Collaborative overreach: Why collaboration probably isn’t key to the next phase of school reform

Research report 7

James Croft

Centre for the Study of Market Reform of Education

23 Great Smith Street, Westminster, London, SW1P 3BL • www.cmre.org.uk
Annual research digest 2015
About the authors

Dale Bassett is Head of Product Reform at AQA awarding body. Previously at AQA, he was Director of Public Policy, and prior to that Research Director at the think tank Reform, where he led on research into public service reform, specialising in education policy.

Daisy Christodoulou is the Research and Development Manager at ARK Schools (Absolute Return for Kids). A Teach First graduate, she taught English in two London comprehensives. She is the author of *Seven Myths about Education*.

James Croft is Executive Director of the Centre for the Study of Market Reform of Education (CMRE) and the author and co-author of several of its reports, including ‘Collaborative overreach: why collaboration probably isn’t key to the next phase of school reform’ (2015).

Henrik Jordahl is Associate Professor of Economics at Uppsala University, and Programme Director at the Research Institute of Industrial Economics (IFN) in Stockholm, where directs the research program ‘The Economics of the Service Sector’.

Jack Worth is Research Manager at the National Foundation for Educational Research (NFER). As part of the NFER Education Trials Unit, he managed the evaluation of the University of Oxford *Improving Numeracy and Literacy* programme, a large cluster-randomised trial in primary schools. Previously he was Researcher at the Centre for Market and Public Organisation (CMPO).
Tim Oates, CBE, is Group Director of Assessment Research and Development at Cambridge Assessment. He is a member of Ofqual’s Standards Advisory Group and following the publication of his paper ‘Could do better’, offering a framework for reform, was chair of the Expert Panel for Review of England’s National Curriculum.
Contents

Introduction 1

‘Mobile Phones in the Classroom: Examining the Effects of Texting, Twitter, and Message Content on Student Learning’
– Jeffrey H. Kuznekoff, Stevie Munz, and Scott Titsworth
Commentary by Daisy Christodoulou 2

‘Can personality traits and intelligence compensate for background disadvantage? Predicting status attainment in adulthood’
Commentary by Dale Bassett 6

‘Does Management Matter in Schools?’
– Nicholas Bloom, Renata Lemos, Raffaella Sadun and John Van Reenen
Commentary by Henrik Jordahl 11

‘An RCT evaluation of the Numeracy and Literacy project’
– Jack Worth, Juliet Sizmur, Rob Ager, and Ben Styles
Commentary by Jack Worth 15

‘“Variation problems” and their roles in the topic of fraction division in Chinese mathematics textbook examples’
– Xuhua Sun
Commentary by Tim Oates 19

‘School Vouchers: A Survey of the Economics Literature’
– Dennis Epple, Richard E. Romano, Miguel Urquiola
Commentary by James Croft 21
Introduction

Evidence-based policy has become somewhat of a catchphrase in politics – everybody is for it and nobody is against it. But there is evidence and there is evidence. For long, education policy has unfortunately been guided by research of poor quality, which in turn has contributed to confusion regarding what works and what does not work to generate higher pupil outcomes.

In order to disseminate high-quality academic research to a wider audience, the CMRE has for the past year produced a free, monthly research digest. It is intended to give interested parties a direct view into the policies and practices that rigorous research (often economic) suggests work, and do not work, in education – both from a macro-policy perspective as well as from the point of view of teachers and headteachers who are looking for more effective classroom strategies. The digest includes abstracts of interesting studies, together with links to free versions of those studies. The abstracts are also accompanied by one ‘Editor’s Pick’, which comments on a selected study that the editor thinks deserves a slightly longer treatment and analysis.

In this publication, we have invited a couple of influential researchers to reflect on a piece of research they think should be acknowledged and discussed in education policy circles. We hope you will find it useful and that you will continue to subscribe to our monthly digest in the coming year.
‘Mobile Phones in the Classroom: Examining the Effects of Texting, Twitter, and Message Content on Student Learning’
– Jeffrey H. Kuznekoff, Stevie Munz, and Scott Titsworth

*Communication Education (July 2015)*

Published version (free)

Commentary by: Daisy Christodoulou

Over the last decades, technology has frequently promised to transform education, but with little actual success in raising pupil performance. Governments have invested in, for example, laptops, tablets, and interactive whiteboards, with disappointing results. More recently, ideas that students learn through a mixture of traditional classroom activities and online content/quizzes – sometimes described as the ‘flipped classroom’ and ‘blended learning’ – have increased in popularity. New apps, such as Khan Academy, Duolingo, and Coursera, have come to offer free educational content in a range of subjects. They also use the data they gather from their tests to improve the effectiveness of their content, and to tailor instruction to the individual student.

Many of these approaches depend on students accessing content through mobile devices. However, mobile devices are banned in many schools and classrooms, with parents and educators often being sceptical of their educational value. Could mobile phones be a platform for the delayed “ed tech” revolution? Or are parents’ and educators’ scepticism warranted?
Research on the use of mobile phones in classrooms tends to finds that mobile phones are distracting – but most it involves students engaging with irrelevant content. But this recent study, analysing an experiment carried out among US undergraduate students, was instead designed so that some students would use their mobiles to engage with content of relevance for the lesson. It also attempted to tease out the separate effects of (1) responding to queries posed by someone else, (2) creating messages, and (3) the frequency of messaging.

The researchers divided students into four experimental groups in which they either had to (1) respond to messages of relevance to the lesson, (2) respond to messages that were irrelevant to the lesson (3) create messages of relevance to the lesson; or (4) create messages that were irrelevant to the lesson. Each of these groups was also divided into a high-distraction condition, with students receiving or responding to messages every 30 seconds, and a low-distraction condition, with students receiving or responding to messages every 60 seconds.¹ The control group was composed of students who did not have any mobile phones at all.

Under the experiment, all students watched a 12-minute video lecture whilst taking notes on it. After the lecture, the students took a free recall test, a multiple-choice test, and their notes were also graded for quality.

The irrelevant messages to which students responded were: “What is your favourite restaurant for dinner?”, “Comment on this photo”, “What is your schedule on most Tuesdays?”, and “Which dorm do you recommend living in?”. Relevant messages, on the other hand, asked: “What are the two

¹ In the paper, the authors refer to creating messages as ‘tweeting’ and responding to messages as ‘texting’, even though it is possible to create text messages and respond to tweets.
types of uncertainty?” and “What is the name of the theory the professor is talking about now?”

Overall, students in the control group and students who responded to relevant messages did better than the others on all three measures. Students who created relevant messages did just as well on the test, but they also took poorer-quality notes, while students who responded to or created irrelevant messages did worse on the test and note-taking.

What can explain these results? The researchers suggest that creating messages is a more cognitively intense task than simply responding to messages – and thus distracted students more, even if the content of the message was related to the lecture. However, they also suggest that both sending and receiving relevant messages may involve similar process as those that occur during note-taking, allowing students to perform well on the free-recall and multiple choice tests.

The results lead the authors to ‘caution against rushing to integrate texting and Twitter into the classroom.’ Since students who engaged with relevant messages did not outperform the control group, while those who engaged with irrelevant messages underperformed, there appears to be limited benefits of allowing mobile phones in the classroom even when they are used as intended – and merely downsides if they are not used as intended.

Still, the paper does offer some hope for smartphone enthusiasts: the authors suggest mobile phones could be successfully integrated into classrooms if students are given a brief break to compose messages about the lesson content specifically.

The experiment confirms one well-known fact about the human mind: in order to learn something, we have to
concentrate on it. Distractions do not help, and multi-tasking impairs performance. However, the paper also asks us to look at education technology in a new and more fruitful way. Whereas much previous research analyses whether mobile devices are intrinsically good or bad, this experiment looks at how they can be good or bad depending on how they are deployed. In doing so, the paper suggests that the crucial aspect of learning is not the medium used to communicate a topic – but the extent to which students are concentrating on the topic.

Yet it is always possible that technology generates more distraction: students equipped with a pen and paper can doodle, but those with a connected smartphone face many more temptations. As the authors acknowledge, experiments of the kind analysed in the paper do not accurately reproduce the way students use their mobile phone in class. If mobile phones in real-life settings lead to more distraction in the classroom, this would be a significant strike against them. Overall, therefore, the paper’s findings make clear that the burden of proof ultimately rests on proponents of education technology – it is up them to show how such technology will enhance, not divert, concentration.
‘Can personality traits and intelligence compensate for background disadvantage? Predicting status attainment in adulthood’


*Journal of Personality and Social Psychology (November 2014)*

**Published version (free)**

Commentary by: Dale Bassett

‘Character education’ is one of the current hot topics in the education policy world. Personality characteristics such as resilience, conscientiousness, and sociability are thought to have an impact on educational attainment and career success. Many policymakers believe that some socioeconomically disadvantaged children lack the opportunities to develop these traits due to the nature of their upbringing, while those from wealthier backgrounds benefit from home environments that encourage children to develop these characteristics – and that this difference contributes directly to the ‘attainment gap’ between disadvantaged children and their better-off peers. Character education – where schools seek to actively support the development of these ‘desirable’ behaviours – is therefore seen as a possible remedy to reduce the attainment gap.

In this paper, five psychologists attempt to answer the question of whether possession of the noted characteristics really make a difference to educational attainment and
success in later life, or whether any positive effects are overshadowed by the impact of socio-economic background. For the first time, therefore, this research seeks to disentangle the impact of personality traits from that of intelligence – and examines the impact of ‘character’ on young people’s educational and life chances compared with the effect of their socio-economic background.

The authors present an analysis of data from Project Talent, a large-scale nationally-representative longitudinal study that tracked US high school pupils until 11 years after graduation. This allows them to evaluate the long-term impact of teenage personality traits on adult outcomes – such as ultimate educational attainment, annual income and occupational prestige – a decade later. The researchers examined the correlations between socio-economic background, intelligence, and the so-called ‘Big Five’ personality traits (extraversion, agreeableness, conscientiousness, emotional stability, and openness) of the subjects, with their future educational attainment, annual income, and occupational prestige.

The analysis suggests that each of the personality traits has a statistically significant effect on outcomes: a 1 standard deviation (SD) difference in conscientiousness is associated with an extra 2.7 months of education, while a 1 SD difference in extraversion is correlated with around a 4 per cent increase in annual income. An increase in parental SES of 1 SD is in turn associated with an additional 8.3 months of education and 8 per cent of annual income. So character matters – but not as much as socioeconomic (dis)advantage. So far, so unsurprising.

The paper gets really interesting when the authors look at the effects of the interaction between pupils’ socio-economic background and personality traits. They consider
three hypotheses: (1) the independent effects hypothesis, which predicts that character affects educational and life outcomes independently of background, (2) the resource substitution hypothesis, which predicts that personality is more important for outcomes among pupils from lower socio-economic backgrounds (i.e. that character ‘compensates’ for disadvantage), and (3) the Matthew effect hypothesis, which predicts a stronger correlation between character and outcomes among pupils from more advantaged backgrounds (i.e. that character reinforces socio-economic advantage).

The research finds that high levels of three of the Big Five characteristics (agreeableness, extraversion and conscientiousness) have a bigger effect on educational attainment among pupils from low socio-economic background. This confirms the resource substitution hypothesis: having these traits makes a bigger difference if you are less well off, thus ‘compensating’ (to an extent) for your socioeconomic disadvantage. The potential impact is significant: going from -2 SD to +2 SD in each of these traits results in the equivalent of about one additional academic year of education among pupils with the lowest socio-economic background, compared with pupils with the highest. Conscientiousness and extraversion showed a similar effect on annual income: among pupils with the lowest socio-economic background, going from -2 SD to +2 SD in the personality traits results in an 18 per cent increase in annual income compared with pupils from the highest socio-economic background. Again, ‘character’ makes a bigger difference for the socioeconomically disadvantaged. Extraversion also has a resource substitution effect on occupational prestige.

In a first for research in this field, the authors then control for intelligence – measured by a composite of
verbal, quantitative and spatial cognitive abilities – to investigate the robustness of these results. Strikingly, the main effects of the personality traits remain statistically significant – and the effect sizes do not change dramatically. However, the interaction between the Big Five traits and pupils’ socio-economic background no longer is no longer statistically significant in terms of educational attainment or occupational prestige, but the effect on annual income of conscientiousness interacting with background remains statistically significant.

So, when controlling for intelligence, character makes a difference to educational and career outcomes. But the effects are independent of pupils’ background: character does not make more difference to the more disadvantaged, with the exception of conscientiousness, which does benefit the annual income of those from lower socio-economic backgrounds more than those from higher socio-economic backgrounds. It is also important to note that even this effect only partially ameliorates the overall impact of background; it is not large enough to fully compensate for socio-economic disadvantage.

Certainly, there are a number of limitations to this research, acknowledged by the authors, chief of which is the age of the data. The Project Talent data were collected between 1960 and 1971, and both the economy and education system have clearly changed considerably since then. But it does nonetheless raise important questions for policymakers.

If (and it remains a big ‘if’) character can be taught – if education can increase the level of the Big Five personality traits demonstrated by young people – this could have a positive effect on pupils’ educational attainment and career success, irrespective of their intelligence. In principle, this
effect would be as significant for advantaged children as for the very disadvantaged ones, so it would not help to close the attainment gap if applied universally. However, implementing character-building interventions successfully in schools serving primarily disadvantaged communities could have a considerable effect.

The remaining question, then, is just how to ‘teach’ character and whether it is truly possible to develop particular personality traits through education. The research suggests that the work of Sir Anthony Seldon, Lord James O’Shaughnessy, and others trying to answer that question could have an important impact on our education system and the life chances of our most disadvantaged young people.
‘Does Management Matter in Schools?’
– Nicholas Bloom, Renata Lemos, Raffaella Sadun and John Van Reenen

*The Economic Journal (May 2015)*

Published version/Working paper version (free)

Commentary by: Henrik Jordahl

Differences in the quality of education are large, both within and between countries. A recently published study by Nicholas Bloom and co-authors suggests that those differences can partly be explained by management quality. The authors use a survey instrument that has been used to study management in other sectors, and apply it to schools. Management quality is evaluated by structured telephone interviews with headteachers at 1,800 schools in eight countries: Brazil, Canada, Germany, India, Italy, Sweden, the UK and the US.

The management practices under study cover operations management, performance monitoring, target setting, and talent management. Most of the 20 included management practices have appeared in previous studies of companies in various sectors, but a few practices are more school specific, such as adopting educational best practices. Each practice is scored from one (worst practice) to five (best practice) according to a pre-determined scoring grid and the 20 practices are then averaged to an index of school management quality ranging between 1 and 5.
The researchers find that management quality is positively correlated with pupil achievement at age 15 across all eight countries. An increase in the index of management quality by one standard deviation is associated with an increase in pupil performance of equivalent 20-40 PISA points. Compared with other studies, the association of management quality with pupil performance is considerably stronger than that of school competition, teacher quality, and class size. However, it should be stressed that because of the research design utilised – involving cross-sectional data and a couple of relevant control variables – the findings are only suggestive. Interestingly, for UK schools, the authors also analyse school-level value added scores, which have an even stronger correlation with management quality.\(^2\)

Another notable finding is that the management quality of schools varies significantly between countries. The UK, Sweden, Canada, and the US obtain the highest average management scores. Germany is located in the middle, whereas Italy, Brazil, and India obtain the lowest scores. Compared with other sectors, such as manufacturing and health care, the between-country variation makes up a larger part of the total variation of management quality in schools – suggesting that national institutions contribute to shaping school management.

Importantly, there is significant scope for improvement, since the average level of management quality in schools is low. For example, 82 per cent of Indian schools score below 2 on the management index. Such schools apply most management practices only in a primitive way or not at all. They would hardly track and review the performance of its

\(^2\) The findings are in line with those in a study of UK university departments: at universities, there is a positive association between management quality and research/teaching performance.
pupils, set perhaps only one target annually at the school level, base promotion on tenure only, and fail to take action about underperforming teachers.

Unlike the situation in India, no UK schools score below 2 on the management index. Still, there is room for improvement also at the top end of the distribution. In fact, there are no schools in any country with a score close to 5, which is only awarded if all management methods are up to best practice standards. This is also different from other sectors; previous studies of manufacturing have typically identified a limited number of firms with excellent management practices across the board.

Interestingly, the best management practices are found in so called “autonomous government schools”. These schools, which include UK academies and US charter schools, are publicly funded but relatively independent from government control – and they have higher management scores than both government and privately funded schools. The researchers establish that this advantage is not due to pupil composition – and that over 50 per cent of it can be attributed to (1) stronger performance accountability and (2) the leadership, in the form of a coherent and effectively communicated long-term school strategy. The difference between autonomous government schools and other schools is especially large in Sweden. This may be due to the country’s voucher reform in the early 1990s, which led to a gradual expansion of (often for-profit) autonomous government schools (friskolor) – the benefits of which may take some time to be realised.

The authors conclude that management quality appears to matter for schools. And unlike input factors, such as class size and teacher quality, altering schools’ management practices is a low-cost way to improve performance. The
study’s findings also indicate the importance of increasing autonomy among state schools – which appears to generate better governance and management, and, as a result, higher pupil performance.
‘An RCT evaluation of the *Numeracy and Literacy* project’

– Jack Worth, Juliet Sizmur, Rob Ager, and Ben Styles

*Education Endowment Foundation (July 2015)*

**Published version (free)**

Commentary by: Jack Worth

Randomised controlled trials (RCTs) are a growing phenomenon in education research. And the debate now appears to have moved beyond the question of whether or not such trials are desirable to one about whether or not it is possible to run robust trials in practice. The NFER Education Trials Unit’s evaluation of the *Improving Numeracy and Literacy* programme demonstrates that it is indeed possible.³

The *Improving Numeracy and Literacy* programme consisted of two separate teacher-training programmes, developed by Professors Terezinha Nunes and Peter Bryant of the University of Oxford, which were designed to improve pupils’ numeracy and literacy ability. Previous, small-scale interventions by the developers indicated that both interventions had promising effects on measures closely aligned to the concepts being taught. The purpose of the RCT was to evaluate the efficacy of the interventions when delivered by teachers in a whole-class situation.

---

³ The trial was one of many randomised trials sponsored by the Education Endowment Foundation, an independent charity set up by the Department for Education,
The interventions were aimed at teachers of 6-7 year old Year 2 pupils, and included a day of teacher training, resources for delivering ten lessons, and accompanying computer games for the pupils to use in class and/or at home. The *Mathematics and Reasoning* programme aimed to develop children’s understanding of the logical principles underlying mathematics, and the *Literacy and Morphemes* programme aimed to improve spelling and reading comprehension by teaching children about sentence structure and morphemes (units of language that convey meaning).

The Oxford team recruited 55 participating schools, and once all pupils had been tested in literacy and numeracy in the autumn, the NFER randomly allocated a third of schools to the literacy intervention, a third to the numeracy intervention, and the rest to the business-as-usual control group. For those randomised to one of the intervention groups, teacher training took place in December and the intervention was delivered in the spring term. After Easter, NFER test administrators tested all pupils again in numeracy and literacy.⁴

The results showed that pupils in the numeracy intervention schools made 0.2 standard deviations more progress in numeracy than pupils in the control schools, which is roughly equivalent to three additional months’ progress (or 20 PISA points higher performance). The pupils who received the literacy programme made slightly less progress in literacy than the control group, but this difference was too small to confidently conclude that it was caused by the intervention.

⁴ Importantly, the test administrators did not know to which group schools had been allocated. The gold standard RCT in medicine is ‘double-blind’, which means that neither participant nor researcher knows whether the patient receives the drug or placebo. Double-blind trials are impossible for most educational trials, but blindness on part of researchers and administrators can be achieved in most cases.
Since pupils in all groups were tested for both numeracy and literacy, we were also able to estimate the ‘spill-over’ effects of the numeracy programme on literacy, and vice versa. However, our analysis shows that there were no statistically significant spill-over effects.

Additionally, we analysed the association between greater use of the accompanying computer games and the effects of the interventions. Interestingly, the results suggest that pupils who played more games also made more progress in numeracy. However, this analysis is susceptible to selection bias: pupils playing more games also had on average higher pre-test scores and were less likely to have low socio-economic background. The positive association may thus be due to these underlying differences rather than the causal impact of playing more games.

Importantly, because all 55 schools in this project completed the testing, the results cannot be biased due to some of their dropping out of the trial prior to completion. Also, test scores were only missing for about one in eight pupils, which is consistent with normal levels of daily absence and pupils leaving schools. In other words, we can be quite confident that the effects detected represent a causal effect of the numeracy intervention on learning.

While the intervention had a promising impact on test scores – and given its low cost also seems to demonstrate good value for money – some questions remain. For example, could the positive results merely be a statistical fluke? And are the results generalisable to all types of school? These are valid questions that cannot be answered by this study.

But this ‘efficacy’ trial is not intended to be the final word on the topic – it aimed to test whether the intervention works under ideal conditions, with intensive support from
its creators. The next step would be to run an ‘effectiveness’ trial, which seeks to test whether the intervention can work at scale in a large number of schools, when others apart from the developers act as deliverers. Testing the intervention using a scalable model in another RCT would (if successful) add further credibility to the relevance of this study’s findings.

The *Improving Numeracy and Literacy* evaluation adds weight to the argument that it is both desirable and feasible to run robust educational RCTs in English schools – and that such evaluations can help identify programmes that can have positive effects on pupil attainment. However, while RCTs are often seen as the gold standard for credibly demonstrating intervention effects, the results of one trial are just the starting point for policy development; the most important contribution of education experiments will be the body of evidence that is generated from the running and replication of many trials over time and in different settings.
Over the past decade, international comparisons of education have assumed an increasingly important role in domestic policymaking. The big international surveys – principally PISA and TIMSS – have come to dominate thinking, not least through league tables that rank countries based on their pupils’ performances. Despite the fact that these league tables are the least analytical element of the surveys, they still have an almost magnetic appeal among policymakers. In recent years, focus has especially been directed at the Asian countries, and the ways in which they conduct mathematics education.

Albeit a couple of years old, this apparently modest paper focuses on the way in which fractions are taught in Shanghai. It presents one of those seemingly small insights, which nonetheless have significant implications for curriculum thinking and practice. It focuses on ‘elaboration theory’: the way in which practice problems are varied to encourage children’s understanding of underlying mathematical relationships, which enable them to attain
secure, and early, mastery of important mathematical operations and techniques.

The paper presents a fascinating illustration of how Chinese children work on all four functions with fractions – addition, subtraction, multiplication, and division – simultaneously in the last two years of primary education. The author does not locate this approach in a new or emerging theory, but in a thousand-year old philosophy of mathematical reasoning. To make children work on all four functions at the same time contrasts with traditional sequencing in English education, where children instead work on the functions in sequences that span over several years – with only a minority of children even getting to, let alone mastering, the division of fractions.

In Shanghai, rather than learning separate techniques, pupils are encouraged to see how adding, subtracting, multiplying, and dividing fractions are related. Previous research on mathematical misconceptions in England shows that many secondary children struggle with the division of fractions. For example, Jeremy Hodgen’s penetrating analysis of English children’s understanding of proportional reasoning shows a decay of facility over the last two decades.

And here, in this small study of Shanghai mathematics, we find a well-theorised, carefully-refined curriculum practice, which challenges our own approach to sequencing of content, age expectations, and practice activities. It contains important insights, which present profound challenges to the ways in which we approach mathematics education in our schools.
The merits, or otherwise, of school voucher programmes have been the subject of (often heated) debate for many years. The issues are complex, and comprehensive, balanced and fair appraisals of the evidence of their effectiveness, potential adverse effects, and trade-offs, are rare. In this substantial body of work, taking in the full range of theoretical, computational, and empirical research on voucher measures, the authors make a sensible case for continued exploration without ignoring the many rigorous studies that have found effects of vouchers on educational outcomes to be insignificant, and the challenges of designing programmes to limit or counteract the adverse effects of sorting.

Beginning with an overview of theoretical arguments for and against vouchers, the authors show how the limited nature of many targeted reform measures means that little headway has been made research-wise in resolving key areas of debate, which tends to assume universal availability. Small-scale, targeted programmes are less likely to produce any sorting effect, so evaluation is correspondingly less likely to address concerns about
stratification. In the same way, measures that do not stimulate extensive new supply are unlikely to result in much curricular variety. Evaluation is therefore not going to yield insight into potential gains from choice effects – that is, whether finding the right learning environment for the child is as important for outcomes as proponents of school choice suggest. This doesn’t mean targeted measures don’t have value. Indeed, empirical evidence to the effect that in some settings, or for some subgroups or specific outcomes, vouchers can have a substantial positive effect, informs an overall optimism about targeted measures. The authors suggest that we may need a different frame of reference when it comes to evaluating their effects.

Attention to context, and the specific characteristics of the programmes being evaluated, is shown as crucial to improving the design of future measures. With reference to characteristics of established programmes in eight different countries, the authors’ commentary on the empirical research undertaken to date shows how eligibility criteria, how vouchers are funded, and the criteria for private school participation, are all key considerations for evaluation.

The case for continued exploration rests on several grounds. While most studies have found learning gains to be insignificant, vouchers are very rarely associated with negative effects for pupils, and in some cases can have significant positive effects. By isolating the conditions under which voucher programmes are most likely to be successful in inducing gains through greater competition, and attending to the circumstances in which sorting is most likely to undermine effectiveness, we can design for improved functioning.

In respect of the authors’ emphasis on the consensus that some sorting is inevitable, this optimism is especially
encouraging. It is all the more so given the limited consideration given to the potential of a ‘positively discriminating’, means-tested voucher designed to compensate for cream-skimming and incentivise efforts towards raising attainment among disadvantaged pupils. This is an especially promising avenue, albeit one that may involve trade-offs in respect of savings to public education generally expected of voucher programmes.