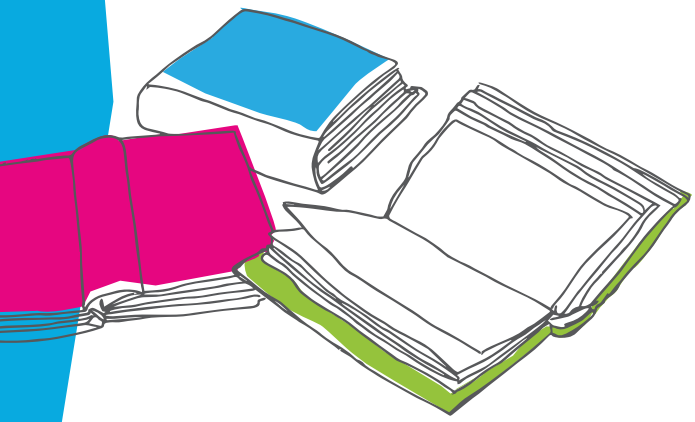
education datalab

research • analysis • evidence

SEVEN THINGS YOU MIGHT NOT KNOW ABOUT OUR SCHOOLS

WHO WE ARE

Education Datalab, hosted by FFT, produces independent, cutting-edge research on education policy and practice using administrative and survey data.



The National Pupil Database in England is one of the richest education datasets in the world. With detailed information about seven million school pupils matched over 12 years, it is used widely to support school improvement, inform education policy and for research. The recent government focus on open data is now encouraging many more researchers to start using the NPD and other administrative datasets.

FFT is a not-for-profit organisation that has been using the national education datasets in England and Wales for 14 years. We provide analyses that help schools, local authorities and academy sponsors to evaluate and benchmark their performance. Our aim is to help schools make the best possible use of education data to improve education outcomes for all pupils. Over 80% of schools in England and Wales regularly use FFT's analyses to support school improvement.

We are launching the Education Datalab to bring together an expert team of academics, researchers and statisticians specialising in the analysis of large-scale administrative and survey datasets. Led by Dr Rebecca Allen, Education Datalab will produce independent, cutting-edge research that can be used by policy makers to inform education policy, and by schools to improve practice. We will work collaboratively with research partners and make sure that our published research is accessible to policy makers and schools.

I hope that the Education Datalab's first report about the 'seven things you might not know about our schools' will raise questions and encourage debate amongst all those working in education.

Paul Charman
Managing Director, FFT

MESSAGE FROM THE DIRECTOR

Education Datalab brings together a group of expert researchers who all believe we can improve education policy by analysing large education datasets. Many of these are administrative databases held by Government, such as the National Pupil Database and School Workforce Census – others are large surveys like the Programme for International Student Assessment (PISA) and Labour Force Survey.

Rich data on our schools has been available for over a decade now but, as we hope this report shows, there is still much we can learn. Every piece of analysis we show in this report is the sort of thing that can be done in an afternoon, provided you have access to longitudinal databases with indicators that have been created in a consistent way to monitor changes as the way we measure attainment changes.

Some of the 'seven things you might not know about our schools' are little 'facts' that fell out of the data whilst looking at something completely unrelated. Some are idle questions people at FFT asked themselves whilst preparing pupil data for other purposes. Other pieces arose from conversations with policy makers.

We have asked respondents to comment on each piece of analysis. We did this because no piece of research produces a definitive answer and usually it throws up more questions worth exploring. We aim to turn curiosity about education into quantitative analysis. We hope you enjoy the report.

Dr Rebecca Allen
Director, Education Datalab

NORTHERN LOCAL AUTHORITIES WILL MAKE HUGE IMPROVEMENTS SIMPLY BY FILLING THE ATTAINMENT 8 SLOTS

Attainment 8 and Progress 8, the new Key Stage Four school accountability measures due to be introduced by Government in 2016, will undoubtedly make a difference to how schools enter their pupils for qualifications. They are reminiscent of the old 'best eight or capped' and 'contextual value added' measures in that they judge a school across a large number of subjects taught. But they are more restrictive because only some subjects count for inclusion: the first two slots must be filled by English and maths, the next three by EBacc subjects of science, computer science, history, geography, or languages, and the final three by any other GCSEs or eligible qualifications.

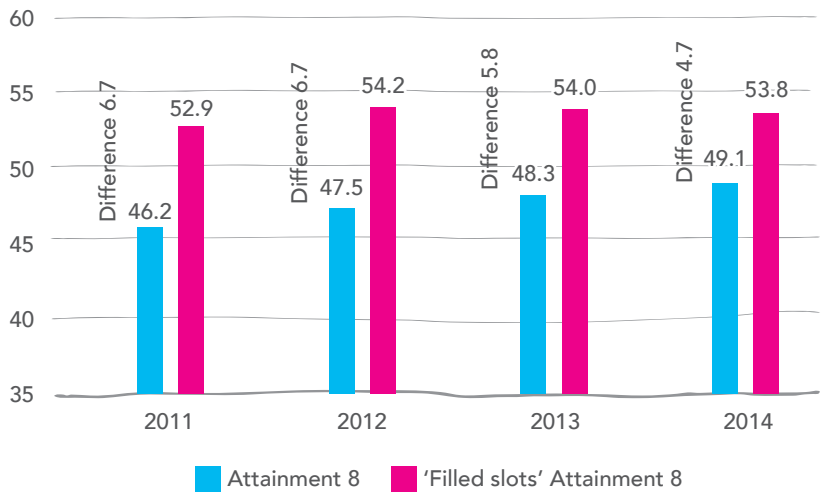
These new performance tables will, overnight, alter which schools are deemed to be doing well or poorly. And inevitably schools will adjust their curriculum offer to raise Attainment 8 (A8) performance. We predict that the new performance measures will change the narrative about which regions and local authorities are improving. Here, we focus on how adjusting subject entry patterns to 'fill the slots' can (and will) yield large improvements in Attainment 8 without any associated improvement in teaching quality. It is better for a pupil to enter at least five EBacc subjects, even if they are likely to do relatively poorly, because every grade from G upwards counts in the new measures.

SCHOOLS HAVE ALREADY STARTED FILLING THE ATTAINMENT 8 SLOTS

Although the Attainment 8 measure has only recently been announced, back in 2010 schools started altering the Key Stage Four curriculum to allow more of their pupils to have the chance of inclusion in the EBacc performance measure. This has resulted in Attainment 8 improving nationally from 46.2 in 2010 to 49.1 in 2014, even though it wasn't yet devised as a performance monitoring tool.

If schools ensured that every pupil was entered for eight eligible qualifications and that the pupil managed to achieve the same average grade in them as they do for their current Attainment 8 subjects, then Attainment 8 would clearly rise further. We call this measure of the potential Attainment 8 achievable, holding constant average grades, the 'filled slots' Attainment 8 measure. In 2014, the Attainment 8 score would have risen by 4.7 to 53.8 if all Attainment 8 slots were filled by all children in all schools.

National actual and 'filled slots' Attainment 8 scores (using best rather than first entry grades)



NORTHERN REGIONS HAVE MORE OPPORTUNITY TO IMPROVE ATTAINMENT 8

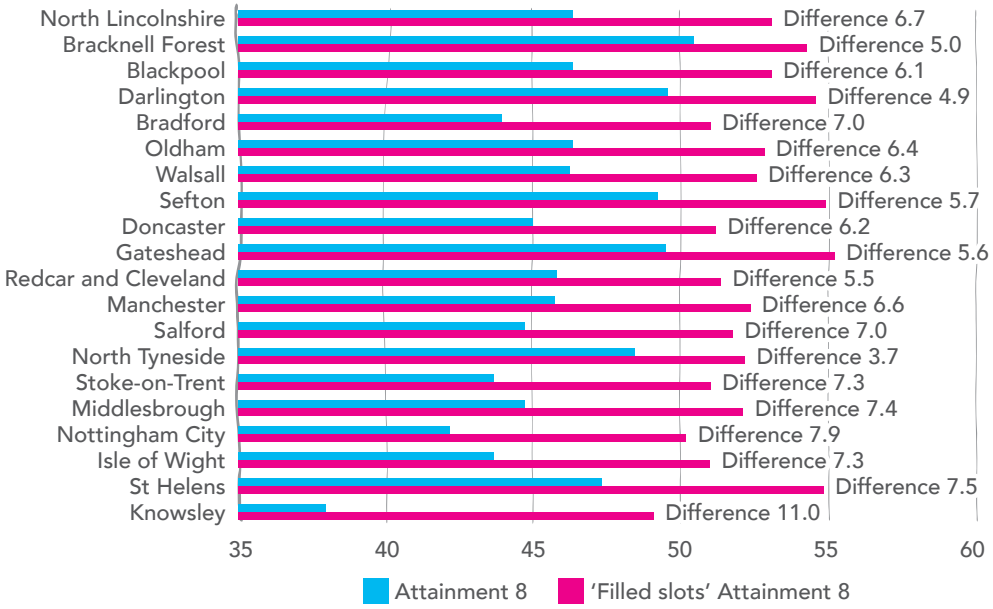
Current regional differences in performance on Attainment 8 follow well-known patterns. Overall, achievement is higher in London and the south of England than it is in the north and the Midlands. However, those regions that are currently performing poorly also have the greatest opportunity to make rapid improvements in performance, simply by 'filling the slots' through curriculum change.

Differences between Attainment 8 and 'filled slots' Attainment 8 by region

WE THINK THESE LOCAL AUTHORITIES SHOULD ACHIEVE THE BEST IMPROVEMENTS IN ATTAINMENT 8

Those local authorities where schools offer a curriculum that is poorly aligned with Attainment 8 have huge opportunities to improve their performance, should they want to. There are no London and few southern local authorities on the list. Some areas will argue that a more traditional curriculum is not appropriate for children in their area or that they are not well equipped to deliver this new curriculum. It is interesting to note that most local authorities were offering a curriculum better aligned with Attainment 8 in 2004 than they were in 2014. At the extreme, in the past decade there has been over a 20 percentage point fall in the number of pupils entered for a full set of Attainment 8-aligned subjects in Bracknell Forest, Oldham, Doncaster and East Riding of Yorkshire. By contrast, Islington has seen a huge 23 percentage point increase over the same period.

Actual and 'filled slots' Attainment 8 points for 20 local authorities with the most potential to improve



BEWARE OF LONDON'S DETERIORATING PROGRESS 8 MEASURE

Progress 8 is calculated for each child by comparing their Attainment 8 score to that of all other pupils scoring the same Key Stage Two fine grade. This is then averaged across all pupils to give the school's Progress 8 measure. As an accountability measure, a value below -0.5 (i.e. pupils scoring on average half a grade below expectations or entering fewer than average number of subjects) will be the new floor standard, triggering scrutiny through inspection.

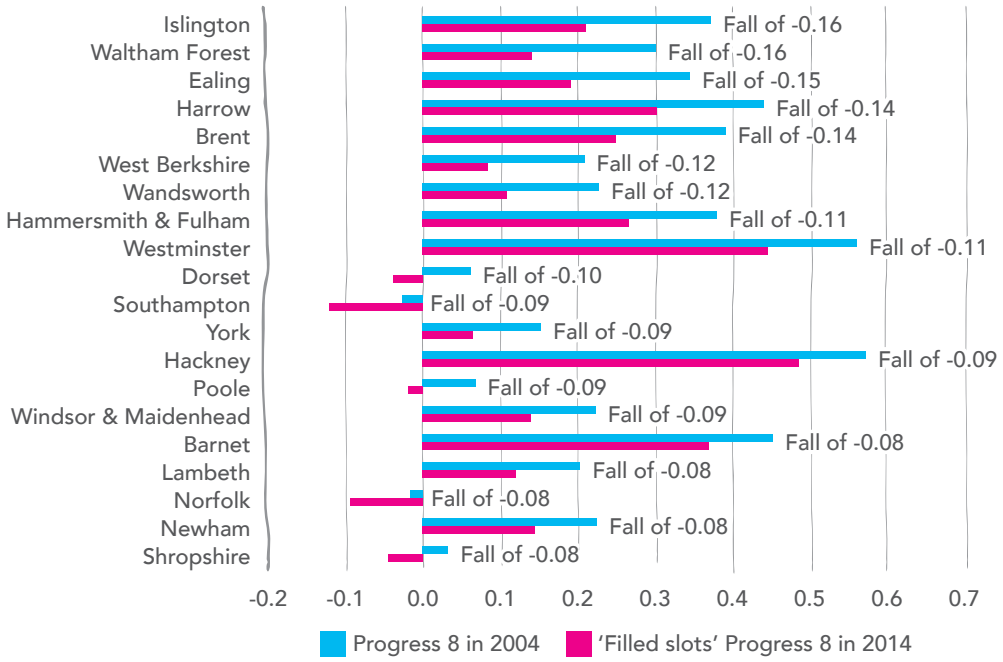
In 2014, almost 300 schools would have fallen below the floor standard (even taking the best entry of the pupil, rather than their first entry in the subject). As more students enter for more eligible qualifications, the distribution of Progress 8 will shrink – at least 100 of these schools below floor standard can remove themselves from scrutiny through curriculum reorientation alone.

London performs well on Attainment 8 and will continue to perform well even as schools in other regions start to realign their curriculum to the new performance measures. However, as some schools enter more students for qualifying subjects, the comparison for those schools and areas already doing well becomes increasingly difficult. The areas where there is little room for improvement through 'filling the slots' are mostly in London and so it is reasonable to assume that many schools here will see deteriorating year-on-year Progress 8 scores.

These new accountability measures will force schools to think again about how they enter pupils for qualifications, and will encourage them to ensure progress is made by pupils at all levels, rather than just at the C/D grade threshold. The measure should allow the rest of England to start to catch up with London by better aligning their curriculum to fill the Attainment 8 slots.

By contrast, schools already offering a traditional curriculum can only improve by raising individual subject grades. This presents a rather unanticipated risk: schools with a traditional curriculum – those with higher entry attainment of pupils, grammar schools, schools in London – focus their energy on improving teaching and grades since this is the only way to raise Attainment 8. Meanwhile, the reorientation of the curriculum distracts others from maximising individual subject grades. Thus, Attainment 8 thus converges, yet average subject grades diverge.

These local authorities will need to explain why Progress 8 is deteriorating, even if Attainment 8 is improving



'Progress and Attainment 8' presents a new challenge in the heads' Herculean game of trying to square what's good and fair for the individual with what's good for the school. They want both, but they can't always have it.

This will tip the odds slightly in favour of getting it right for the pupils, particularly those who at present gain least from the system: the biggest points gain for the school is going to lie in raising the results of those pupils who arrive with less rather than more prior attainment. And not publishing five or more will change the media-driven public's perception of how a school should be judged on exams, which in an ideal world should be taken when pupils are ready, rather than at a pre-determined age.

Tim Brighouse, Norham Fellow at University Department of Education Oxford

Measurement is rarely neutral. People and organisations respond to being measured. Sometimes, if policy makers have done a good job, they respond in the way intended. But people are smart and will respond to the fine detail of whatever is measured, rather than to the spirit.

This analysis explores the implications of some possible reactions of schools to the new accountability framework. It assumes that schools will make use of an easy hit to raise their Attainment 8 and Progress 8 scores by filling any currently-empty Attainment 8 slots – they're almost certainly right.

To understand the full effect of such a move by schools, we need to know what schools and pupils would be giving up to move to what the piece calls a more traditional curriculum. Was what the pupils were doing before truly valueless? Could teachers of vocational qualifications that have been 'Wolfed' simply turn to teaching history instead? While the broad prediction of a relative improvement in parts of the north of England seems right, there will be enough twists and turns to keep all us analysts very busy.

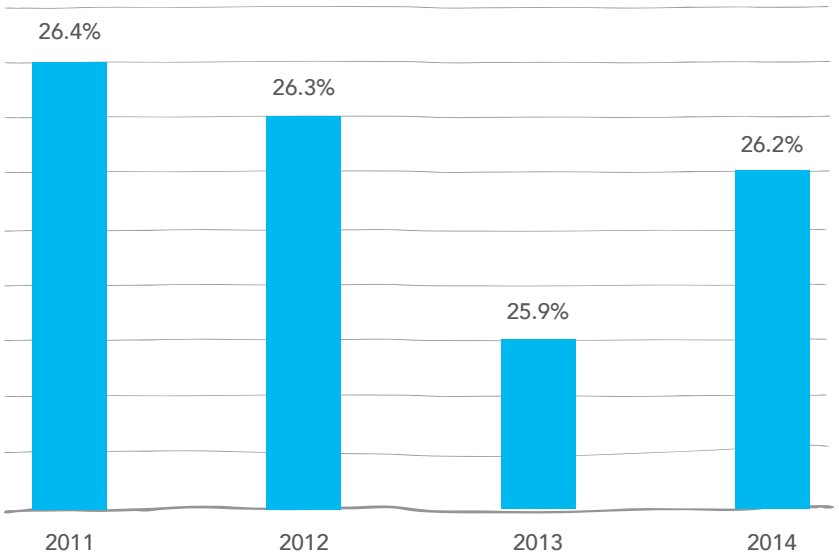
Simon Burgess, Professor of Economics, University of Bristol

WE ARE CLOSING THE PUPIL PREMIUM GAP – IF WE LOOK IN THE RIGHT PLACES

This Government has invested enormous amounts of money and political capital in closing the attainment gap between children from low-income families, and everyone else.

They give schools a pupil premium for children eligible for free school meals (and some other vulnerable groups) now worth £1300 for primary pupils and £935 for secondary pupils. They gave £125 million to create the Educational Endowment Foundation to find out how best to spend this money. Schools are now judged by Ofsted on what they are doing to close the gap and how far they have succeeded. But despite all this investment, commentators are arguing that the policy is failing because the pupil premium attainment gap isn't closing. Using the performance metric of the proportion of children achieving five or more A*-C grades at GCSE including English and maths (5+A*-C EM), commentators appear to be right: the gap is falling so slowly that it will take about 250 years to succeed in closing the gap.

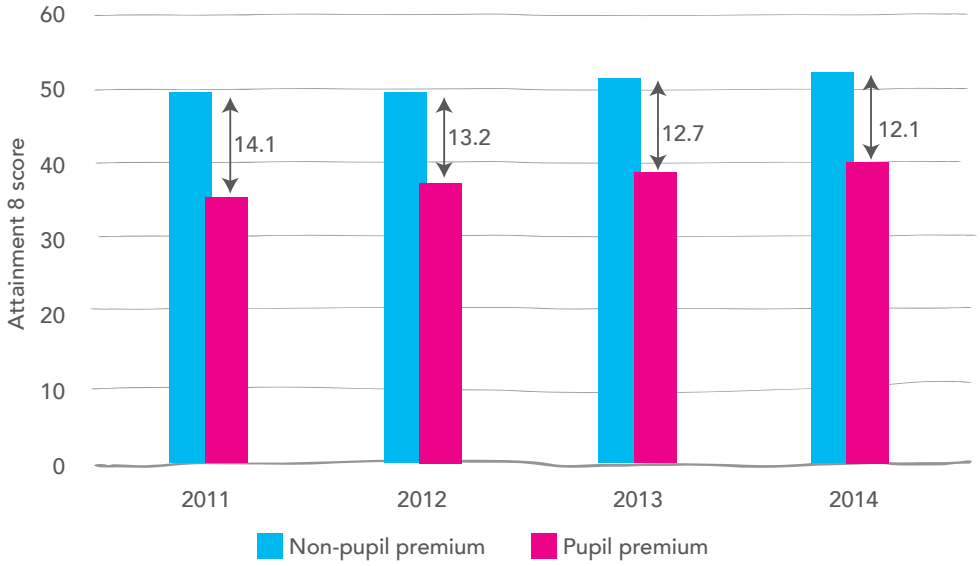
At first glance the gap between % 5+ A-C GCSEs, including English and maths, for non-pupil premium and pupil premium children isn't closing*



ON NEW ACCOUNTABILITY MEASURES THE GAP IS CLOSING SO FAST THAT IT WILL BE ZERO BY 2032!

From 2016 onwards, school performance will be judged on pupil grades across eight subjects: English and maths, three subjects from science, computer science, history, geography and languages, plus any other three subjects. On this Attainment 8 measure, the gap has been narrowing fairly consistently each year. This gap has been closing particularly rapidly for children achieving a Level 4B or better in Key Stage Two tests at age 11.

The Attainment 8 pupil premium gap has been steadily falling



WE SHOULD NOT USE FIVE OR MORE A*-C TO MONITOR THE GAP SINCE IT IGNORES ALMOST ALL IMPROVEMENTS

Obviously getting five or more A*-C grades at GCSE is important for a pupil because it enlarges their future educational opportunities, but it is a poor measure for describing how an education system is performing.

It is a threshold measure only capable of changing when a student successfully achieves a C grade instead of a D grade, and not if they achieve an E rather than F or indeed an A rather than a B grade. For many children, it is their grade in English or maths that prevents them achieving five or more A*-C, including English and maths. This means the school's performance in this threshold measure hangs on the performance of one maths and one English teacher, each teaching the C-D borderline ability set for their subject.

Since some pupil premium children are very low attaining, it is very hard for a school to bring large numbers over the five or more A*-C threshold, even if they make very substantial improvements to teaching. By contrast, the grades of all pupils across a wide range of subjects contribute to Attainment 8 success, so it successfully identifies improvements even where they are happening for those pupils at the bottom (or top) of the attainment distribution.

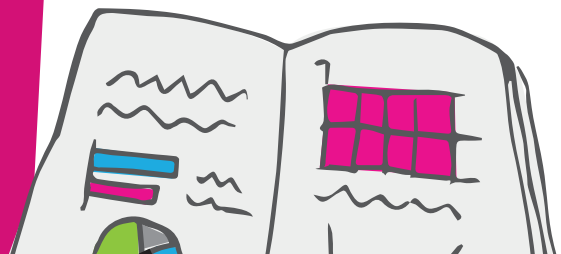
ATTAINMENT 8 IMPROVES BECAUSE GRADES IMPROVE AND BECAUSE SUBJECT ENTRY MIX IS CHANGING

Schools have been entering pupils for more traditional or academic qualifications as a consequence of the English Baccalaureate performance measure. This is truer for their disadvantaged pupils simply because this group was less likely to be following this curriculum in the past. The gap in the number of Attainment 8 qualifying subjects has fallen from 1.13 subjects in 2011 to 0.81 subjects in 2014. By simply entering their pupils in for more qualifications, some of the gap can be narrowed. In fact, the pupil premium gap in entry patterns has now almost closed entirely for pupils with very high prior attainment.

However, performance in exams, as measured by average point score (APS), also matters. Here, the gap is also narrowing but more slowly from an average APS gap of 1.15 in 2011 down to 0.99 in 2014. In 2014, the closing of the APS gap stalled, perhaps due to changes in entry policy resulting from the Government's announcement that they would only count a subject first entry in performance tables. However, overall both the narrowing of the entry gap and the narrowing of the average point score gap have contributed to the closing of the Attainment 8 gap. This is true for groups of children across all Levels of Key Stage Two prior attainment.

That said, the pupil premium gap will not close in the next 20 years. At some point, subject entry mix across different types of schools and pupils may largely converge, but grades achieved in subjects will not. Over the next couple of years, it is possible that Ofqual's comparable outcomes policy will enforce an approximately similar distribution of grades awarded in a subject each year. This means that the attainment of pupil premium children can only improve if that of some other children falls.

We need to be realistic about the extent to which schools can compensate for differences in social background and parental support that families are able to provide, regardless of how much money or how many incentives we give them to close the gap.



Schools are putting enormous effort into making the pupil premium succeed. The gap at 11 is closing and the news that the gap at 16 is closing year-on-year too will bring great encouragement. It is very important that this success is recognised in debates about the future of the pupil premium.

John Dunford, Pupil Premium Champion

It feels unfair to ask schools alone to close the pupil premium achievement gap when differences in out-of-school support for learning and development are so stark. This is particularly true in rural areas. Beyond schools, children need an enriching pre-school experience, great role models, a place to be able to complete homework, access to local services and public transport, a stable home life with parents that care about schooling, broadband and computer access, and access to wider community facilities such as libraries. We hope that the introduction of Progress 8 will provide secondary schools such as ours a more realistic target of equalising attainment for children with a similar starting point.

Liam Collins, Headteacher,
Uplands Community College

Measuring the attainment gap in terms of a threshold has always carried the danger of incentivising schools to focus too much on their middle attainers, those pupils most likely to cross the D/C grade boundary. So it is a pleasant surprise that the gap has in fact been narrowing in terms of the new broader academic Attainment 8 measure, which captures poorer pupils right across the academic spectrum, from high achievers to those really struggling in the classroom.

This gap however also only tells part of the story – there also exists a stark and widening divide between the most privileged pupils and the rest of children from low and middle-income homes. The evidence reveals a recurring trend: gaps take time to close, and other gaps emerge when they do. In the social mobility arms race, the privileged find new ways to gain advantage both within and increasingly outside the school gates. To nurture all our children, we will always need to mind not one but several gaps.

Dr Lee Elliot Major, Chief Executive of the Sutton Trust

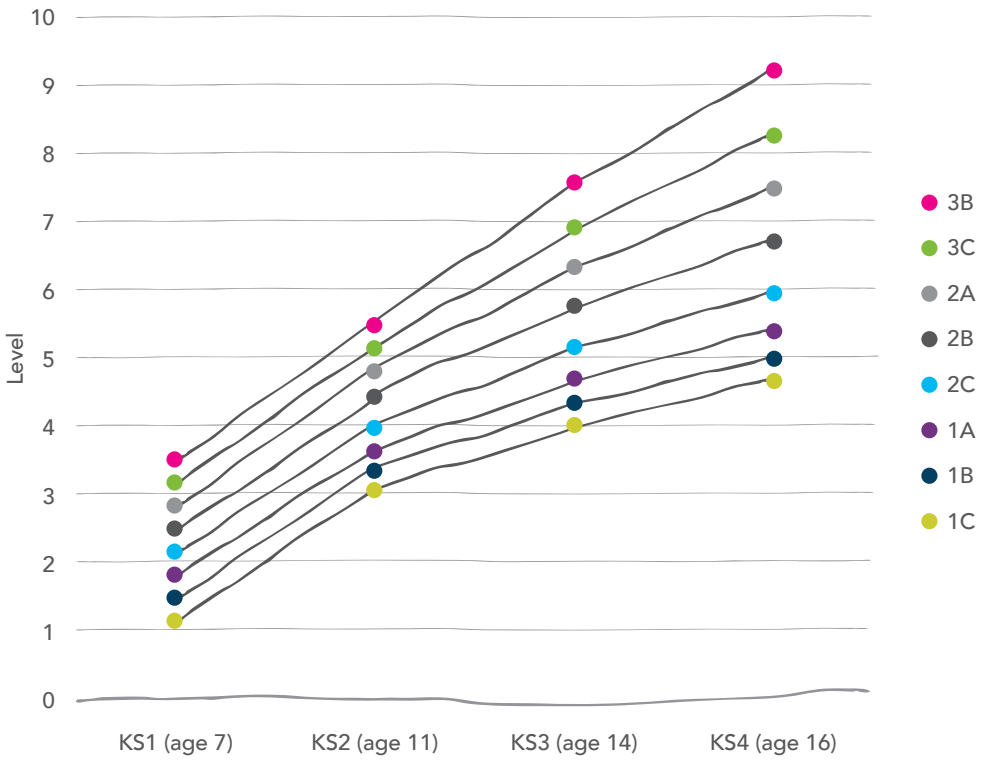
WHY MEASURING PUPIL PROGRESS INVOLVES MORE THAN TAKING A STRAIGHT LINE

We have an accountability system that has encouraged schools to check that children are making a certain number of sub-levels of progress each year. This is the basis on which headteachers monitor (and now pay) teachers and on which Ofsted judges schools. Yet there is little hard science underpinning the system in use: take a child's attainment at Key Stage One (age 7), look up the average attainment for children at the same level by Key Stage Two (age 11) and draw a straight line between the two assuming that linear progress will be made in each of the four intervening years.

For example, a child deemed to be working at Level 2C at Key Stage One is expected to reach Level 4C by their Key Stage Two tests, Level 5C by Key Stage Three and at least Level 5A by Key Stage Four.

But do children normally take such smooth learning journeys as they acquire knowledge and understanding in a subject as our accountability system assumes? And is it reasonable to deem children as 'on target' or 'in need of intervention' using this approach?

Linear progress of pupils from their Key Stage One assessment forms the basis of targets



Note: Average attainment, grouped by Key Stage One sub-level, for children age 16 between 2008 and 2010

Key Stage One = average English and maths levels; Key Stage Two and Key Stage Three = average English and maths test fine grades; Key Stage Four = average English and maths GCSE grades (A*=10; G=3)

ONLY 1 IN 10 PUPILS MAKE LINEAR PROGRESS AT EACH KEY STAGE

Linear progress at each Key Stage can be defined as attaining within one third of a Level (i.e. +/- a sub-Level) of the national average for all pupils who start with the same Key Stage One attainment. Yet by reviewing the data, we find that only 9% of pupils take the expected pathways through Key Stage Two, Key Stage Three and Key Stage Four Levels. If we look at each Key Stage separately, 55% of pupils make the anticipated linear progress to reach the Key Stage Two Level that is predicted for them from their Key Stage One score. The rest either overperform or underperform.

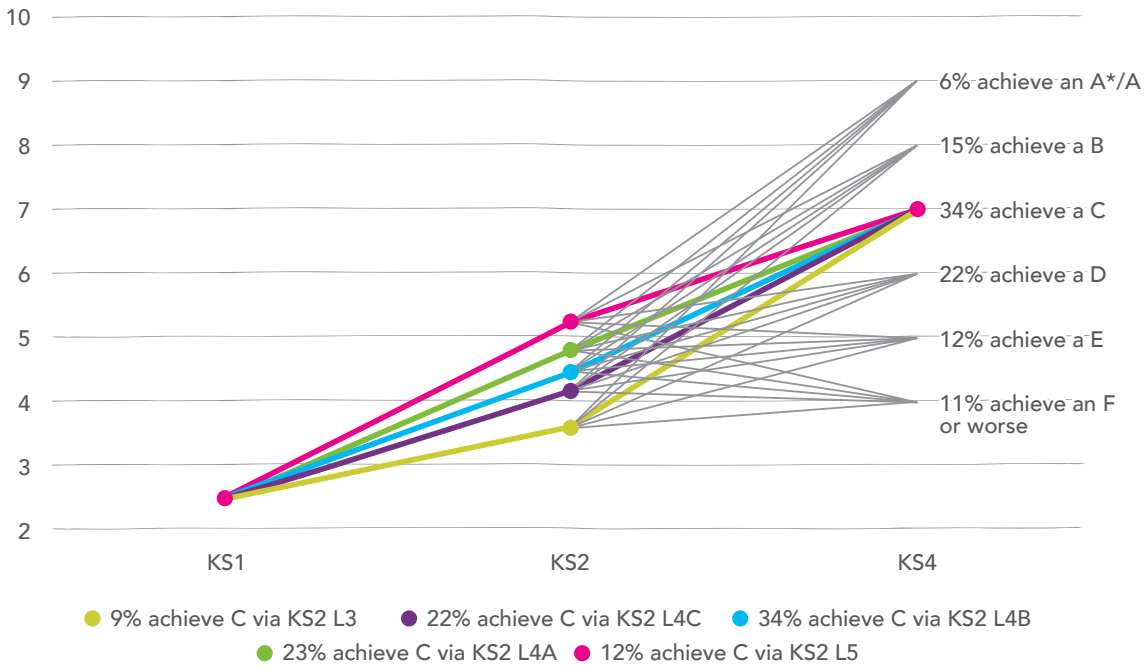
Moreover, the capacity of the model to accurately predict pupil's attainment falls in secondary schools, with 45% of pupils making the anticipated linear progress between Key Stage Two and Key Stage Three, and just 33% making the anticipated linear progress between Key Stage Three and Key Stage Four. This suggests that the model's assumption of consistent progress for a group of children at the same starting point is already weak between the first two stages of schooling, and that the numbers of children judged to be outperforming or underperforming targets is higher still in secondary school.

THERE ARE MANY PATHWAYS TO ACHIEVING 'EXPECTED' ATTAINMENT

By using this model of consistent linear progression, most children will perform better or worse than their expected Key Stage attainment on one or two occasions, and some consistently outperform or underperform throughout their school career. Others do achieve the level of GCSE attainment that we might expect, given their age 7 starting point, but their route to doing so is far from linear and predictable. Here, we show the numerous pathways to achieving an 'expected' age 16 attainment for the large group of children receiving a Level 2B at Key Stage One, tracking their progress through to their average GCSE grade achieved. These children would be expected to achieve a Level 4B at Key Stage Two, a Level 5A at Key Stage Three and average GCSE grades just below a grade C.

One third of these children will indeed get an average grade C at Key Stage Four. But of these children who meet their predictions, the majority will do so via a route that includes periods of both slower and more rapid progress. If targets are simply set based on the last Key Stage test results available, this leads to almost all children being deemed as underperforming at some stage of their schooling career.

More children get to the 'right' place in the 'wrong' way, than get to the 'right' place in the 'right' way!



CHILDREN WITH LOW INITIAL ATTAINMENT HAVE PARTICULARLY UNPREDICTABLE FUTURE ATTAINMENT

The predictability of progress and attainment as children pass through schools is particularly poor for those with low levels of attainment at Key Stage One. For children achieving a Level 1C, B or A at this stage, their development is so unpredictable that most will either outperform or underperform any Key Stage Two target that might be set. It would seem important that these children are not unthinkingly receiving curriculum restriction through placement in lower ability teaching groups or given low targets for attainment, because many of them will go on to achieve later in their school career.

THE SETTING OF EXPECTATIONS AND TARGETS FOR PUPILS MUST BE DONE WITH CARE

Our evidence suggests that the assumptions of many pupil tracking systems and Ofsted inspectors are probably incorrect. The vast majority of pupils do not make linear progress between each Key Stage, let alone across all Key Stages. This means that identifying pupils as "on track" or "off target" based on assumptions of linear progress over multiple years is likely to be wrong. This is important because the way we track pupils and set targets for them influences teaching and learning practice in the classroom, contributes to headteacher judgements of teacher performance and is used to judge whether schools are performing well or not. Providing pupils with the curriculum diet that is deemed suitable for the 'Level' they are working at may be doing them a profound disservice, if in reality their trajectories are much more varied.

Of course, this data tells us nothing about why pupil learning trajectories are so diverse. Children are likely to make cognitive leaps and pauses at different times for a variety of reasons. Indeed, if researchers were able to observe termly test score data, the patterns we show above would be accentuated even more. There are also more practical reasons as to why children do not track through the Key Stage levels as smoothly as we ask them to: all tests have a considerable measurement error and assessments may not be good matches for the knowledge and skills we might ideally measure.

All of the points raised above highlight the importance of using tracking systems carefully, putting to one side 'average progress' as the key target by which children should be judged where it clearly doesn't mirror teacher experiences of the child's potential. Monitoring systems that trigger rewards or warnings if deviation from the mean average takes place can only work if these deviations are relatively rare. The way that children learn is too idiosyncratic to do this and so pupil target setting should be more flexible and take into account a range of likely outcomes rather than a single number.

The variation in children's learning trajectories varies by initial starting point of the child

		Proportion of pupils making linear progress between Key Stages:			
Key Stage 1 sub-levels	Number of pupils	Key Stage One - Key Stage Two	Key Stage Two - Key Stage Three	Key Stage Three - Key Stage Four	Between every Key Stage
1C	54,830	31%	33%	20%	3%
1B	159,933	36%	38%	24%	4%
1A	211,280	40%	41%	27%	5%
2C	473,986	44%	43%	34%	7%
2B	1,036,126	54%	46%	34%	10%
2A	689,539	60%	47%	35%	11%
3C	691,381	69%	48%	36%	12%
3B	3,598	86%	63%	44%	24%
All children		55%	45%	33%	9%

Note: we exclude Key Stage One Levels W, 3A and 4C for brevity in this table but they are included in the total

For example, we could monitor whether pupils are making progress within the range of attainment levels that is the case for, say, 60% of pupils with similar prior attainment. Pupils making progress in the 20% above or 20% below these ranges could then be more reasonably identified as overperforming or underperforming and in need of investigation to understand why. How wide we should set these ranges will depend on how the target is to be used: the higher the penalties to the teacher or school for underperforming the range, the wider the range must be.

This shows that expected progress measures should never be seen as more than indicative. Yet unfortunately it seems they are treated as a science – and indeed teacher performance and even pay are being evaluated on the supposition that progress is linear. My interest is in social disadvantage, which is of course correlated with low initial attainment, and I would like to see how these trajectories differ by social class. The analysis shows that many low initial attainers go on to achieve very good education success, but most do not. This highlights the urgency of focusing policy and practice on supporting lower attainers.

Becky Francis, Professor of Education and Social Justice, King's College London

Numbers need to be treated with care. Statisticians know this. In particular they recognise that the single number summary of a disparate dataset through the production of an average score hides as much as it reveals. At a stroke, it removes from view the variation in the dataset. Yet good teaching relies on keeping in view the variation in pupils' responses to what and how teachers teach. We lose sight of this when numbers set a uniform path for every pupil to follow. It is refreshing to see the first fruits of a more sophisticated analysis here that is challenging some of the core assumptions upon which much monitoring of teachers' practice rests and is reminding the profession to expect the unexpected, not rely on simplistic mantras to keep children on track.

Gemma Moss, Professor of Education, University of Bristol

WE WORRY ABOUT TEACHERS' INFLATING RESULTS; WE SHOULD WORRY MORE ABOUT DEPRESSION OF BASELINE ASSESSMENTS

It's not uncommon to hear schools express a view that the attainment of their intake is systematically over estimated or 'inflated' in some way. Schools are not accusing another of cheating; simply that high stakes accountability pushes teachers to ensure children achieve the best result possible. And where a feeder school outperforms expectations for an entire class, those receiving the pupils in September are given a very difficult hurdle to jump over.

Is there any systematic evidence that certain teachers assess children optimistically on Key Stage tests in England? (There is plenty of evidence they do in the US). Here, we look at Key Stage One assessment, which is particularly interesting for two reasons. First, it is teacher-assessed, although this was not always the case, and so there is room for a great deal of discretion as to how Levels are assigned. Second, it takes place at age 7 and is treated as a baseline metric for judging progress in primary schools, and yet is an outcome measure for those in stand-alone infant schools.

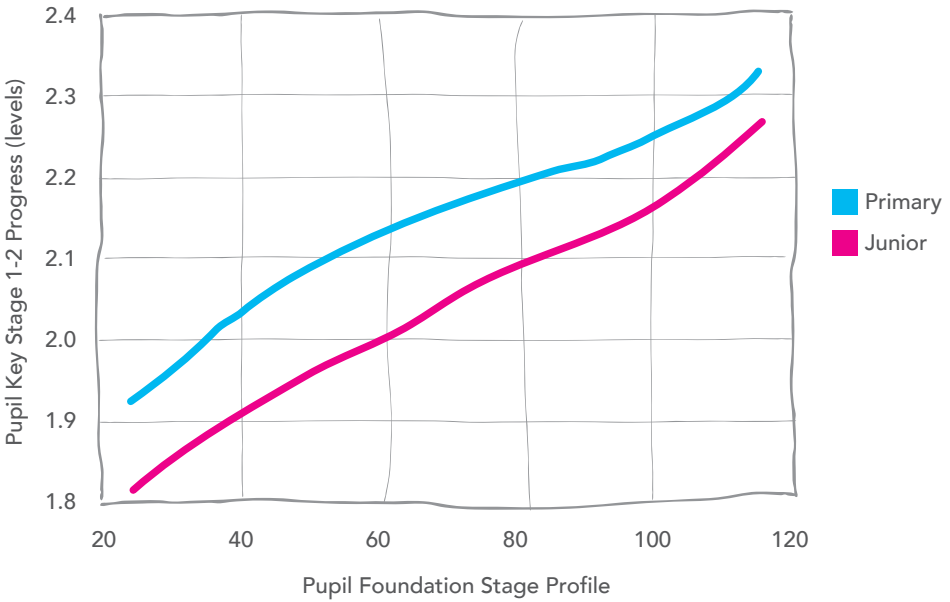
WHY IS KEY STAGE ONE ASSESSMENT SO DIFFERENT IN INFANT AND PRIMARY SCHOOLS?

Junior schools regularly claim that the Key Stage One scores assigned by infant schools are unrealistically high, making it hard for them to achieve good Key Stage One-Key Stage Two progress compared to all-through primary schools. Junior schools do make relatively poor progress at Key Stage Two, but does the blame lie with infant schools 'optimistically' inflating their pupil grades?

At first glance, junior schools appear to have a point: infant schools look suspiciously effective when judged on their Foundation Stage Profile-Key Stage One progress relative to primary schools. And the split infant-junior systems are no more or less effective than all-through primary schools when progress from age 5 to age 11 is measured.

It seems highly unlikely that infant schools are systematically effective institutions whilst junior schools are systematically ineffective institutions. So, are infant schools indeed inflating their results, or does the problem lie elsewhere?

Pupil progress in junior schools is poor



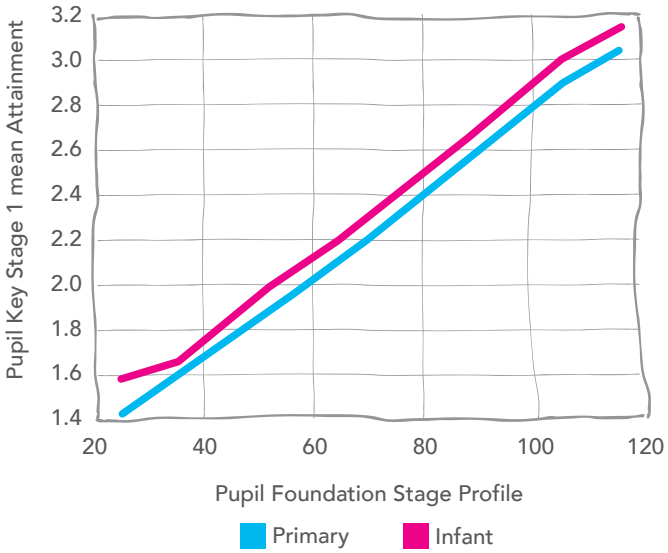
Note: we band the Foundation Stage Profile total point scores into groups of 10

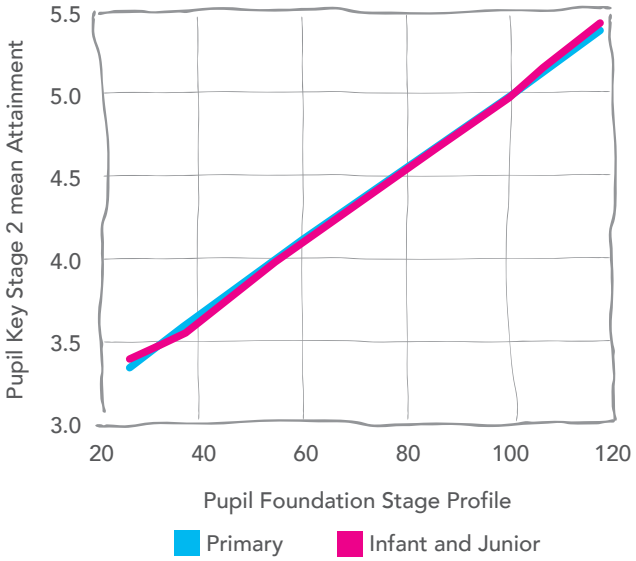
SCHOOLS RESPONDED DIFFERENTLY TO THE SWITCH FROM EXTERNAL MARKING TO TEACHER ASSESSMENT

Key Stage One assessments were externally marked tests until 2003, after which it was left to schools themselves to measure pupil attainment. Before 2003, infant schools achieved only slightly higher Key Stage One scores than primary schools, but after teacher assessment was introduced, their scores started to diverge strongly.

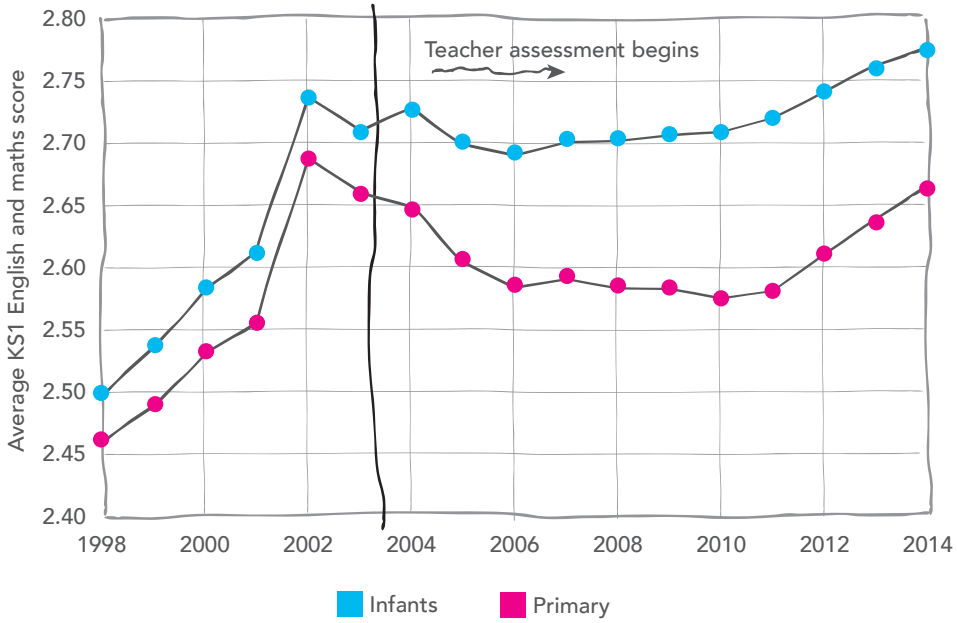
The pattern of this divergence is very clear. There is little evidence that infant schools are taking advantage of teacher assessment to inflate the scores they give pupils. Instead, teacher assessment in primary schools produces lower judgements of Key Stage One attainment, thus lowering their bar to show impressive pupil value-added at Key Stage Two.

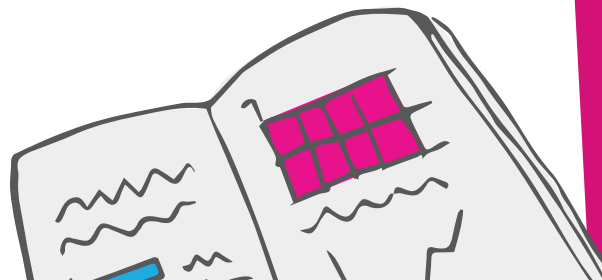
Infant schools achieve high progress, yet the split infant-junior system is no different overall





Infant and primary school Key Stage One scores really diverge after teacher assessment begins





INCENTIVES CHANGE IF INFANT SCHOOLS ARE REORGANISED INTO PRIMARY SCHOOLS

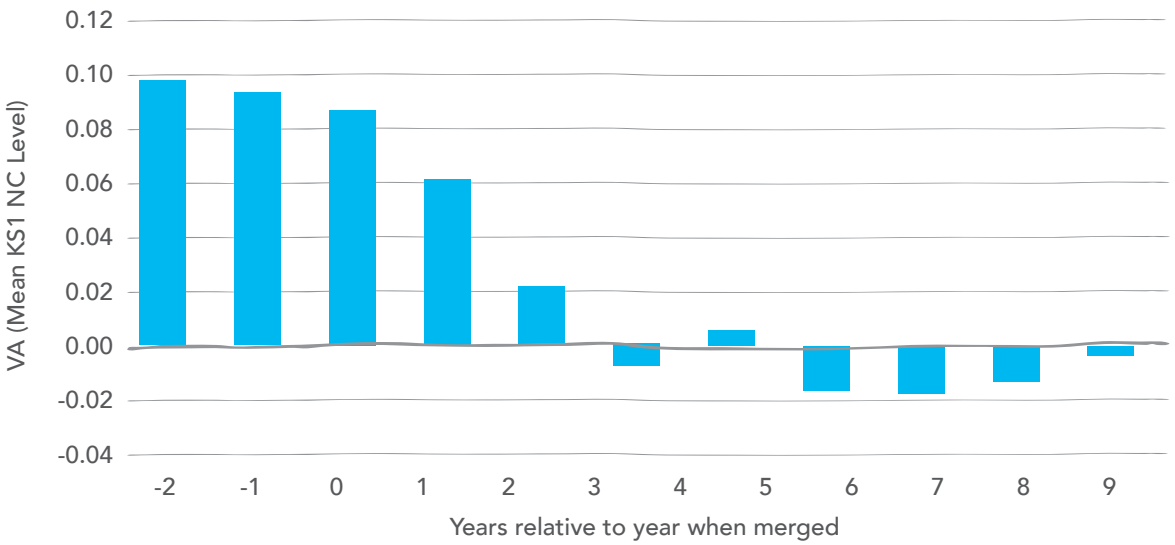
We can also use the large number of reorganisations from split infant-junior schools into all-through primary schools to see how Key Stage One assessments change over time. We see exactly the same pattern: the apparent effectiveness of infant schools, as measured by their Foundation Stage Profile-Key Stage One value-added, slips away once they become part of a primary school.

Looking across the individual Key Stage sub-levels, it is clear that this pattern is most pronounced for the assessment of high-attaining 7-year-olds. Once teachers are asked to assess pupils themselves, they become particularly cautious in assigning a Level 3. Primary school headteachers have often complained the conversion of Levels into marks by the Department for Education encourages this, since all Level 3 children are judged equal; at Key Stage One there is no such thing as a sub-Level 3C. Furthermore, the conceptualisation of what it means to be 'working at Level 3' may be different in a primary school where teachers work across Key Stages One and Two.

So, junior schools are, in part, vindicated by our analysis. It is indeed more difficult for them to achieve high Key Stage One-Key Stage Two progress, having received their Key Stage One assessments from infant schools. But perhaps they are wrong to point the finger of blame at their feeder school partners. Instead, the problem appears to lie with primary schools who depress their teacher-assessed Key Stage One baselines to achieve the best possible progress results at Key Stage Two.

This is not cheating of the sort we see in the US: teacher assessment of young children is highly subjective and faced with enormous incentives not to give a child the benefit of the doubt for an answer to a question, it is not surprising that primary schools choose to err on the side of caution.

High value-added of infant schools fades away when reorganised into primary schools



One of the quirks of the National Curriculum for England is that while levels of attainment were always intended to be independent of the age of the student, the programmes of study were age-specific. The result is a strange system in which Level 3 (say) is meant to denote the same level of achievement whether it is based on the programmes of study for Key Stage One or that for Key Stage Two. The question, therefore, is what the 'correct' interpretation of Level 3 should be. Junior schools have claimed that the assessments made by infant schools are somehow inflated, but the analysis above presents a strong argument that these claims have little merit. Indeed, if there is any distortion in the system, it is that levels given to students at the end of Key Stage One in primary schools are lower than they should be, so that it is the estimates of value-added in Key Stage Two that are inflated.

Dylan Wiliam, Emeritus Professor of Educational Assessment, UCL Institute of Education

This analysis shows how difficult it is for someone making a judgement not to be influenced by their knowledge of the use to which the judgement will be put. As assessment data is put to more uses, it becomes even more important to make sure that the context is recognised when the data is interpreted. Understanding context can also help to make sure that we have realistic expectations of assessment programmes.

Amanda Spielman, Chair of Ofqual and Education Advisor to ARK

I have been around long enough to remember the introduction of the National Curriculum and Key Stage testing in the early 90s, and the words of one trainer have stayed with me. She put some water into a glass and then put a pencil into it to measure the level – Key Stage One. She topped it up with more water and measured it again – Key Stage Two. Then she repeated the exercise – Key Stage Three. And her comment was, "But, as we all know, children leak…". Knowing this, it is understandable that teachers in primary schools feel cautious about assessing high levels of attainment at Key Stage One that then translate into higher targets for all of their teaching colleagues.

Jill Berry, Educational consultant and former headteacher

TEACHERS WITH A PHYSICS DEGREE MAY IMPROVE ENTRY RATES TO GCSE PHYSICS, BUT DON'T APPEAR TO AFFECT ATTAINMENT

We want the nation's children to be taught by teachers who are passionate experts in the subject they teach. There is widespread concern that many children are taught science and maths by teachers without an academic degree in the subject. This shortage is most acutely felt in physics, with large numbers of unfilled teacher training places, despite the offer of substantial financial incentives to train, and where a Department for Education survey reports one third of physics teachers do not have a degree in the subject.

Are we really sure that we want more graduates with a physics degree in the classroom? What types of careers should they forgo to do so and at what costs to the industries they currently serve? Central to these important policy discussions must be the demonstration that teachers with a physics degree are more effective in delivering the GCSE curriculum than those who simply have an A-Level.

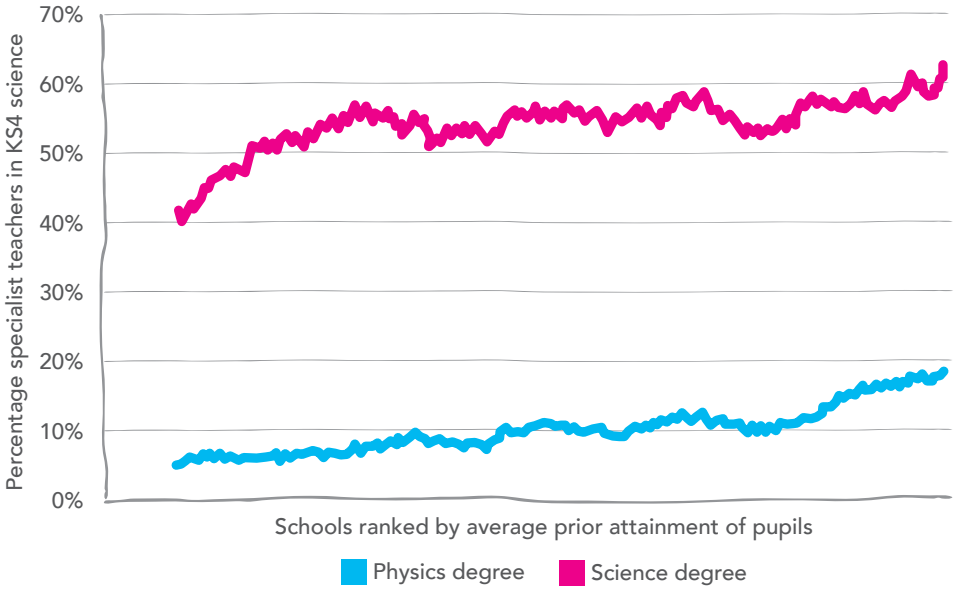
In this piece, we use the School Workforce Census to explore where teachers with a physics (or engineering) academic degree are currently teaching. Although theoretically a census of all schools, only about a third of secondary schools have completed the qualifications and curriculum parts sufficiently well for us to feel confident in using the information. In these schools we focus on the teachers who report they are teaching Key Stage Four science classes in 2013.

In this sample of 1128 secondary schools, about 45% of all GCSE science teaching time appears to be with somebody who doesn't have a science first degree (or Masters or PhD), using Department for Education mapping of degree subject to curriculum area. We find just 10% of Key Stage Four teaching time is with a teacher who has a physics or related engineering degree. For comparison, the Institute of Physics survey claims fewer than 1 in 5 science teachers have a specialist physics degree and a Department for Education survey says one third of physics teachers do not have a degree in the subject. In our sample, as many as 40% of schools are delivering their Key Stage Four curriculum without a teacher with a physics degree on the teaching team.

SCHOOLS WITH A HIGHER ATTAINMENT PUPIL INTAKE HAVE MORE PHYSICS SPECIALIST TEACHERS

We first look at the kind of schools that have greater numbers of science and physics specialists delivering the Key Stage Four science curriculum. Perhaps not surprisingly, they are more prevalent in schools with a higher entry attainment of pupils.

The location of teachers with physics and science degrees by ability of pupil intake



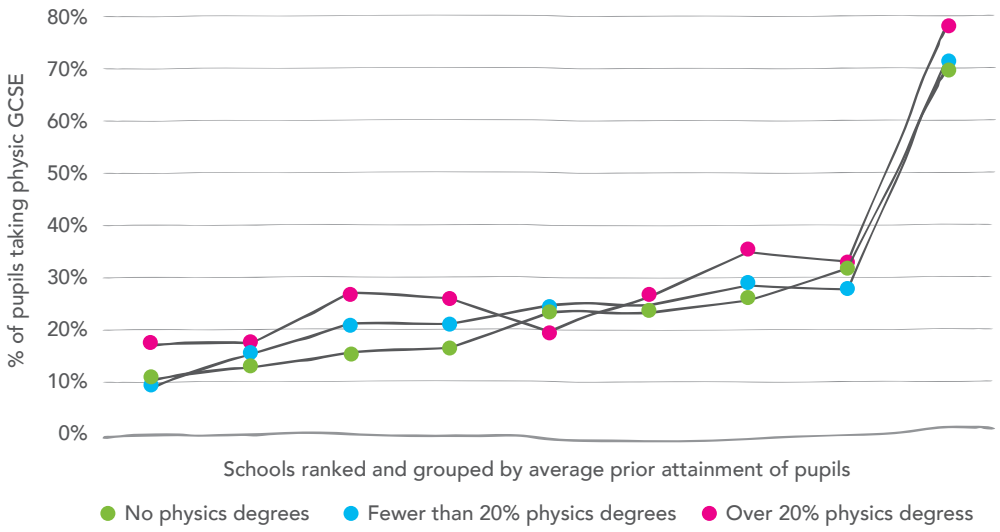
THE MORE SPECIALIST PHYSICS TEACHERS YOU HAVE, THE MORE STUDENTS TAKE GCSE PHYSICS, BUT WHICH IS THE DIRECTION OF CAUSATION HERE?

Schools will vary a great deal in how many students are encouraged to take the 'triple sciences' GCSEs of biology, chemistry and physics, rather than a double or single-award science. We explore this by grouping schools into whether they have (i) no teachers claiming to have a physics degree in their Key Stage Four teaching team, (ii) up to 20% of teaching time with a teacher who has a physics degree, or (iii) over 20% of teaching time with a teacher who has a physics degree.

There is some evidence that schools with larger numbers of physics specialists have slightly higher entry rates to GCSE Physics – two and five percentage point higher entry rates for modest and high numbers of physics specialists, respectively (holding constant entry attainment at school).

However, it is not obvious which comes first. Do specialist physics teachers encourage GCSE Physics take-up, or do high triple-science entry levels attract specialist physics teachers to apply to teach at the school and does the school need to work harder to recruit them?

Proportions taking GCSE Physics by entry attainment of school intake



THERE IS NO RELATIONSHIP BETWEEN HAVING PHYSICS SPECIALISTS AND GCSE OUTCOMES IN SCIENCE OR PHYSICS

We cannot look at the effectiveness of a department's teaching by observing their GCSE Physics grades because not everyone takes the subject. Instead, we look for any relationship between the presence of Key Stage Four teachers with a physics degree and GCSE attainment in two ways.

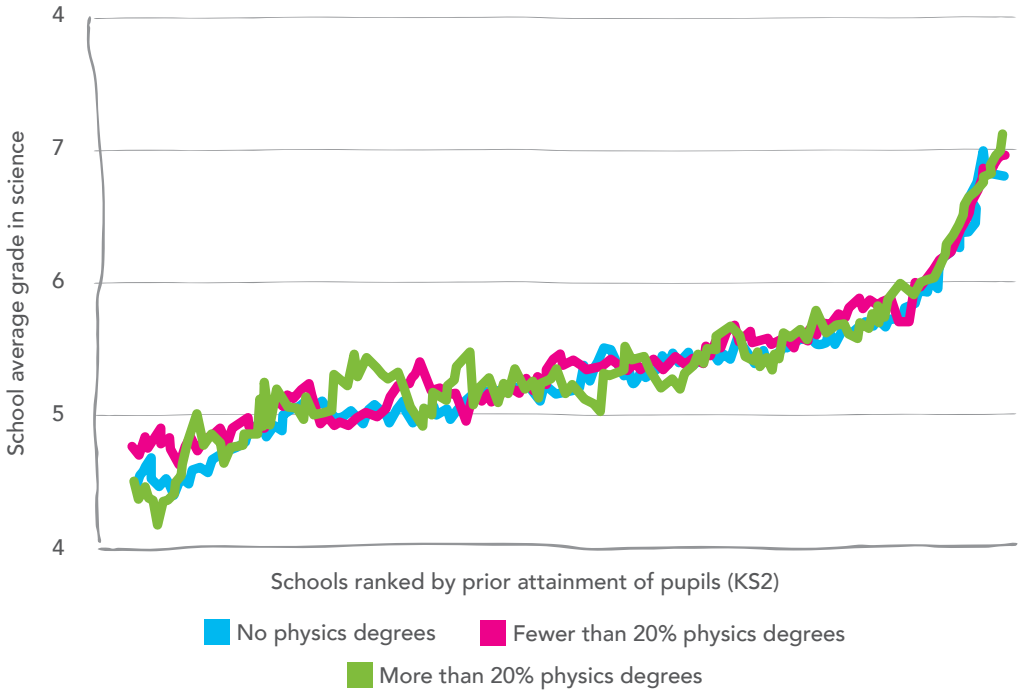
First, we look at whether the Key Stage Four teams that are well-endowed with physics specialists have a higher average point score in science, controlling for entry attainment. We find no such relationship. Obviously a pupil's science point score can be achieved through a variety of different qualifications and so it isn't an ideal measure of teaching quality or pupil mastery of physics.

Second, we use a physics contextual value-added score to assess performance, taking into account all prior attainment and pupil demographic characteristics that we can observe. It is constructed to have a mean of zero, even though more able pupils are more likely to take GCSE Physics. Once again, there is no overall relationship between school physics CVA and the number of physics specialists in the school, either on average or for any particular type of school.

Given the undeniable shortage of teachers with a physics degree delivering the science curriculum (note – in our sample, 25% of those with a physics degree report they are teaching maths), schools face the very real trade-offs about how to manage. Can they really cope with offering GCSE Physics to most of their pupils? And if they are hiring new staff, should they favour the biologist who performs outstandingly well at interview or a physicist to rebalance the science teaching team?

All our intuitions tell us that teachers with physics degrees should be better at teaching physics than those without. This is akin to saying that teachers with the greatest mastery of the subject should be the greatest teachers. Once we generalise to this level, we can draw on the wealth of evidence that suggests teacher quality appears to be largely unrelated to academic credentials. It is a surprising and little understood finding – it seems that the ability to engage and impart knowledge is quite a different skill than the ability to understand and store information yourself. With this in mind, why should physics teaching be any different?

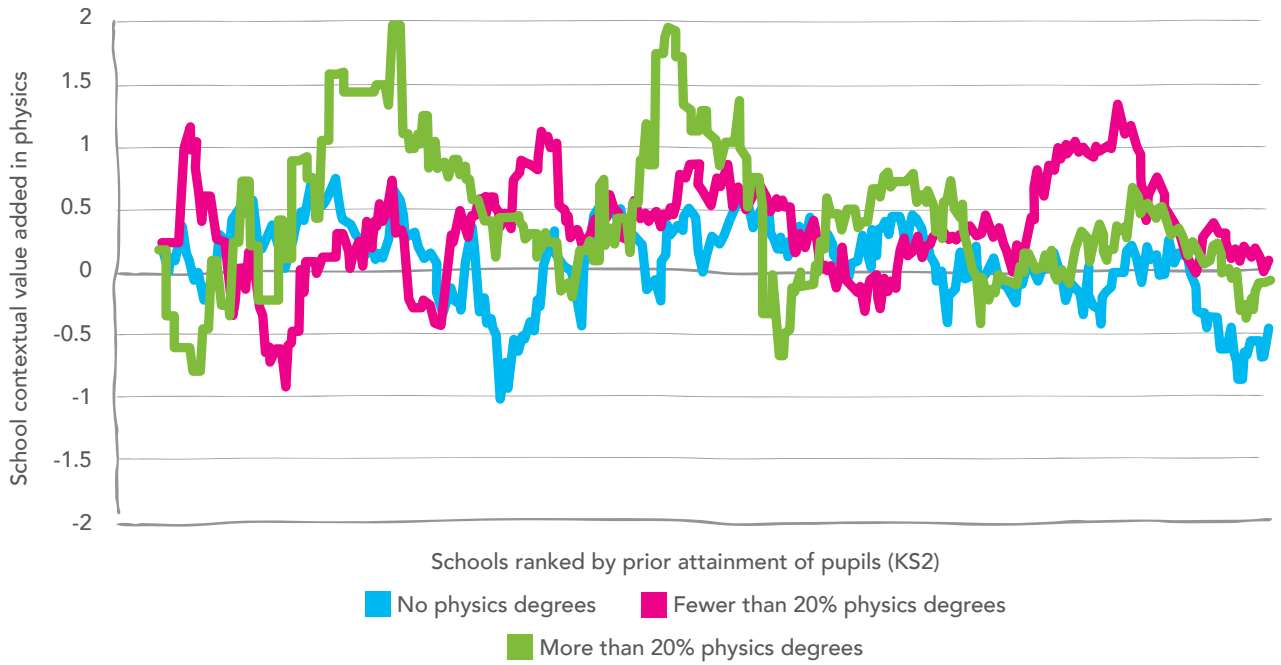
School average point score in pupil's best Science GCSE



Is high GCSE attainment what defines 'good' in physics teaching? Evidence from PISA 2006 suggests that internationally there is an inverse relationship between attainment and interest in science. In countries with the lowest science attainment (typically less developed countries like Mexico and Brazil), students show high interest in science, while in countries with the highest attainment (like Finland and Netherlands) students show low interest. There are several possible explanations for this effect, but one thing is for sure: there are many ways teachers can get high GCSE attainment, and not all of them are inspiring nor encourage to progression to A level study. For a subject like physics, where demand for qualified physicists exceeds supply, it is arguable that inspiring teaching is even more important than teaching that gets high GCSE scores.

Sir John Holman, Senior Adviser in Education, The Wellcome Trust

Physics contextual value-added score also appears to be unrelated to teaching by specialists

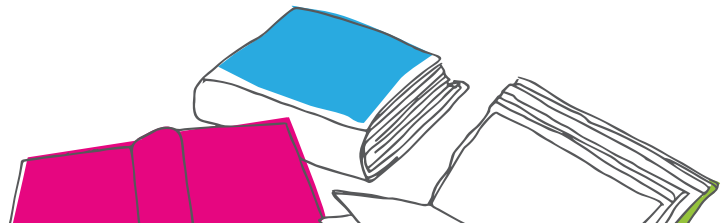


This analysis is consistent with a wealth of US research and our own findings at Teach First. We recruit a number of participants to Teach First without a degree in the subject they'll be teaching – as long as they have an A-Level in the subject and pass a knowledge assessment – and have never found any difference between these recruits and others on our quality measures (e.g. QTS score, likelihood to drop out).

It's not really that surprising. The type of content that a physics undergraduate deals with is very different from that on the GCSE syllabus. It does not mean, though, that academic ability doesn't matter in teacher selection. US research by Rockoff et al. shows that teachers' scores on cognitive tests can be modestly predictive of pupil outcomes.

Nor does it mean that subject knowledge doesn't matter. Heather Hill's research on Knowledge For Teaching shows a strong relationship between teachers' ability to explain mathematical concepts and predict students' misconceptions and those students' outcomes. Perhaps teacher training needs to focus more on translating ITT recruits' existing knowledge into 'knowledge for teaching'?

Sam Freedman, Director of Research, Evaluation and Impact, Teach First

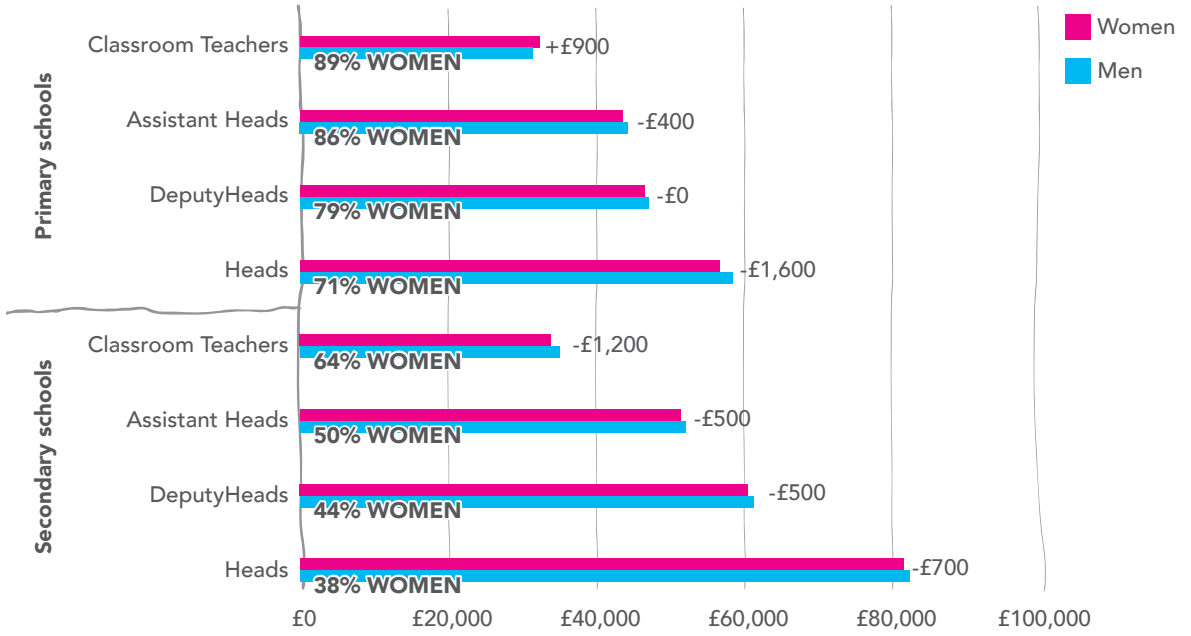


WOMEN DOMINATE THE TEACHING PROFESSION, BUT MEN ARE WINNING THE PAY GAME

It has been well reported that women fall behind men in their rates of promotion to school senior leadership positions.

The well-versed reasons for this are similar to other professions, with women taking more time than men out of the labour market when they become parents, many deliberately choosing to take lower responsibility roles and others never returning to teaching at all. However, even for those female teachers who do achieve senior leadership roles, their wages are lower than men with the same level of responsibility. In this piece we look at teachers working full-time across two years of the School Workforce Census to try to understand how these pay differences emerge.

There are average wage differences between male and female teachers



WOMEN ACHIEVE A SMALLER ANNUAL PAY RISE THAN MEN

If we compare teachers working full-time in both 2010 and 2011, women achieve a smaller annual pay rise than men at all levels of seniority. This is true when controlling for initial level of pay, age of teacher, tenure at school and region (unfortunately we do not have the teacher's total years of experience in this version of the dataset).

Differences in annual pay rises between men and women

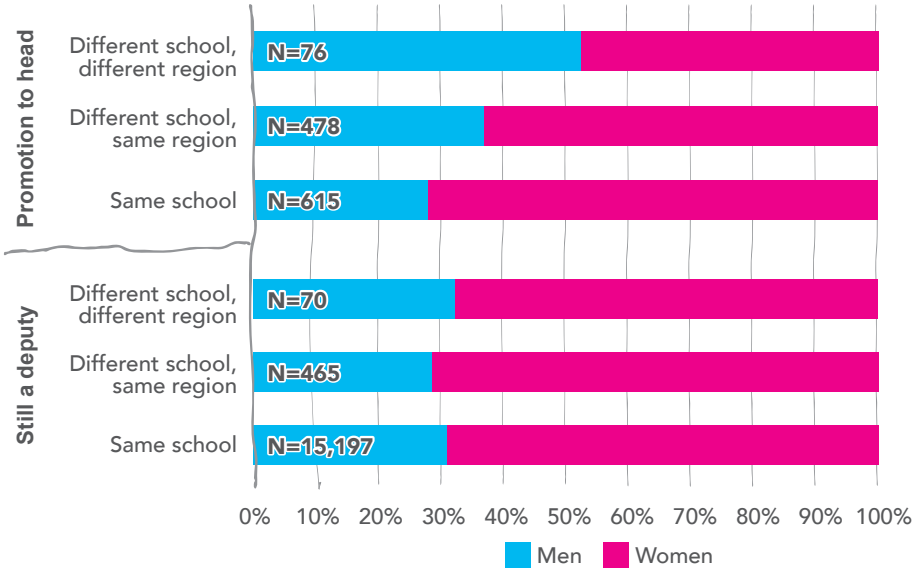
	Heads	Deputy Heads	Assistant Heads	Classroom Teachers
Men	£1,280	£1,287	£764	£934
Women	£1,122	£898	£677	£1,063
Difference	-£158	-£389	-£87[n.s.]	£129
Difference, holding constant pay level	-£145	-£386	-£153	-£167
Difference, holding constant pay level, age, tenure at school, region	-£147	-£353	-£122	-£199

CAREER MOVE DIFFERENCES ARE PARTICULARLY PRONOUNCED FOR DEPUTY HEADS

The gender differences for those in deputy head positions in 2010 is particularly large, at almost £400, and so we explore what type of career moves are taking place to explain these differences. In any one year, most deputy heads remain as deputy heads in the same school. Similar proportions of men and women make a sideways move to another school, but the proportion achieving promotion to headship looks quite different.

Women are a little more likely to achieve internal promotion to head within the same school, but are far less likely to be promoted to a different school in the same region or in another region. They may feel less confident that they are ready to seek promotion to headship or may have other life commitments that mean they feel unable to take it on. They may feel highly geographically constrained by a spouse job or childcare arrangements. Alternatively, selection panels may frequently have an unconscious bias towards a male candidate when they are previously unknown to the school.

Types of career moves for deputy heads

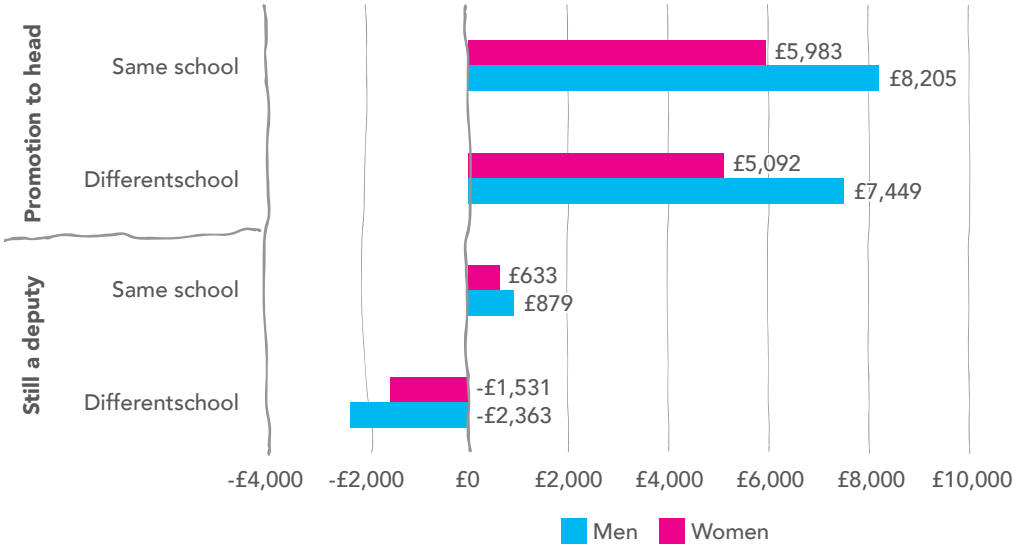


MORE WORK IS NEEDED TO UNDERSTAND THE STARK DIFFERENCES IN PAY RISES ASSOCIATED WITH PROMOTION

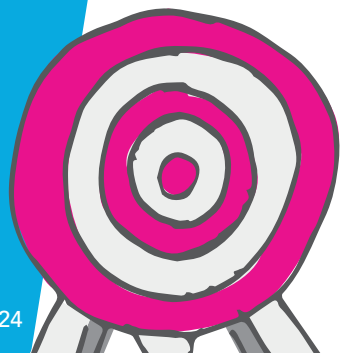
Some of these different sorts of job moves are associated with a gender difference in wage rises. Men achieve substantially greater pay rises on promotion to head than women do, and this is true whether they do so via internal or external promotion. They actually see a greater wage fall if they decide to make a sideways move to a deputy head post at a new school – these moves may be forced by household relocations and some of this wage fall is explained by loss of London weighting. But even for those remaining as a deputy head within the same school, the wage rise advantage of men remains. All these patterns hold for the teachers' initial pay, age, tenure in school and region.

The apparent wage bargaining advantage for men is much stronger in secondary schools than in primary schools. We cannot show why this is. It may be explained by greater wage variation overall for secondary headteachers or result from lower levels of guidance in wage setting from local authorities. Or it may simply be that men receive these higher wages in return for the types of roles they take on, whether they be more complex schools or risky headships of previously underperforming schools.

Annual pay rise for deputy heads by career move



Note: all differences are statistically significant



This depressing analysis mirrors the picture we have seen among the almost 100 participants on the Future Leaders programme who have so far secured headship. Female Future Leaders heads are far more likely than their male counterparts to be in an interim or acting role – which is more likely to happen in an internal appointment. It is almost as if women have to prove themselves to governors (and sometimes to themselves) in a way that isn't expected of men before being able to take on the role substantively. And those who do look elsewhere for headships can face discrimination from governing bodies looking to recruit the "right man" for the job, evidenced both in anecdotal feedback and figures suggesting our female Future Leaders make more applications on average than male Future Leaders before securing their first headship.

Kate Chhatwal, Chief Programme Officer, Future Leaders

Both career breaks and a lack of mobility may affect the pay of women leaders in schools, leaving them earning less than men.

Prof John Howson, Chairman, Oxford Teacher Services Limited

By looking at average pay rises rather than just average salaries, this analysis suggests that lower pay for women in senior roles is not caused solely by slower career progression (due to family commitments for example). Female leaders seem to be offered lower pay rises for the same roles. It is not a great finding for a profession that can claim with some pride to have one of the largest densities of female chief executives in the country.

A more professionalised recruitment and pay-setting process for senior roles might help; this is an especially urgent need now that significant discretion on pay has been delegated to governing bodies.

Russell Hobby, General Secretary, National Association of Head Teachers

THE 'LUCKY' CHILDREN WHO JUST GET INTO GRAMMAR SCHOOLS DON'T APPEAR TO ACHIEVE MORE THAN THEIR PRIMARY SCHOOL CONTEMPORARIES WHO JUST MISS OUT

It is very hard to say anything about the impact of grammar schools. There is no such thing as selective and non-selective areas: 1 in 5 grammar school students cross over a local authority border on the way to school. And perhaps as many as 1 in 5 grammar school students were not in a state school at age 10. We can, however, watch what happens to children who attend the same state primary in an unambiguously selective area. It is likely that most able children will sit the 11+ – some will pass and proceed to a grammar school – others will end up in non-selective state or private schools.

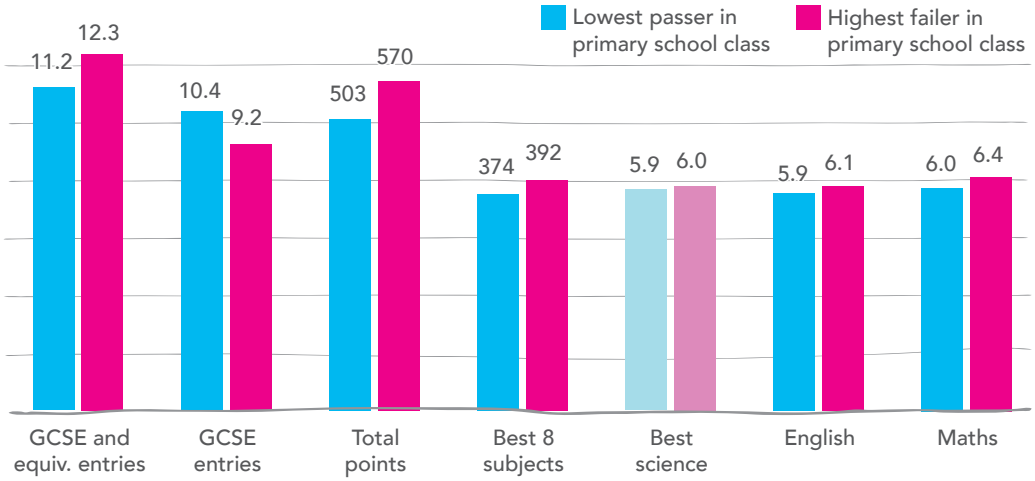
Here, we look at children in just over 500 primary schools in grammar school areas that routinely see their pupils pass the 11+. Children who have come from the same primary school will have access to a similar set of grammar schools with similar pass rates, similar practice for 11+ tests within the school, similar classroom experiences and so only differ in the amount of parental support for their learning and private tutoring they receive.

We don't have these children's 11+ scores (which in any case are skewed by test preparation investment and missing for non-takers), but we do know how they did in their Key Stage Two tests just before leaving primary school. We look in these primary schools and find the child who scored highest in their Key Stage Two tests and yet did not attend a grammar school, calling them our 'highest failer'. (We exclude children who go onto private schools since we do not know whether their decision to do so was influenced by their 11+ performance). We then find the child who scored lowest in their Key Stage Two tests and who went on to attend a grammar school, calling them our 'lowest passer'. It is worth noting that our highest failers are twice as likely to be eligible for free school meals as the lowest passers (6% vs 3%).

WHAT HAPPENS TO OUR 'HIGHEST FAILER' AND 'LOWEST PASSER' BY THE AGE OF 16?

The child scoring highest at Key Stage Two who goes onto a non-selective school outperforms their primary school peer who 'just' passes their 11+, i.e. scores the lowest Key Stage Two mark of all those in the primary school going onto a grammar. They outperform in both broad measures of attainment – total point score and best eight subjects – and in GCSE English and maths (though not in science). The 'highest failer' takes more qualifications at 16 than the 'lowest passer', though fewer GCSE examinations.

The child who 'just' gets into grammar school does worse at GCSE than their primary school peer at a state non-selective school

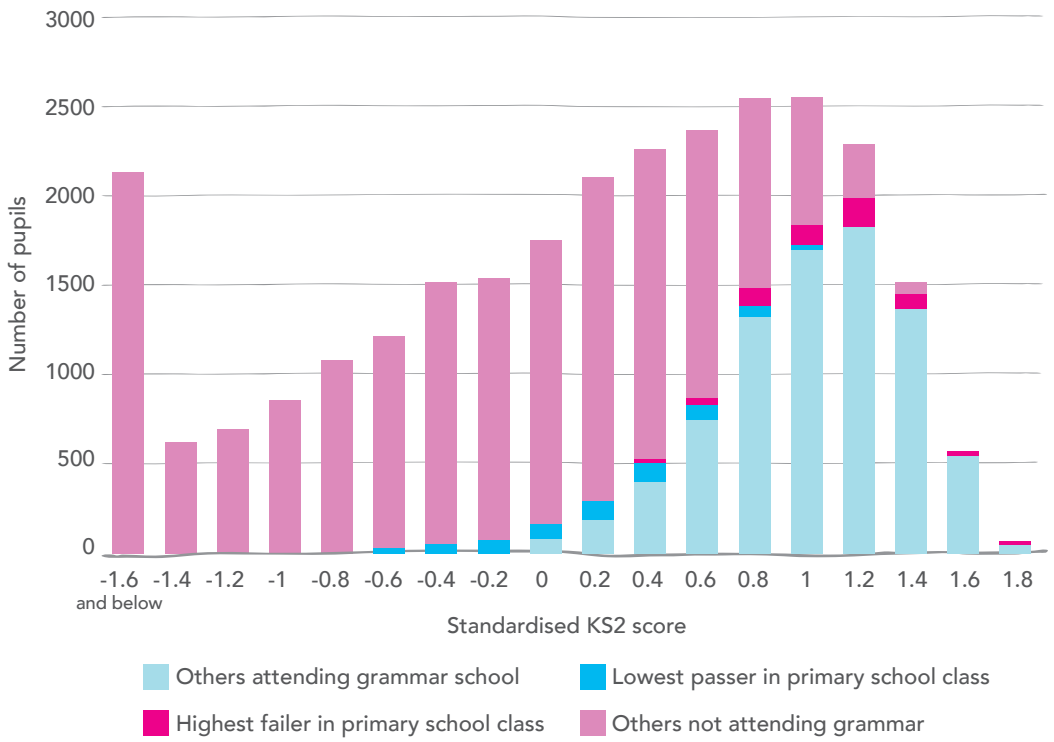


HIGH KEY STAGE TWO ATTAINING CHILDREN SOMETIMES FAIL THE 11+ WHILE LOW KEY STAGE TWO ATTAINING CHILDREN SOMETIMES PASS

Should we take this to mean that secondary modern and grammar schools are equally well-suited for high-attaining pupils? Or that 11+ selection isn't a problem because everyone does well, regardless of destination? No. Even though the 'worst passer' must have passed an 11+ test, their average Key Stage Two scores are lower than the 'highest failer'. It seems hard to believe, but there are children at grammar schools who only achieved a Level 3 at English or maths.

Equally, some children at non-selective state schools in grammar school areas have Key Stage Two scores equal to the smartest pupils at grammar schools. We don't know why the 11+ is a poor match for Key Stage Two attainment and the local authorities where the discrepancy is particularly high have nothing obvious in common.

Somehow children with very low Key Stage Two scores manage to pass the 11+



PASSING THE 11+ LEADS TO A VERY DIFFERENT SECONDARY SCHOOL EXPERIENCE

How lucky are these 'lowest passers' for passing the 11+? We already know they don't do particularly well at GCSE, compared to their 'highest failers' from their primary school class, so it isn't possible to argue that the 11+ is a better indicator of academic potential than the Key Stage Two tests. We can also show that they tend to find themselves amongst the weakest performers academically at the grammar school they attend. More worryingly, many of them will be labelled as having Special Educational Needs (School Action) by their grammar school, a status commonly given to children on the basis of emotional and behavioural difficulties.

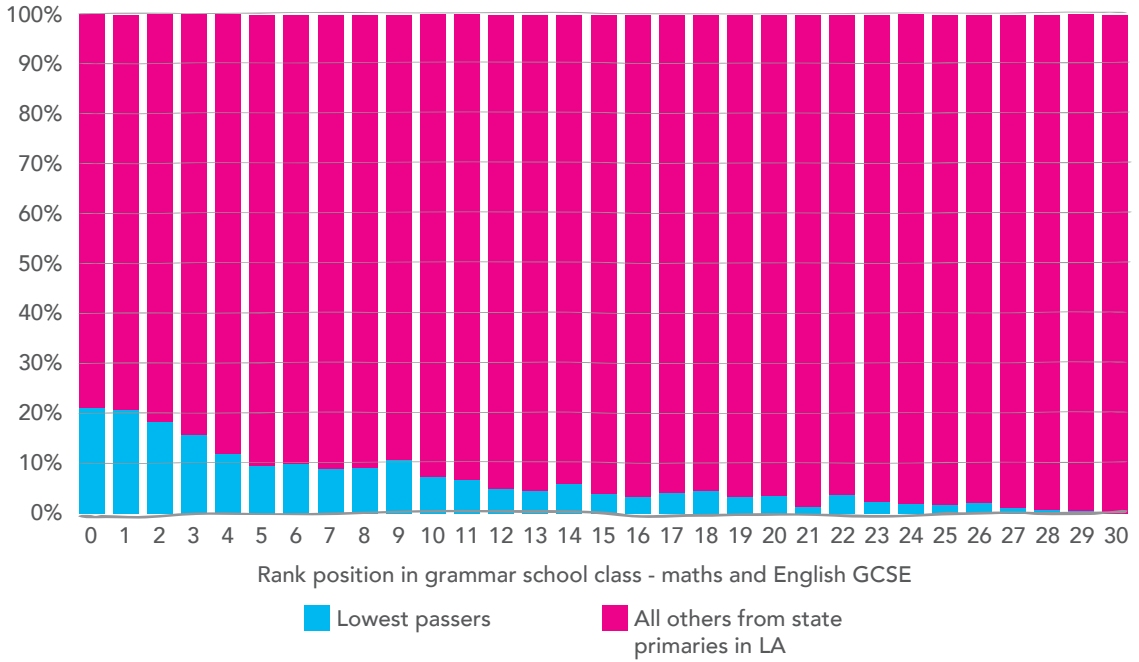
At primary school, about 4% of our 'highest failers' and 'lowest passers' have the SEN School Action status; by age 16, about 10% of the 'lowest passers' at grammar school will have this status compared to 4% of the 'highest failers' at a non-selective state school.

How might our 'highest failers' have performed had they had the opportunity to go to a grammar school? Rather than compare them to the 'lowest passer', instead we find the child in their primary school class with who they most closely match on Key Stage Two scores but who attends a grammar. With this matched comparison we see that the grammar school attendee outperforms by about half a GCSE grade in their core subjects. They also appear to take a more traditional curriculum, with a greater number of GCSE subjects and fewer GCSE equivalent qualifications.

We cannot be sure these advantages are entirely down to the grammar school. Although they attended the same primary school and had the same Key Stage Two score, the very fact they passed the 11+ may not have been down to chance. They may have had greater academic potential that we cannot identify in Key Stage Two tests, or more likely came from the kind of family who helped them prepare for the 11+ exam and has equally supported their educational progress since.

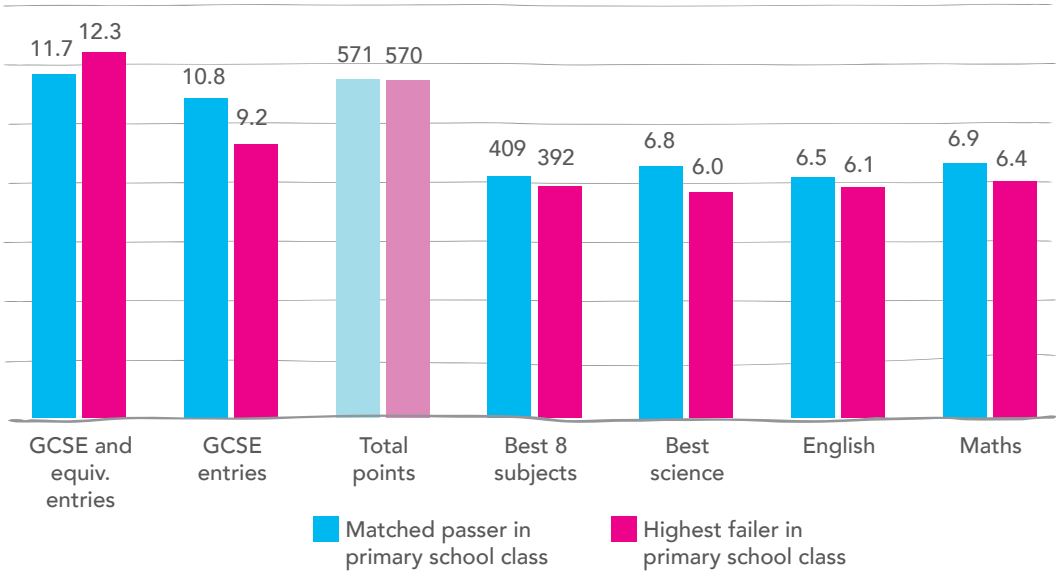
This analysis doesn't say anything about whether selective schooling systems are better for the average child, the low-attaining child, the high-attaining child, or anyone else. It simply points out the shortcomings of the 11+ exam across every highly selective local authority, regardless of whether they use verbal reasoning, non-verbal reasoning, numeracy, English or any other test paper.

The 'lowest passers' tend to find themselves struggling against their new grammar school peers



There are children who routinely manage to pass the 11+ exam ahead of primary school peers who both score higher at Key Stage Two tests and must have greater academic potential since they go on to achieve more at GCSE. This means the 11+ exam is frequently less successful than Key Stage Two tests in selecting the highest potential children from primary schools. Differences in the social background of these pupils who 'just pass' compared to their brighter peers who fail suggest this is not simply due to chance.

The grammar school pupil outperforms a pupil from the same primary school with the same Key Stage Two score who did not attend



This interesting analysis highlights the real challenges posed by trying to measure children's academic achievement at age 11, let alone their potential. For some children, it appears that the 11+ tests that are used to gain entry into a grammar school do not measure their academic potential as well as Key Stage Two test scores. As a consequence of these measurement problems, undoubtedly many high-ability children miss out on a place at a grammar school. It remains unclear from this analysis however, whether this really matters for their future academic success.

The other issue this work throws up is the problem of the 'big fish little pond' effect (Marsh, 1987). This stems from the finding that lower-achieving children in a group of higher-achieving pupils will tend to have lower academic self-esteem, with negative consequences for their subsequent achievement. Children who perhaps are wrongly identified as high-achieving according to the 11+ test (due to the measurement error problem described above) are likely to suffer from the 'big fish little pond' effect in a grammar school full of higher achievers. They are then likely to have an even lower achievement as a result.

In a nutshell, measuring achievement at an early age is difficult and our selection systems need to reflect this.

Professor Anna Vignoles, Professor of Education (1938), University of Cambridge

education datalab

research • analysis • evidence

Education Datalab
19th Floor, Portland House,
Bressenden Place, London SW1E 5RS

e: educationdatalab@fft.org.uk
t: 0203 761 6959