

**CMRE Research Digest
2016-07 (5)**

Editor: Gabriel Heller-Sahlgren

Table of Contents

Editor's Pick.....	1
Effects of Policy and Practice - Developed World.....	3
Effects of Policy and Practice - Developing World.....	5
General Education	7

Editor's Pick

The Causes and Consequences of Test Score Manipulation: Evidence from the New York Regents Examinations

By: Thomas S. Dee, Will Dobbie, Brian A. Jacob, and Jonah Rockoff

NBER Working Paper No. 22165 (April 2016)

[Published version](#)

[Free version](#)

In most countries, externally marked tests have come to play an increasingly important role in education. In the UK nations, examinations determining pupils' grades have for long been centrally marked. Until recently, however, many GCSE and A-level courses have included some teacher-assessed elements – and many have lamented the government's determination to eliminate these components.

Educationalists often deplore standardised testing because it reduces teacher autonomy in the classroom and because it supposedly offers an unreliable snapshot of pupil performance. Yet any form of teacher assessment of importance gives rise to the possibility of score manipulation. Indeed, the lack of reliability in such assessment is the key argument against it.

In this paper, Thomas Dee and colleagues study the causes and consequences of allowing teachers to mark the Regents Examinations in New York State. These exams are high-stakes tests measuring performance in accordance with the state's secondary-school curricula. Up until and including 2010, the tests were marked locally by teachers in the pupils' own school. Then, in 2011, schools were not allowed to re-grade exams with scores just below the determined proficiency cut-off – a practice that previously was required – and in 2012, a reform centralised grading, which abolished any ability to manipulate test scores whatsoever.

The authors find clear evidence that teachers inflated scores in the period when it was possible to do so. There is a clear jump in the distribution of scores that fall just above the relevant thresholds for specific grades, compared with those that fall just below these thresholds – which would not be expected without manipulation.

Indeed, the authors find that about 40 per cent of scores that were close to the thresholds were inflated. This manipulation was reduced by 80 per cent in 2011 and then eliminated entirely in 2012. Both re-scoring policies and local grading were therefore important factors behind grade inflation.

Importantly, pre-reform manipulation affected pupils in different schools differently. Also, minority pupils with low initial performance and worse behaviour benefited from manipulation overall, simply because their scores were more likely to fall around the thresholds. However, conditional on scoring near the thresholds, these pupils were actually less likely to have their scores inflated.

Furthermore, using the reforms in 2011 and 2012, the authors analyse the causal impact of manipulation on future outcomes, finding that pupils who have their scores inflated above the relevant thresholds are much more likely to graduate from high school. This also has egalitarian implications: the black-white test score gap would have been 1.3 percentage point higher, while the share of pupils graduating from high schools would have decreased by 1.2 percentage points, without any manipulation taking place.

At the same time, however, pupils who have their scores inflated are less likely to meet the requirements for obtaining an advanced high school diploma and pass more advanced examinations. This indicates that the impact of grade inflation depends on pupils' prior attainment. Pupils who are close to dropping out benefit from score inflation, as they are not forced to retake a class they may fail. But pupils on the margin of receiving the advanced diploma may be hurt by inflation if this means they are tricked into believing they can get away with studying less than they have to in order to succeed.

Finally, the authors also provide evidence of mechanisms driving the score manipulation. Inflation is just as prevalent in subjects that were not included in the accountability system, indicating that pressure from government is not a key driver. Intriguingly, the authors also find that teacher performance pay has no impact on score manipulation, indicating that pecuniary motives among teachers are unlikely to be the main reason either. Since they do find that inflation occurs both for pupils on the pass/fail margin and for those on the margin of secondary benefits, such as eligibility for university credits, it is likely that altruism ultimately drives manipulation.

Overall, the paper highlights the dangers of teacher assessment. It indicates that any such assessment of importance is likely to be subject to arbitrary inflation, whether or not the scores are tied to accountability systems, and that most forms of teacher assessment are likely to be compromised by the lack of comparability. In times of opposition to the government's reforms to reduce or eliminate teacher-assessed components in GCSEs and A-levels, the paper provides fresh evidence indicating that such opposition may be misplaced.¹

¹ This does not mean that all forms of teacher assessment is necessarily is bad. Rob Coe and I have previously [suggested](#) that such assessment could potentially be reconciled with demands for comparability. This would require the total distribution of marks to be fixed at the exam centre level, based on results in the external components of the qualification. This would ensure a fixed sum of teacher-assessed marks (e.g. 10 As, 20 Bs etc.), which means that teacher assessment would only redistribute marks among pupils

Effects of Policy and Practice – Developed World

The Effects of Delaying Tracking in Secondary School: Evidence from the 1999 Education Reform in Poland

By: Maciej Jakubowski, Harry A. Patrinos, Emilio E. Porta, and Jerzy Wiśniewski

Education Economics (forthcoming)

[Published version](#)

[Working paper version \(free\)](#)

Delaying tracking, extending students' exposure to a general academic education and increasing their time on task on basic competences (reading, mathematics) could improve academic outcomes. To test the hypothesis that delayed vocational streaming improves academic outcomes, this paper analyses Poland's significant improvements in international achievement tests and the restructuring of the system which expanded general schooling. Estimates using propensity-score matching and difference-in-differences estimates show that delaying vocational education and increasing time on task have a positive and significant impact on student performance on the order of a standard deviation.

Effective Teacher Retention Bonuses: Evidence From Tennessee

By: Matthew G. Springer, Walker A. Swain, and Luis A. Rodriguez

Educational Evaluation and Policy Analysis

[Published version](#) | [Forthcoming version \(free\)](#)

The authors report findings from a quasi-experimental evaluation of the recently implemented \$5,000 retention bonus program for effective teachers in Tennessee's Priority Schools. They estimate the impact of the program on teacher retention using a fuzzy regression discontinuity design by exploiting a discontinuity in the probability of treatment conditional on the composite teacher effectiveness rating that assigns bonus eligibility. Point estimates for the main effect of the bonus are not different from zero. However, for teachers of tested subjects and grades, the program has a consistently positive effect that is both statistically and substantively significant. The authors hypothesise that the null finding for the main effect is driven by teachers of untested subjects and grades given the amount of weight Tennessee's teacher evaluation system attributes to school-level performance. This creates a strong incentive to exit the Priority Schools that are by definition low performing. Implementation concerns,

including the timing of application process and observed noncompliance in bonus distribution, present obstacles for both the program's effectiveness and its evaluation.

Crime, Compulsory Schooling Laws and Education

By: Brian Bell, Rui Costa, and Stephen Machin

Economics of Education Review (forthcoming)

[Published version](#) | [Working paper version \(free\)](#)

Do compulsory schooling laws reduce crime? Previous evidence for the US from the 1960s and 1970s suggests they do, primarily working through their effect on educational attainment to generate a causal impact on crime. In this paper, the authors consider whether more recent experience replicates this. There are two key findings. First, there is a strong and consistent negative effect on crime from stricter compulsory schooling laws. Second, there is a weaker and sometimes non-existent link between such laws and educational attainment. As a result, credible causal estimates of the education–crime relationship cannot in general be identified for the more recent period, though they can for some groups with lower education levels (in particular, for blacks).

Accountability Pressure, Academic Standards, and Educational Triage

By: Douglas Lee Lauen and S. Michael Gaddis

Educational Evaluation and Policy Analysis (March 2016)

[Published version](#) | [Forthcoming version \(free\)](#)

Despite common conceptions, evidence on whether No Child Left Behind (NCLB) has had adverse effects for low achieving students is mixed. The authors hypothesise that the incentive to shift attention away from the lowest achieving students increases with the rigour of state standards. Using panel data from students in North Carolina, they exploit two natural experiments: increases in the rigour of standards in math in 2006 and then again in reading in 2008. They report an increase in test score gaps between low and high achievers and students near grade level. Adverse effects on low achievers are largest in the lowest achieving schools. They discuss the policy implications of our findings given the widespread adoption of more rigorous Common Core Standards.

Effects of Policy and Practice – Developing World

Maternal Education, Parental Investment and Non-Cognitive Skills in Rural China

By: Jessica Leight and Elaine M. Liu

NBER Working Paper No. 22233

[Published version](#)

[Free version](#)

The importance of non-cognitive skills in determining long-term human capital and labour market outcomes is widely acknowledged, but relatively little is known about how educational investments by parents may respond to non-cognitive skills early in life. This paper evaluates the parental response to variation in non-cognitive skills among their children in rural Gansu province, China, employing a household fixed effects specification; non-cognitive skills are defined as the inverse of both externalising challenges (behavioural problems and aggression) and internalising challenges (anxiety and withdrawal). The results suggest that on average, parents invest no more in terms of educational expenditure in children who have better non-cognitive skills relative to their siblings. However, there is significant heterogeneity with respect to maternal education; less educated mothers appear to reinforce differences in non-cognitive skills between their children, while more educated mothers compensate for these differences. Most importantly, there is evidence that these compensatory investments lead to catch-up in non-cognitive skills over time for children of more educated mothers.

Decentralization of Health and Education in Developing Countries: A Quality-adjusted Review of the Empirical Literature

By: Anila Channa and Jean-Paul Faguet

World Bank Research Observer (forthoming)

[Published version](#)

[Working paper version \(free\)](#)

The authors review empirical evidence on the ability of decentralisation to enhance preference matching and technical efficiency in the provision of health and education in developing countries. Many influential surveys have found that the empirical evidence of decentralization's effects on service delivery is weak,

incomplete, and often contradictory. The authors' own unweighted reading of the literature concurs. However, when they organise quantitative evidence first by substantive theme, and then—crucially—by empirical quality and the credibility of its identification strategy, clear patterns emerge. Higher-quality evidence indicates that decentralisation increases technical efficiency across a variety of public services, from student test scores to infant mortality rates. Decentralisation also improves preference matching in education, and can do so in health under certain conditions, although there is less evidence for both. They discuss individual studies in some detail. Weighting by quality is especially important when quantitative evidence informs policy-making. Firmer conclusions will require an increased focus on research design, and a deeper examination into the prerequisites and mechanisms of successful reforms.

General Education

Returns to Education: The Causal Effects of Education on Earnings, Health and Smoking

By: James J. Heckman, John E. Humphries, and Gregory Veramendi

NBER Working Paper No. 22291

[Free version](#)

This paper estimates returns to education using a dynamic model of educational choice that synthesises approaches in the structural dynamic discrete choice literature with approaches used in the reduced form treatment effect literature. It is an empirically robust middle ground between the two approaches, which estimates economically interpretable and policy-relevant dynamic treatment effects that account for heterogeneity in cognitive and non-cognitive skills and the continuation values of educational choices. Graduating college is not a wise choice for all. Ability bias is a major component of observed educational differentials. For some, there are substantial causal effects of education at all stages of schooling.

Immigrant Student Performance in Math: Does it Matter Where You Come From?

By: Gianna G. Gianelli and Chiara Rapallini

Economics of Education Review (June 2016)

[Published version](#) | [Working paper version \(free\)](#)

The performance gap in math of immigrant students is investigated using PISA 2012. The gap with respect to non-immigrant schoolmates is first measured. The hypotheses that first (second) generation students coming from (whose parents come from) countries with a higher performance in math fare better than their immigrant peers coming from lower-ranked countries are then tested on a sample of about 13,000 immigrant students. The estimated average immigrant-native score gap in math amounts to -12 points. The results show that immigrant students coming from higher-ranked origin countries have a significantly lower score gap, and are thus relatively less disadvantaged. For example, coming from a country in the top quintile for math and having attended school there for one year improves the absolute score gap by more than 33 points, the highest coefficient among the variables that reduce the gap, such as parental education and socio-economic status.