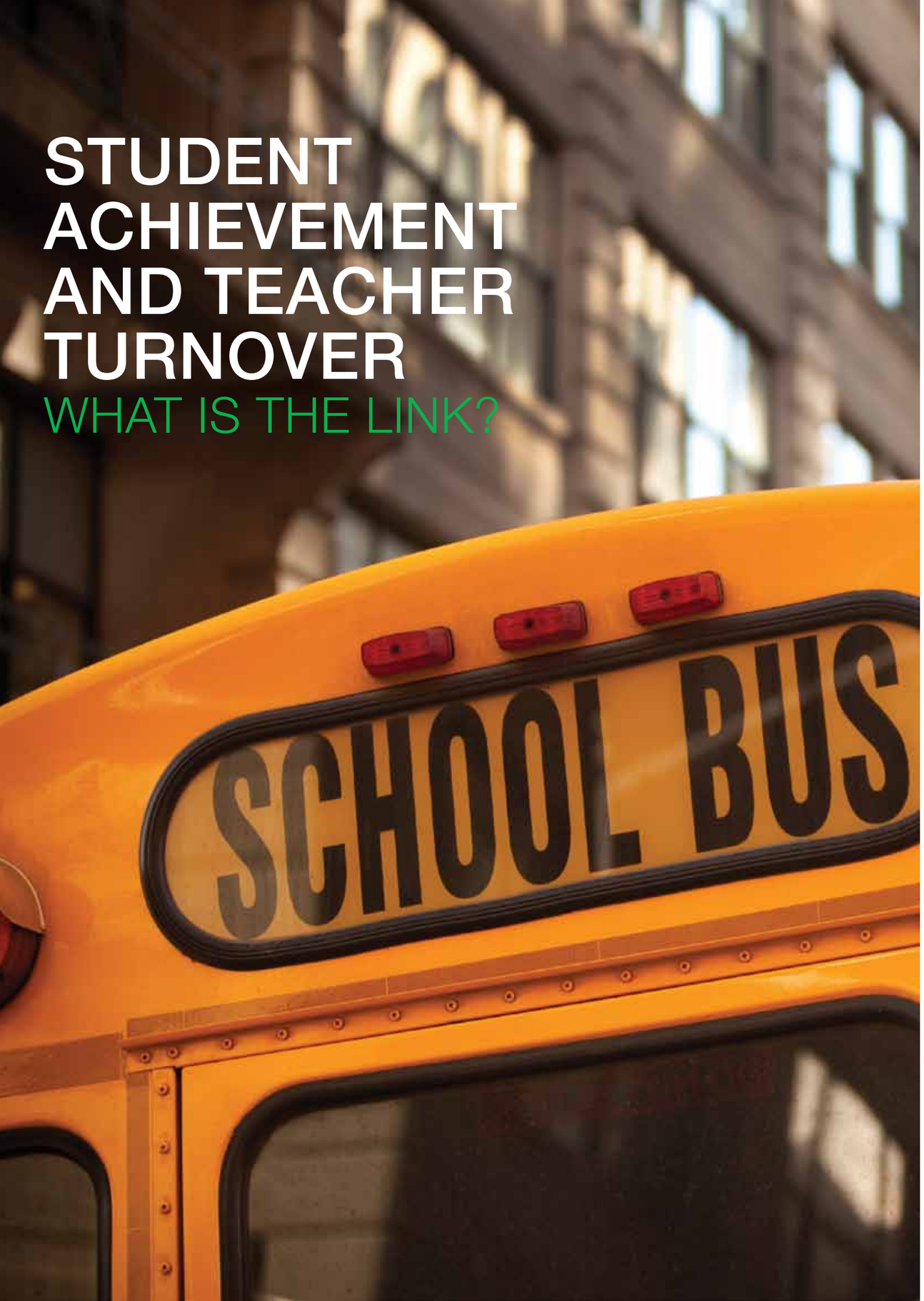
Simple back of the envelop calculations suggest that overall performance of schools could be improved with a simple restructuring of the current wage budget: our estimates suggest that an additional 11,000 pupils could attain five A*-C GCSE grades across the country with a cost neutral strategy. An overall increase in the teacher wage budget could yield higher still returns and could also be used to address the equity issues which would arise from reducing wages in the low outside wage areas in these already tough times.

This article is based on the CMPO working paper 12/239 *'Does Wage Regulation Harm Kids? Evidence from English Schools'* by Carol Proper and Jack Britton (http://www.bristol.ac.uk/cmpo/publications/papers/2012/abstract293.html).

Carol Propper is Professor of Economics of Public Policy and Jack Britton a post-graduate research student at the University of Bristol.

References

Propper, C. and van Reenen, J. (2010) *'Can Pay Regulation Kill? Panel Data Evidence on the Effect of Labour Markets on Hospital Performance'*, Journal of Political Economy, 2010, vol. 118, no. 2



STUDENT ACHIEVEMENT AND TEACHER TURNOVER

WHAT IS THE LINK?

Teacher turnover is often cited as problem for schools yet there is little evidence documenting the actual effect on student achievement. **Susanna Loeb**, **Matthew Ronfeldt** and **James Wyckoff** analyse data from New York City to investigate the impact of teacher turnover on student achievement.

Researchers and policy makers often assume that teacher turnover is a problem for schools, particularly for schools with a high population of disadvantaged students. Theoretically, turnover could be either problematic or beneficial for students. Turnover changes the composition of the teachers at the school and depending on whether the new teachers are higher or lower quality than the teachers who left, the overall 'compositional' effect of turnover on student learning could be either positive or negative.

Turnover may also have broader implications for schools. For example, it could have a 'disruptive' effect on staff cohesion and community, which would affect all teachers, including those retained from year-to-year, and their students. This disruptive effect could be positive if the new teachers brought new and productive ideas or it could be negative if important institutional knowledge was lost.

Given the theoretical uncertainty surrounding turnover, we empirically investigate the effects of turnover on student achievement and begin to explore possible mechanisms by which turnover might affect students.

This study draws on extensive administrative data from New York City. Our analyses focus on approximately 850,000 observations of 4th and 5th grade students over eight academic years. The data include student test scores in math and reading, as well as class, school, and teacher characteristics.

First, we investigate the effects of differences in turnover rates between years in the same school-grade. Then we examine the effects of differences in turnover rates between grades in the same school-year. These approaches minimises the possibility that our results are driven by influences other than turnover. For instance, by comparing school-grades within the same year, we eliminate the possibility that our results are actually picking up the effect of a new principal who affected both turnover and achievement, because this principal would affect all grades at once.

Turnover has a negative effect on student achievement beyond what can be explained by experience and teacher migration.

As expected, we find that there is a significant amount of turnover in New York City. Each year, school-grades experience an average of 11-13 percent turnover and around one percent of 4th and 5th grade level teams experience 100 percent turnover each year.

This turnover, on average, appears to be bad for students. The results consistently indicate that student achievement is lower when turnover is higher, particularly in math. For example, the results imply that in a given grade level with five teachers, reducing teacher attrition from two teachers to none leaving would increase student math achievement by two to four percent of a standard deviation. This effect is similar in magnitude to the difference in learning between low-income students (those eligible for free or reduced price lunch) and other students. Furthermore, turnover appears to be particularly problematic at schools with large populations of disadvantaged students.

We are unable to examine exactly how turnover affects students so we ran additional analyses to get a sense of the whether the effects are compositional or disruptive in nature. One way new teachers may differ from the teachers they are replacing is in terms of experience. When we control for teacher experience in our models, the effects of turnover shrink but remain substantial, suggesting that changes in the distribution of teacher experience (i.e. novice teachers replacing tenured teachers) explain some, but not all, of the effects of turnover on student achievement.

It is also possible that teachers who transfer across schools, regardless of experience, are generally less effective in their first year at the new school. Again, while this does appear to play some role, turnover has a negative effect on student achievement beyond what can be explained by experience and teacher migration.

An alternative way to capture the compositional changes brought about by turnover is control directly for teacher effectiveness. While we do not have a perfect measure of effectiveness we can measure the test-score gains of students in the teacher's

When teachers leave schools at high rates student learning drops.

classrooms in prior years adjusted for the students' background characteristics.

Similar to the teacher experience results, we find that differences in teachers' >

effectiveness as a result of turnover can explain some, but not all, of the effects of turnover on student achievement. This partial explanation is true for both the entire population of schools and lower-achieving schools in particular.

While it appears that under-served schools do to tend to fill vacancies with relatively less effective teachers, this is not the entire story. To examine this further, we separately investigate the effects of turnover on the students of teachers who remain in the same school-grade from year-to-year and are not actually a part of the turnover. We consistently find that students of 'stayers' perform worse when turnover is greater, particularly in lower-performing schools. This result bolsters our previous findings and points to teacher turnover having a wider disruptive effect on the school which is harmful to student achievement.

Although teacher turnover may be beneficial in some cases, we find that, on average, turnover is detrimental to student learning. When teachers leave schools at high rates student learning drops. Furthermore, the effects of turnover extend beyond the compositional changes to teachers and include broader-reaching disruptive effects.

Given the harmful nature of turnover, polices designed to increase teacher retention are a promising avenue to improve student achievement. For instance, schools could implement incentive structures aimed specifically at highly effective teachers who might otherwise leave. Such policies may be especially important in schools with large populations of disadvantage students, where turnover is particularly problematic. A second potential approach to mitigating the effects of turnover is to plan for turnover, assuring that the institutional knowledge that traditionally has left with the departing teacher remains in the school and is more easily passed on to new workers.

Polices designed to increase teacher retention are a promising avenue to improve student achievement.

This article is based on *'How Teacher Turnover Harms Student Achievement'*, a working paper by Ronfeldt, M., Loeb, S., Wyckoff, J. (2012), www.cepa.standford.edu.

Susanna Loeb is the Barnett Family Professor of Education at Stanford University. Matthew Ronfeldt is Assistant Professor at the University of Michigan. James Wyckoff is Director of the Center for Education Policy and Workforce Competitiveness at the Curry School of Education.

HOW LONG ARE TEACHERS STAYING?

The popular view in the UK is that teachers at schools in poor communities serve their time and quickly move on. But is this actually true or an urban myth? Using new School Workforce Census data collected from all schools in England in Autumn 2010 **Rebecca Allen** and **Simon Burgess** have started to explore this question systematically, moving beyond a collection of local anecdotes to find out how long teachers actually stay in specific schools.

Overall, it seems that teaching is not a low turnover profession, and like many urban myths, there is a hint of truth in this view that deprived schools experience greater teacher turnover, but not much.

In a typical school, about a fifth of teachers have been there for less than two years, and over half of the teachers have been in that school for less than five years. On the other hand, nearly a fifth have been there at least 10 years, and in fact over five percent have stayed over 20 years. Of course, teachers vary and we compare different groups (see Table 1, page 22). There is very little difference in tenure between female and male teachers, nor between primary and secondary schools.

Averaging over all teachers, the mean time in a school is 6.7 years. Here we need to introduce a technical issue. The data come from teachers in schools, so this is job tenure so far, elapsed tenure. Obviously, a teacher who has just arrived at a school may go on to spend the rest of her career there. Under certain circumstances, the statistical model implies that the average completed tenure is double the average elapsed tenure.

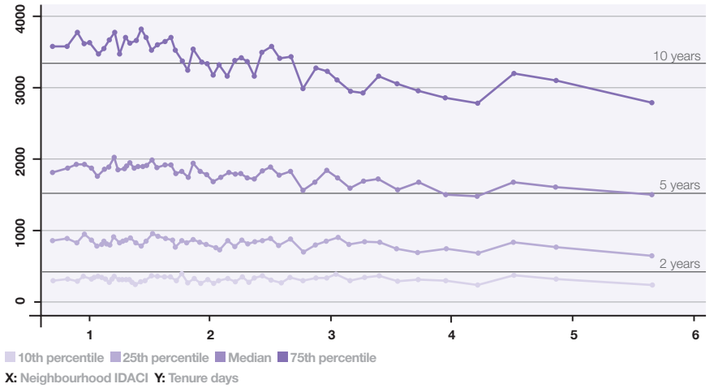
That is the overall picture, what of the differences between disadvantaged and affluent neighbourhoods? We find systematic and statistically significant differences in turnover: schools with many poor pupils do have more short-tenure teachers and fewer experienced teachers.

On average, however, the differences are small: 18 percent of teachers in the least disadvantaged schools have tenure of 0-2

The differences are small: 18% of teachers in the least disadvantaged schools have tenure of 0-2 years, compared to 22% in the most disadvantaged.

years, compared to 22 percent in the most disadvantaged. At the other end of the scale, 20 percent of teachers in schools in the >

Figure 1
Distribution of teacher tenure by school deprivation



Poorer schools do indeed experience slightly higher turnover and have a higher proportion of young and inexperienced teachers, to find out why this happens is a task for future work.

most affluent neighbourhoods have tenure of at least 10 years, whereas the figure in the most deprived neighbourhoods is 17 percent.

Figure 1 gives a flavour of the results. It shows the 10th percentile of tenure in school in days (the lowest line in the figure), across the full range of communities in England, from the richest two percent in the far left-hand side point to the poorest two percent in the far right-hand point.

The 10th percentile comes out at somewhat less than two years, but more interestingly, is flat. The number barely changes across the entire distribution. There is a very slight slope in the 25th percentile and in the median values, reinforcing the point that there are systematic differences but that they are quantitatively small. There is a more noticeable difference in the 75th percentile: in schools serving poor communities, there are slightly fewer experienced teachers.

We also use the richness of the data to breakdown the relationship between turnover and poverty. We show that part can be accounted for by pupil characteristics, perhaps because students in schools in more deprived areas are harder to teach. Part also is accounted for by differences in the local teacher labour market around each school. Controlling for school, student and teacher labour market factors reduces the association between school poverty and turnover, but does not eliminate it.

The remaining association between teacher turnover and poverty is largely accounted for by teacher characteristics, with the poorer schools hiring much younger teachers on average. How should we understand this?

We interpret this as either deriving from the preferences of young teachers, or as reflecting the low market attractiveness of disadvantaged schools. There are a number of possibilities.

First, it could be that this is a desired career path for young teachers. New teachers may look for their first jobs near to where they trained, which implies predominantly urban and therefore on average deprived, schools. Alternatively it could be a desired career path deriving from younger teachers possibly having more idealistic preferences, and welcoming the opportunity to work in deprived schools.

New teachers may look for their first jobs near to where they are trained.

Under these interpretations, the allocation reflects the desire of younger teachers to work in deprived schools, and the higher turnover in such schools derives from this.

The alternative interpretation is a matching story in which the more effective teachers sort on average into the more affluent schools, and the disproportionate number of inexperienced teachers in the poorer urban schools reflects the realities of the market that these schools face.

So, though we know that poorer schools do indeed experience slightly higher turnover and have a higher proportion of young and inexperienced teachers, distinguishing why this happens is a task for future work. It requires further sweeps of the data and possibly attitudinal data from teachers as well.

It is now widely acknowledged that teacher effectiveness is the single most important factor in raising attainment. Attainment gaps arise in part from students' exposure to teachers of differing effectiveness. The process by which different teachers end up at different schools in front of different children is little understood. Indeed teacher labour markets as a whole are not well understood and we intend to spend the next few years utilising this new data to address this research programme.

Rebecca Allen is a CMPO Research Associate and a Reader at the Institute of Education, University of London. Simon Burgess is a Professor of Economics at the University of Bristol and Director of CMPO.

Table 1
Years in current school (%)

	All teachers	Male	Female	Primary	Secondary
0-2 years	**19.4**	**20.4**	**19.2**	**19.5**	**19.4**
2-5 years	**36.8**	**37.0**	**36.7**	**36.6**	**37.0**
5-10 years	**24.8**	**23.5**	**25.2**	**24.8**	**24.8**
10 years or more	**18.9**	**19.1**	**18.9**	**19.1**	**18.8**
Number	343,547	80,704	262,843	172,137	171,410

Note: classroom teachers only, excluding assistant, deputy and headteachers

CMPO CONFERENCE SERIES

MAXIMISING THE IMPACT OF POLICY EXPERIMENTS

There is something of a revolution going on in evidence-based policy-making in the UK. A small team in the Cabinet Office – the Behavioural Insights Team – has been conducting field experiments to test ways of improving policy delivery in the UK. A one-day workshop, hosted by the CMPO, brought together academics and policy-makers to discuss the role of such experiments and how to ensure that they are designed and implemented to maximise their potential impact for policy-making.

Randomisation allows the experimenter to learn about the true causal effect of an intervention.

Through a series of randomised controlled trials the Behavioural Insights Team have demonstrated the effectiveness of text messages to improve payment of fines and used social norms to increase the promptness of tax payments. Outside of government, a number of UK academics are also turning to experiments to learn more about whether policies work and why.

The experiments discussed at the CMPO workshop ranged from a five-year trial run in Ireland to evaluate the effect of an early-years intervention programme (The Preparing for Life intervention), to trials in schools testing healthy eating incentives and performance payments, to experiments run in the laboratory to look at how people respond to increasing levels of complexity in a stylised tax system. Across this wide range, however, there were a number of common themes.

Above all, presenters emphasised the importance of randomisation. The power of experiments comes through the random assignment of the 'treatment'. Randomisation allows the experimenter to learn about the true causal effect of an intervention because the treatment is the only thing that differs between the treatment and control groups.

Yet, there have been some well-known field experiments – such as the Perry Pre-School Programme, trialled in the US in the early 1960s – where there is now a concern that the randomisation may not have been >

CMPO CONFERENCE SERIES

perfect. A big challenge is that often those involved in service delivery may want to give the 'best' treatment to those that they think are in the greatest need. In the experiments discussed, therefore, a lot of care had been paid to achieving a truly random allocation of participants.

Often there may be a prior ethical concern with the idea of random allocation. Yet, in practice, the opposition was found to be stronger among the gatekeepers than among the participants themselves. One message that was effective with the participants was that their involvement would help the researchers to learn about what worked and so benefit others.

There is always a need to obtain informed consent from participants and this may lead to a potential selection bias into the experiment when that consent is not forthcoming from everyone.

Experiments have the potential to have huge impact with policy-makers because they can provide a clear demonstration of policy effect.

In an experiment run in schools to test the effects of incentivising children to eat fruit and vegetables, for example, not all children in the selected schools could take part (in the treatment or control groups) because consent was not given by their parents. The result is that the group of people taking part in the experiment may not be a randomly selected sample of the population – even of the relevant sub-population – and this may have implications for the potential wider applicability of the findings.

There are obvious parallels between policy experimentation and what happens in the medical profession where new medicines are subject to extensive testing. Prior to clinical trials, there is also a stage of development and testing in the lab, often involving rats, which drew comparisons with the role of the lab in policy development (with the role of the rats played by students). The strength of the lab is that it offers a low-cost option to researchers and policy-makers to enable them to 'try out crazy stuff' and to test and tweak designs.

The two approaches – lab and field – should be seen as complements in the policy development process. Researchers can go to the lab before going to the field to try out ideas; but in the policy process, lab experiments may also provide an opportunity to gain greater insights into the findings from field experiments by, for example, bringing the relevant players into the lab to find out why things might have worked and to test responses.

Experiments have the potential to have huge impact with policy-makers because they can provide a clear demonstration of policy effect. Comparing outcomes among two groups is much simpler than explaining a structural model or an instrumental variable identification strategy.

Maximising impact may also require continual engagement with stakeholders – particularly in the context of a longer-term study where the findings may not be known for several years. The political cycle may mean that politicians come and go before the experiment has run its course and this may require researchers to engage with the civil servants who are more likely to remain in post.

Of course, politicians may not be interested in funding something that will not deliver results until they are out of office, but non-government funders are an alternative investor in longer-term studies. Co-funding for the early years interventions study in Ireland came from Atlantic Philanthropies. One of their rationales was legacy-building and the idea that evidence from an experiment, if it proved the programme's success, would make the programme harder for policy-makers to over-turn.

It seems almost inevitable that experiments will play a bigger role in the policy process in the future. For the Behavioural Insights Team, success is breeding success. While there was initial scepticism inside and outside Whitehall about running trials, their clear demonstrability has led to demands for doing something similar from other departments.

Increasing localisation of service provision is also providing opportunities for local governments to run trials. And non-government bodies, charitable foundations such as Atlantic Philanthropies and other organisations such as the Education Endowment Fund, which is funding the performance payments trial, are also increasingly keen on this approach.

Academics emphasised the need to pay careful consideration to what can be learned from individual policy experiments.

Going forward, the academics emphasised the need to pay careful consideration to what can be learned from individual policy experiments. While field experiments provide strong evidence of treatment effects, rolling out a programme nationally is not the same as running a field experiment – even the treatment may vary from what it was in a small scale tightly controlled experiment as well as obvious issues with differences in the affected population. This also means thinking seriously about the potential selection issues that are likely to affect the sample of people who take part in a trial. Replication of lab experiments is relatively straightforward but may be much harder to achieve in the field.

At the very least the findings from different trials should be brought together in one place, e.g. in a trials database, and be made available to allow people to gain the maximum knowledge from each trial both individually and collectively.

For further information please contact Sarah Smith at CMPO.

THE ECONOMICS OF PUBLIC SERVICE REFORM

23-24 MAY 2013

CMPO is organising an international economics conference on public service reform. The event will feature invited talks, contributed sessions and policy panels at which policy makers, practitioners and academics will debate current issues, including the regulation of healthcare markets.

Confirmed keynote speakers are:
– **Oriana Bandiera** (London School of Economics)
– **Marty Gaynor** (Carnegie Mellon)
– **Grant Miller** (Stanford)
– **Karthik Muralidharan** (University of California, San Diego)
– **Fiona Scott Morton** (Yale)

We welcome contributions on the following themes, including applications in both developed and developing country settings, within the topic of public sector reform:
– Incentives and motivations
– Public services markets in health and education
– Not-for-profit and private providers
– The use of field experiments

There will be no charge to attend the conference. There will also be some limited funds available to help with travel costs.

Please email paper submissions (ideally a full paper or a lengthy abstract) to **cmpo-admin@bristol.ac.uk** by 28th February 2013. Selected participants will be contacted by 21st March.

To register for this event please email:
cmpo-admin@bristol.ac.uk

CMPO

About this publication

Research in Public Policy is the publication of the Centre for Market and Public Organisation.

The Centre for Market and Public Organisation (CMPO) is a leading research centre, combining expertise in economics, geography and law. Our objective is to study the intersection between the public and private sectors of the economy, and in particular to understand the right way to organise and deliver public services.

The Centre aims to develop research, contribute to the public debate and inform policy-making. CMPO started its second five years as an ESRC Research Centre in October 2009. The Centre was established in 1998 with two large grants from The Leverhulme Trust, and in 2004 was awarded ESRC Research Centre status.

Keep updated

Visit the CMPO blog 'Viewpoint':
bristol.ac.uk/cmpo/blog

Follow us on Twitter:
@cmpobristol

Watch us on YouTube:
youtube.com/user/CMPObristol

Listen to our podcasts:
bristol.ac.uk/cmpo/audio

bristol.ac.uk/cmpo

CMPO University of BRISTOL E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

CMPO working papers

Recent additions to the CMPO Working Paper Series, all of which, along with the rest of the series, can be downloaded at: bristol.ac.uk/cmpo/publications/papers

12/297 Free to Choose? Reform and Demand Response in the English National Health Service Martin Gaynor, Carol Propper and Stephan Seiler

12/296 Competition, Equity and Quality in Health Care Maija Halonen-Akatwijuka and Carol Propper

12/295 Earnings Mobility and Inequality: An Integrated Framework Paul Gregg, Rosanna Scutella and Claudia Vittori

12/294 The teacher labour market, teacher turnover and disadvantaged schools: new evidence for England Rebecca Allen, Simon Burgess and Jennifer Mayo

12/293 Does Wage Regulation Harm Kids? Evidence from English Schools Carol Propper and Jack Britton

12/292 One Kind of Democracy Siwan Anderson, Patrick Francois and Ashok Kotwal

12/291 Increasing charitable giving – what can we learn from economics? Sarah Smith

12/290 Peer effects in charitable giving: Evidence from the (running) field Sarah Smith, Frank Windmeijer and Edmund Wright

12/289 Diversity and Donations: The effect of religious and Ethnic Diversity on Charitable Giving James Andreoni, Abigail Payne, Justin Smith and David Karp

12/288 The Impact of Fathers' Job Loss during the 1980s Recession on their Child's Educational Attainment and Labour Market Outcomes Paul Gregg, Lindsey Macmillan and Bilal Nasim

12/287 How should we treat under-performing schools? A regression discontinuity analysis of school inspections in England Rebecca Allen and Simon Burgess

12/286 Ethical and Economic Perspectives on Global Health Interventions Sonia Bhalotra and Thomas Pogge

12/285 Separation of Powers and the Size of Government in the U.S. States Leandro de Magalhães and Lucas Ferrero

12/284 Incumbency Effects in Brazilian Mayoral Elections: A Regression Discontinuity Design Leandro de Magalhães

12/283 The Effect of a Hospital Nurse Staffing Mandate on Patient Health Outcomes: Evidence from California's Minimum Staffing Regulation Andrew Cook, Martin Gaynor, Melvin Stephens Jr. and Lowell Taylor