England: England and PISA—The Long View

Tim Oates

Abstract This chapter does not drill down in the minutiae of the PISA results for England. For that, readers can go to the NFER's excellent report (Sizmur in *Achievement of 15-year-olds in England: PISA 2018 Results DFE-RR961* 2019) which comprises the UK Government's commentary on the PISA outcomes. Rather, it tries to do something unique—it places the PISA results in the context of policy changes which may be associated with PISA outcomes, and seeks to explain the factors which determine the trends present in the PISA data. It looks briefly at the other administrations of the UK (Scotland and Wales in particular), but highlights the vital differences between those administrations. I maintain that 'The UK' cannot be treated as a unitary system.

1 Introduction

This chapter does not examine methodological issues associated with the PISA data for England. For this, I refer readers to the work of John Jerrim (particularly his 2011 paper on the limitations of both TIMSS and PISA), Benton and Carroll (2018) and Micklewright and Schnepf (2006).

The time series 2000–2018 is hampered by two issues. Firstly, possible mode effects following the switch to on-screen administration (Jerrim et al 2018) and secondly, the failure of England in 2000 and 2003 to meet the full sample criteria. The 2000 data are regarded as problematic, however the 2003 data are available in Micklewright and Schnepf 2006, are considered by them to be adequate, and are included in Table 1 here.

The curriculum- and instructional-sensitivity of the PISA items has been compared with TIMSS (Schmidt 2018), and I make here no assumption that PISA is

T. Oates (✉)

Director Assessment, Research and Development, CBE, Cambridge Assessment, Cambridge, England

e-mail: oates.t@cambridgeassessment.org.uk

© The Author(s) 2021

N. Crato (ed.), *Improving a Country's Education*,
https://doi.org/10.1007/978-3-030-59031-4_4

Table 1 PISA scores 2000–2018 for England

	2000	2003	2006	2009	2012	2015	2018
Maths	531	507*	495	493	495	493	504
Reading	526	507*	496	495	500	500	505
Science	535	516	516	515	516	512	507

Source OECD *England failed to meet the 85% participation school sample criterion in 2003, but met the 80% student response criterion

an entirely unproblematic or infallible measure of underlying improvement or deterioration of educational standards in England. But what I look at here is a set of correspondences—between changes—or not—in PISA scores, and policy actions. When timelines are aligned with sensitivity, and plausible time lags taken into account, some interesting possible relationships emerge. It is these on which this chapter focuses.

2 The National Scene

It's November 2019 in England, and some of the educational journalists are getting restless. They are beginning to make enquiries about the expected PISA results. In December, immediately after publication of the results, their stories will enjoy a brief flurry of prominence and they know they will swiftly move onto other things. In England, the PISA results subsequently will be cited in scattered articles which talk about educational performance and government policy (for example see *Schools Weeks* 2019; TES 2019a). Various politicized organisations will issue immediate forthright comment, seldom agreeing (Financial Times 2016). Domestic researchers will of course then begin their wide-ranging scrutiny of the outcomes, but their reflections and insights will take time, and will most likely only be seen in academic journals.

It's not that PISA in England is treated as a trivial matter. John Jerrim continues to provide influential methodological critique on both survey approach and interpretation (Jerrim 2011). The National Foundation for Educational Research provides important comparisons of PISA, PIRLS and TIMSS (e.g. NFER 2018). The Department for Education's international unit provides thorough and reflective time series perspectives. And Cambridge Assessment's researchers and policy advisers—Tom Benton in the foreground (e.g. Carroll and Benton 2018)—review method and carefully weigh up the OECD PISA reports against their own extensive transnational comparisons of education system performance around the world.

But for English policy makers, while important for international benchmarking, PISA is by no means the most important body of data for domestic discussions of educational performance. There are a number of reasons for their views, which now I will explore.

PISA has the known limitation of being a cross-sectional analysis of 15-year olds, lacking the quasi-longitudinal structure of TIMSS—‘quasi-longitudinal’ since on a four-year survey cycle TIMSS tests a sample from year 4 and year 8, although the pupils in the year 8 sample are not the exact individuals tested in the previous cycle’s year 4 sample. This presents specific limitations on inference from the PISA data. Timeframes of system reform have to be taken carefully into account, a heavily contested issue which I will deal with in more detail later in this chapter. Even seasoned analysts can forget to think about the total experience 0–15 of those 15-year olds—what was their experience in life, and what were the characteristics of the education system as they passed through it? What exactly did the 10 + years of school experience prior to the assessment point in PISA at age 15 contain? Is what is happening at 15 consistent with earlier education and experiences? This provides a different take on the unsound conclusions which can float around PISA. The following provides a telling example.

An April 2018 headline declared *‘Exclusive: England held back by rote-learning, warns PISA boss—England’s schools system is losing ground to the Far East because of an emphasis on rote-learning and a narrowing of the curriculum, says the official behind the Pisa international education rankings’* (TES 2018a). The article emphasises the 2015 PISA finding that “...Britain comes out right on top...” in terms of the amount of rote-learning in its schools. But this highlights starkly the limitation of inference from a survey of 15-year-olds—quite the wrong conclusions about the system as a whole can be assumed. These pupils are close to their GCSE examinations, taken at age 16. In this, England is typical: most systems have high stakes assessments at 16 (Elliott et al. 2015). England is atypical in using externally set examinations for these. The GCSE examination results are high stakes for schools as well as pupils—schools are measured on the grades obtained, and the ‘value added’ which each school’s programme presents. Particular subjects are counted towards a target—the English Baccalaureate ‘basket’ of GCSE qualifications grades. Rote learning as a feature of education in this exam-focussed phase is widely acknowledged in England (Bradbury undated; Mansell 2007) but its origins are complex. The 2010 Coalition Government introduced new GCSEs in English and Mathematics—first teaching from September 2015. In sharp contrast to the older qualifications—which included large amounts of coursework—the new GCSEs require extensive memorization—for example of poetry and segments of drama. Whilst the 2015 PISA cohort were not on programmes directed to these new qualifications, there was widespread discussion permeating schools about the dramatic rise in memorization required. Staff had seen new sample assessment materials and were busily preparing new learning programmes of high demand. Fifteen-year olds measured in PISA 2018 were highly sensitized to the reformed qualifications about to be taught in the system. Understandably, memorisation was a preoccupation of these students, and their teachers, as they approached their public examinations. But this should not be generalized to all pupils of all ages; it speaks of the reality for 15-year-olds.

There is further fundamental background to the PISA finding. Rote learning is seen as an essential component of learning in some Asian systems (Cheong and Kam 1992; Crehan 2016)—not as an end in itself, but in enabling knowledge to be

retained in long term memory and therefore immediately available for higher-level and complex problem-solving (Christodoulou 2014; Au and Entwistle 1999). This approach is endorsed by contemporary cognitive science (Abadzi 2014; Kirschner et al. 2006). However, it was an approach which was strongly discouraged in primary schools in England as a consequence of the recommendations of the 1967 Plowden Report.

Concerned to improve reading and maths attainment, the 2010 Coalition Government re-emphasised the importance of rote learning, particularly in respect of elements of mathematics—and particularly in respect of multiplication tables (BBC 2018). The action subsequently taken in the revised National Curriculum of Sept 2014 to strategically re-introduce rote learning into primary education was irrelevant to the 15-year-olds in the 2015 PISA survey. But this highlights an important *absence* of rote learning from the 2015 PISA cohort's primary education. With a general absence of rote learning of the form which supports higher level functioning (Sammons et al. 2008), it is unsurprising that the PISA cohort's perception is one of a highly pressured year immediately prior to high stakes examinations, characterized by subject content which need to be memorized. An appropriate reading of the PISA findings is the exact reverse of the headline '...a system dominated by memorization...'. That is, the findings should not be read as an indication of the prevalence of rote learning throughout the compulsory school system in England, but as a sign of its general neglect and absence—and an urgent preoccupation with it as pupils approach demanding national assessments.

This illustrates the extent to which extreme care needs to be taken in the interpretation of the outcomes of a cross-sectional survey of 15-year-olds. Historical context, time lags, domestic analysis all need to be taken into account.

Whilst aware of the limitations, and whilst sceptical of some of the top line conclusions from PISA reporting, policy makers and politicians in England certainly maintain a keen interest in the underlying measurements of mathematics, reading and science—and certainly the trend data. But they view PISA data as an important *part* of the picture. A part, not the whole—and whilst important for a short time around publication, in England the PISA data quickly become of secondary importance. Why? Not because of PIRLS or TIMSS—although they too are of course of interest when the reporting from them begins. No, for understanding domestic performance, PISA is of secondary interest because of the quality and comprehensiveness of England's National Pupil Database (NPD).

While exact equivalents of the PISA data on classroom climate and social/familial background contextual data is not collected in the NPD, the NPD includes essential school and pupil characteristics (birthdate, gender, school location, etc.) and attainment data (phonics check, national tests, national qualifications) for every child and every educational setting. It is massive, comprehensive, underpinned by law, of high quality, and well-curated (Jay et al. 2019). The NPD supports vital and wide-ranging analysis of equity and attainment throughout England. It is now linked to educational databases for further and higher education, and to data from the labour market. Scrutiny of these data allow sensitive analyses of the distribution of attainment within pupil groups and across geographical areas, the performance of schools,

right through to analysis of the comparability of qualifications and standards in qualifications over time. It is these massive domestic datasets which are at the forefront of policymakers', politicians' and researchers' enduring interest (Jay et al. op cit; TES 2018b). Sitting in the sidelines there also are the domestic cohort studies: the 1958 National Child Development Survey (NCDS)—all children born in the first week of February 1958- the British Cohort Survey 1970 (BCS), and the Millennium Cohort Survey 2000 (MCS). These are genuine longitudinal surveys, following life outcomes in health, education, employment—allowing extraordinary insight into the role of education in society, and society's impact on education.

3 PISA 2018 in 2019

So...PISA is interesting, but not sole reference point for English commentators and analysts. Let's go back to November 2019 and the growing anticipation in the countdown to the publication of the 2018 PISA results. What was anticipated...and why? Since the election of a coalition government in 2010, there have been genuine structural changes in provision and shifts in aims.

Principal amongst these have been:

- (1) From 2010, massive shifts of schools from local authority (municipality) control to direct contractual relations with central government. This 'academy' policy originated in the 1997–2010 Labour administrations, with the first 'academies'—essentially a change in governance—appearing in 2001. The policy was extended to potentially include all schools from 2010. 'Free schools' also were introduced as a new category of school from 2010; schools suggested and constituted by parents and community groups, approved by the State (Alexiadou, Dovemar & Erixon-Arreman 2016). Again, these are in direct contractual relationship with central government.
- (2) In Sept 2014, a new National Curriculum. A new curriculum for secondary was designed in 2007, but was rejected by the 2010 Government as poorly theorised and lacking international benchmarking. By contrast, the 2014 revisions emphasised clear statement of demanding content—concepts, principles, fundamental operations and core knowledge—with this content strongly benchmarked to high performing jurisdictions. It should be noted that the National Curriculum is not a strict legal requirement in all schools in England. Independent schools (private schools) are not under legal obligation, although their performance tends to be judged by their outcomes in national examinations at 16 and 18. Other classes of schools are required to participate in national testing and national examinations at 16 and 18, but the 72% of secondary schools and the 27% of primary schools which are academies are not required to follow the programmes of study of the National Curriculum (NAO 2018).
- (3) From 2010, a strong emphasis on improving reading in primary schools. While the previous Labour governments 1997–2010 had put in place the Literacy

and Numeracy strategies, only in the closing years of the Strategies had there been a move from diverse approaches to reading towards more evidence-based methods (Chew 2018). This followed Andrew Adonis' commission for the Rose Report on early reading (2006). From 2001 to 2006, outcomes fell in PIRLS—the Progress in International Literacy Study—which tests year 5 pupils. Scores fell from 553 to 539. They then rose between 2006–2011, climbing back to their 2001 level (score 552 in 2006) and continuing to improve for the 2016 survey (score 559)—with the 2016 figures representing a substantial closing of the gender gap—previously, in 2011, England had possessed one of the largest gender gaps in PIRLS.

Although synthetic phonics increasingly had been the focus of the last years of the Literacy Strategy, the lack of professional consensus around methods was clear in the vigorous and adverse reaction to the incoming 2010 Coalition Government's emphasis on phonics. This was considered to be highly controversial and was widely discussed in the media. Many then-prominent educational commentators were critical of this strong emphasis and argued that specific techniques should be decided upon by schools (Guardian 04 03 14; Clark 2018). Government was not deterred, and asserted phonics-based reading schemes through textbook approval procedures and, from 2012, a statutory 'phonics screening check' was introduced for Year 1 pupils (a 40-item test with a threshold score of 32). Notably, highest attainers in the 'phonics check' also were high performers in the 2016 PIRLS survey (McGrane et al. 2017). The Government's commitment to phonics subsequently was justified by both a new community of researchers (Willingham 2017; Machin et al. 2016) as well as by the continued improvement in PIRLS results for year 5 pupils. The phonics check for year 1 children also saw escalating scores—from 31.8% reaching the threshold score in the trial test of 2010, climbing to 81% in 2017. Reading attainment improved; the gender gap was significantly reduced.

- (4) From 2010, alongside the focus on enhancing reading, maths education was enhanced through a series of measures. Targeted funding was allocated to the National Centre for Excellence in the Teaching of Mathematics (NCETM, founded in 2006 by non-government bodies). Under specific contract to Government, NCETM has managed a wide ranging programme of support to schools through designated 'Maths Hub' schools (37 in total, working with over 50% of schools in the country, with over 360,000 registrations for information from the Centre), managing a research-based teacher-exchange with Shanghai, providing teaching resources and professional development, and participating in Government textbook approval processes. This emphasis on professional development and high-quality textbooks is a highly distinctive feature of the policy work on maths.
- (5) In 2015, revised GCSEs (the key examinations at the end of lower secondary, typically taken at age 16) in Maths, English and English Literature were introduced, following national debate about declining assessment standards in key examinations (Cambridge Assessment 2010). Teaching started in September

2015, with first examinations in 2017. These dates are important for interpretation of the 2018 PISA survey. New GCSEs in all remaining subjects (sciences, geography, languages etc.) were introduced in 2016, with first examinations in 2018. Significant changes were made to content, assessment methods, grading, and overall level of demand. Increasing ‘content standards’ was fundamental to the reform, with a recognition that a significant increase in demand was required to ensure alignment with high performing jurisdictions (Guardian 2013a). At the same time, a new grading scale was introduced (Ofqual 2018), moving from the old system of ‘A*-G’ (with A* being highest) to 9–1 (with 9 being highest). Great attention was paid to managing, through statistical means, the relationship between grading standards in the old GCSEs and the revised GCSEs (Ofqual 2017).

- (6) Revised targets: with the introduction in 1988 of the National Curriculum and allied national testing, there emerged policy options regarding the use of the data from those tests and from the existing national examinations at age 16 (GCSE) and 18 (GCE A Level). A national survey-based analysis of standards had been operated in previous decades by the Assessment of Performance Unit (Newton 2008) but with the introduction of national testing for every pupil, government sensed that greater accountability for each school could be introduced into the system, rather than simply the production of policy intelligence on overall national standards. The publication of individual school results, and the strong use of assessment data in school inspection became a feature of the system from 1992. For over a decade, the published data focussed on simple measures such as the percentage of pupils achieving specific ‘levels’ in national assessments in each primary school (with Level 4 being the ‘target’ level in Maths and English) and in each secondary school achieving 5 GCSEs at grades A*-C. More elaborated measures were added in 2002 (‘value-added’) (Leckie and Goldstein 2016). The evolution of school league tables in England occurred during the period 1992–2016: ‘contextual value-added’, ‘expected progress’ and ‘progress 8’), modified in 2006 (‘contextual value-added’) and 2011 (‘expected progress’) and significantly remodelled in 2016 (‘Progress 8’). The early GCSE targets had driven damaging behaviour in schools, with many schools focusing on grade C/D borderline pupils, to the detriment of both higher attaining and lower attaining pupils—and prior to 2010, successive governments had persisted with these crude targets despite clear research regarding adverse impact (Oates 2014). Later refinements, particularly Progress 8, are a committed effort to improve the validity of the representation of the performance of individual schools (principally, accounting for intake), and to drive desirable behaviours in schools. As a response to ‘gaming’ previous cruder GCSE qualification targets and measures, the Government in 2010 introduced the ‘English Baccalaureate’ (EB) performance measure—designed to ‘increase the take up of ‘core’ academic qualifications which best equip a pupil for progression to further study and work’ (House of Commons Library 2019). Typically pupils take 10 GCSE examinations. The EB measure requires pupils to achieve specific grades in English Literature and Language, Mathematics, History or Geography, two Science and a Language.

- This allows a degree of choice in the 3 or more GCSEs which pupils take in addition to the EB requirement. The national target was to reach 90% of pupils reaching the EB target by 2020, but in 2017 a new target was set at 75% by 2022. It is not a legal requirement, but it is viewed as high stakes by schools.
- (7) Changes in teacher supply: contrary to popular press stories around PISA 2000, Finnish teachers are not the most respected in the world, with the ranking from NIESR data giving Brazil 1 (lowest) Finland 38, England 47, and China 100 (highest) (Dolton and She 2018). Likewise, while starting salaries are higher in Finland, there is far greater post-qualification growth in salaries in England (OECD 2003). Yet in England teaching generally is portrayed as an undesirable occupation, with commentary focusing on (i) the high level of direct government scrutiny of schools via accountability arrangements and (ii) a high level of non-teaching activities which detract from quality of professional and personal life (NFER 2019).

Government rightly has seen the problem as twofold—a problem of supply (both in the nature of training and in the quantity of supply) and a problem of retention (content of the professional practice). Principally, in respect of supply, from 2010 the focus of initial training was switched from universities to schools and school-university partnerships (Whitty 2014) and a flagship scheme, Teach First, was launched to encourage high performing graduates into teaching (Teach First undated). In respect of professional practice, a teacher workload survey and review was commissioned, to both understand and act on the reported problems of workload and role.

These initiatives and developments—and the problems to which they are a response—are major aspects of the context which needs to be taken into account when interpreting the 2018 PISA results. A lot has changed in the period prior to the 2018 PISA data capture on 15 year olds. Both the scale of change and, crucially, the timing of the impact of the changes are embedded in the pattern of the PISA outcomes.

The scale of the post-2010 changes—so many aspects of the education system simultaneously being re-cast by new policies—were criticized by Parliamentary Select Committee as ‘hyperactivity’ on the part of the then-Secretary of State Michael Gove (Guardian 2013b). But the need to work across all aspects of arrangements was driven by explicit theory, not mere personal tendency of the Secretary of State. The commitment to international benchmarking and to ‘policy learning’ from international comparative research including examination of the characteristics of high performing systems. Bill Schmidt’s work on ‘curriculum coherence’—where instruction, assessment, standards and materials carefully and deliberately are aligned—was extended into a wider examination of coherence across all key aspects of arrangements (Cambridge Assessment 2017). In 2010 this mobilized wide ranging policy review by the Secretary of State. This emphasised the need to ensure coherence across accountability, curriculum standards, professional practice, institutional development and so on—aiming to remove in particular the tensions between accountability (targets, reported data, inspection) and curriculum aims which had been evident for so long (Hall and Ozerk 2008). The use of international evidence on effective pedagogy

in reading and maths was focussed intently on securing ‘curriculum coherence’. All this was not so much ‘hyper-activity’ as a perceived need to align key elements of arrangements quickly and effectively; a demanding act of public policy.

4 PISA as a Reflection on Post-2010 Policy Moves

So, it was in the context of these changes which journalists and researchers waited for the 2019 publication of the 2018 PISA results. It was clear that a simple look at timing of the policy actions—emphasis on reading, new National Curriculum, new qualifications—should give everyone pause for thought. Timing of change, and the time lag involved in genuine system impact is essential in interpreting international survey data such as PISA, TIMSS and PIRLS. This seemed entirely to be forgotten in the noise in 2001 after Finland was announced as ‘first in the world’ (BBC 2015). Time lags and the necessity of relating the timing of actions to effects were ignored by the majority of commentators. Profoundly misleading narratives have been created as a result.

It is a necessary statement of the obvious that the pupils in the PISA survey were 15 years of age. In England, most of the surveyed pupils were in their 11th year of schooling. For those in year 11, they had only experienced the new National Curriculum for their secondary education—from year 7 (age 11). They had not experienced a full education aged 5–15 under the new curriculum. The new curriculum was introduced in September 2014, when the PISA cohort already had entered secondary education. But it is important to note that the revised National Curriculum was intended to have its biggest impact in Primary Education. The PISA cohort was not exposed to this new curriculum. And no new national curriculum, especially one committed to a radical reshaping of learning, is implemented perfectly in its first year of operation. The intention of National Curriculum policy in England was to shape the curriculum in secondary education through the revised, more demanding specifications of national qualifications, along with national targets and accountability instruments—particularly the English Baccalaureate requirement.

The policy around years 7, 8 and 9—the first 3 years of secondary education—was controversial. The new National Curriculum was stated in a very specific and detailed form for Primary education, including a move to a new year-by-year format. Years 7, 8 and 9 were stated as a much more general requirement and, unlike the new Primary specification, was treated as a ‘Key Stage 11–14’ rather than as separate years. The controversial assumption of the policy was that the period from year 7 to year 11 would be seen by schools as a continuum of learning, culminating in GCSE examinations in around ten selected subjects. This attracted criticism, since there was a dominant notion in educational discussion that ‘learning already was far too heavily determined by exams and assessment’ (Mansell *op cit*). But the 2010–13 policy advisers committed to a ‘continuum of learning 11–14’ principle, with the detail of learning targets linked to the detailed and carefully designed objectives of national examinations rather than a set of detailed year-by-year National Curriculum

statements. They felt that this would not lead to narrow, restrictive learning, and by contrast should equip pupils with the wide reading, extended writing, critical thinking, rich discussion etc. which would lead to enhanced outcomes in national exams at 16. Policy advisers noted the evidence, from teachers, of the strong lower secondary ‘washback effect’ from the demands of national examinations and, rather than working against it, intended that it be used to intensify and focus the learning in the first years of secondary education. There was no strong, explicit statement of this principle, since the washback effect was so strongly evident in the system. While national inspection reports noted that years 11–13 frequently were the ‘lost years’ in the system (Ofsted 2015), the policy assumption was that the significantly increased demand of the new national examinations and accountability requirements would intensify these early secondary years.

But for interpretation of the PISA results, it is essential to note that the new examinations were only introduced in 2015 (maths and English) and 2016 (other subjects). The 2018 PISA cohort therefore did not experience either the new National Curriculum during their primary education 5–11 nor the intended intensification of 11–14 education. They also experienced the new qualifications only in the immediate years after first implementation, a period typically associated with sub-optimal performance of the system due to (i) lack of established, stable and refined teaching approaches, (ii) uncertainty regarding exact expectations, (iii) unrefined professional support (Baird et al 2019).

Thus, in November 2019, when expecting the PISA results, some researchers and commentators—including this author—were taking these factors into account and anticipating the possibility of static or even depressed PISA outcomes for England. Yet, on publication, England’s results showed significant improvements in mathematics, an apparently stable position in reading but improved performance relative to other benchmark nations, and high but static performance in science. How should we interpret this? The simplest explanation is also a positive one; that despite the depressive effects of system changes, real gains in attainment were being secured.

The improved maths performance comes after decades of protracted flat performance. Again, carefully considering timelines, it was anticipated that pupils in previous PISA cycles might benefit from the Numeracy strategies of the late 1990s—but no such effect is obvious in the data, unless the time lag is unfeasibly long. The increase corresponds more exactly with the post-2010 emphasis on mathematics—the wide and varied practical policy combined with high profile public discourse, demanding targets, and some ‘washback’ from elevated standards in public examinations.

The reading scores demand careful interpretation. PIRLS data shows an increase 2006–2011 which suggests some elevation of reading prior to 2010. The high level of reaction against phonics in 2010 suggested that varied methods for early reading were established in the system. School inspection reports reinforce this view (Ofsted 2002; Department for Education 2011). The strong emphasis on synthetic phonics, re-enforced by the ‘phonics check’ in Year 1 co-incides with an elevation of performance and closing of the gender gap in PIRLS. In PISA—assessing 15-year-olds—with only small variation in scores since 2006, performance of the education system

appears moribund. But this is deceptive. Background trends need to be taken into account. As we know from the Finland context, reading is significantly influenced by factors outside the school system (Tveit 1991). The 2018 PISA elevated reading scores in the USA are all the more remarkable in the light of the significant gradient of decline in reading speed and related comprehension since the 1960s (Spichtig et al. 2016). Likewise, England's score should not be seen solely in terms of the country's static trend in PISA, but in the relation between England's trends and those of other high-performing nations. In 2015, 12 countries outperformed England. Germany, Japan, New Zealand and Norway all outperformed England in 2015, but had similar scores to England in 2018. Those that had similar scores in 2015—Belgium, France, Netherlands, Portugal, Russian Federation, Slovenia, and Switzerland) were outperformed by England in 2018. These comparisons cast an interesting light on the seemingly static performance in England. In addition, the gender gap is significantly lower than the OECD average. However, equity remains challenging. Notably, the increase particularly has been amongst higher performing pupils. The lowest achievers' scores have remained static, but the difference between high and low achievers in 2018 is similar to the OECD average. The international picture, from both PISA and other sources, suggests a *significant international widespread decline* in reading—and this without measuring the technology-driven switch in reading habits and family environment which is occurring over a dramatically compressed timeframe (Kucirkova and Littleton 2016). With the PIRLS data showing similar relative improvement in performance for England, the results appear to endorse the post-2010 policy emphasis on reading—similar to the macro and micro policy emphasis on maths—and substantial benefit in the practical action which has been put in place with schools.

A sub-domain in the 2018 sweep, science results in England provide a different story to maths and reading, and suggest that government should both sustain its approaches in those two subjects *and* attend to policy action aimed at primary and secondary science performance. With very few specialist science teachers in Primary, and science testing withdrawn from national assessment prior to national qualifications at 16, incentives and drivers have declined in the both the pre- and post-2010 period. While over this period over 400 initiatives from various sources (House of Lords 2006), of various types and various scales have been implemented across arrangements, TIMSS Grade 4 data in 2011 showed a fallback to late 1990s performance levels (Sturman et al. 2012).

Performance in science is not crashing, it just remains static—and shows challenging equity outcomes—a high gap between high and low achieving pupils, and with a higher proportion of pupils achieving at the highest proficiency levels. England bucks the international trend in terms of gender: no significant gap in attainment, unlike the OECD gap in favour of females. But it is essential to recognise that overall the gender gap data in England is not positive: national qualifications data in science subjects at 16 and particularly at 18 remain highly gendered (Cassidy et al. 2018).

The 2018 mean score for England remains higher than the OECD average, and 10 nations had overall scores higher than England. But 12 others in 2018 had a significant

drop in performance: Australia, Canada, Denmark, Finland, Japan, Norway, Spain and Switzerland. Only two secured an increase: Poland and Turkey.

5 England Within the United Kingdom; the Devolved Administrations

All of this gives the post-results overview for England, linking it to policy actions and long term timelines. But this is England—what of the United Kingdom, of which England is a part? Usefully, the survey design and the implementation requirements result in PISA providing valid data for the devolved administrations of the UK—Scotland, Northern Ireland and Wales. I have said nothing of these latter territories. And deliberately so. They are different from England in vital ways, and different from each other (with an emerging important exception in the apparent increasing convergence of Wales with Scotland). All of this provides the most extraordinary natural experiment—the late David Raffe’s ‘Home international comparisons’ (Raffe 2000).

Scotland has for the past two decades worked on the ‘Curriculum for Excellence’; an increasingly controversial approach to defining, specifying and delivering the primary and secondary curriculum which emerged from a 2002 consultation exercise. Implemented in 2010–11, with new qualifications in 2014, it uses models strongly contrasting with those in England (Priestly and Minty 2013). For Scotland, an increase in reading in PISA 2018 accompanies an unarrested decline from 2000 in maths and science.

Wales is undertaking radical reform, in the light of a previous history of low results relative to the rest of the UK and of previously declining scores. It is looking to the Scottish model in the ‘Curriculum for Excellence’ (Priestly op cit; TES 2019b) rather than the post-2010 actions and models in England. But timelines remain important. The new curriculum model in Wales has not yet been fully designed and specified, let alone implemented—the target date for enactment being 2020. Yet in 2018, reading scores improved over 2015, science reversed a severe decline, and maths continued the improvement which was first seen in the 2015 PISA outcomes.

Northern Ireland remains distinctive and different—its arrangements heavily shaped by its history and its size and geography. Pupils there performed better than pupils in Wales, but slightly lower than England in all three domains. Performance was below Scotland in reading—Scotland’s improved domain—but above Scotland in science and maths. However, science in Northern Ireland has shown significant decline since 2015, despite an unchanged position 2015–2018.

When considering the relative performance of England, Wales, Northern Ireland and Scotland, I have emphasised just how essential it is to avoid lapsing into any assumptions or view that ‘The UK’ can be regarded as a unitary system. It cannot. Indeed, far from it: the systems in the different administrations are highly distinctive, increasingly driven by very different assumptions, models and policy instruments.

6 Diversity and Difference

And it is on this note of ‘difference’ which this chapter ends. David Reynolds, co-author of the influential and incisive transnational 1986 report ‘Worlds Apart’ (Reynolds and Farrell 1996) has repeatedly emphasised the importance of within-school variation in England, which remains amongst the highest in high-performing PISA nations. Arrangements in England also manifest high between-school variation. This hints at a level of variation in educational forms which has been under-analysed and under-recognised. Diversity of institutional forms, curriculum assumptions, professional practices is extraordinary in England. In a country of 66.4 million (ONS 2019) there are extremely large and very small schools and everything in between. There are free schools, state schools, independent schools, academies and academy chains. There are selective authorities and non-selective areas. The transfer age between primary, lower secondary and upper secondary varies from area to area. Setting and streaming is managed in very different ways in different schools. Micro-markets have emerged in different localities with different patterns of schools (Gerard 1997). There is historical legacy which gives large variations in school funding from area to area. There is a possibility of parental choice of school in some areas and operationally, none in others. Regional variation in growth and economic activity shows similar extreme variation (ESCOE 2019).

The picture of diversity in educational performance in England is rendered complex by the peculiar distribution of the ‘unit of improvement’. In some cases, schools in ‘academy chains’ (allied groups of schools) are clustered in a locality—in other instances, they are widely geographically distributed. In addition, city development strategies of the early 2000s—the most prominent being ‘London Challenge’—lent policy integrity to a specific urban localities. While the underlying causes and extent of improvement associated with London Challenge are contested (Maddougall and Lupton 2018), the ‘City Challenges’ indicate a period of change where key ‘units of development’ were at the level of large urban centres, rather than the nation as a whole.

Research which the author undertook across England during the National Curriculum review showed not only high variation in forms of education and professional practice, but high variation in assumptions and values regarding curriculum, assessment, and pupil ability. Few other nations appear to possess such high structural and internal variation across all dimensions of provision. This poses a massive challenge to policy makers, who must anticipate very different conditions for the impact of national policy measures, and high resilience and resistance (Cambridge Assessment 2017). Sustained improvement therefore suggests particular potency and design integrity in the forms of public policy which legitimately can causally be associated with that improvement. The Reading and Maths results should be seen as signs of genuine policy achievement in a highly diverse and challenging context. When improvement bucks the trends present in society, as in the case of literacy, policy makers can be doubly satisfied with their and teachers’ endeavours.

References

- Abdazi, H. (2014). How to define education goals for the post-2015 period? Guidance from memory functions unpublished working paper.
- Alexiadou, N., Dovemark, M., & Erixon-Arreman, I. (2016). *Managing Inclusion in Competitive School Systems: The Cases of Sweden and England Research in Comparative and International Education*. <https://doi.org/10.1177/1745499916631065> accessed 030320.
- Au, C., Entwistle, N. (1999). Memorisation with understanding. In *approaches to studying: cultural variant or response to assessment demands?* University of Edinburgh Education Line Database.
- Baird, J. A., Caro, D., Elliot, V., El Masri, Y., Ingram, J., Isaacs, T., et al. (2019). *Examination Reform: Impact of linear and modular examinations at GCSE*. Oxford/Coventry: OUCEA/Ofqual.
- BBC. (2015). *Finns aren't what they used to be 15 04 20 BBC News* <https://www.bbc.co.uk/news/business-32302374>. Accessed March 27, 20 18:20.
- BBC. (2018) Times tables check trialled ahead of rollout 14 02 18 BBC News <https://www.bbc.co.uk/news/education-43046142>. Accessed March 27, 20 14:10.
- Benton, T., & Carroll, M. (2018). *The link between subject choices and achievement as GCSE and performance in PISA 2015*. Cambridge: Cambridge Assessment.
- Bradbury A undated Pressure, anxiety and collateral damage—the headteachers' verdict on SATs, UCL Institute of Education, London.
- Cambridge Assessment. (2010). *Exam standards: The big debate*. Cambridge: Cambridge Assessment.
- Cambridge Assessment. (2017). *A Cambridge Approach to improving education*. Cambridge: Cambridge Assessment.
- Cassidy, R., Cattán, S., & Crawford, C. (2018). *Why don't more girls study maths and physics?* London: Institute of Fiscal Studies.
- Cheong, A. Kam, H. (1992). *Singapore students are rote learners; Fact or Myth? ERA conference paper*, September 24–26, 1992, Educational Research Association of Singapore (ERAS).
- Christodoulou, D. (2014). *Seven myths about education*. Abingdon: Routledge.
- Clark, M. (2018). How justified is it to dominate government policy on early literacy with synthetic phonics and the phonics screening check part II a critique of the evidence 14. *Education Journal, Issue, 352*, 14–19.
- Crehan, L. (2016). *Cleverlands: The secrets behind the success of the world's education superpowers*. London: Unbound.
- Department for Education. (2011). *The National Strategies 1997–2011*. London: Department for Education.
- Dolton, P., & She, P. (2018). *Global Teacher Status Index*. London: National Institute of Economic and Social Research.
- Elliott, G., Rushton, N., Darlington, E., & Child, S. (2015). *Are claims that GCSE is a white elephant red herrings?* Cambridge: Cambridge Assessment.
- ESCOE. (2019). *Differences in economic growth across the UK remain substantial in the year to Q1 2019, Economic Statistics Centre of Excellence* <https://www.escoe.ac.uk/differences-in-economic-growth-across-uk-remained-substantial/>. Accessed February 14, 20 12:50.
- Financial Times. (2016). *Ambitious education reforms fail to lift UK Pisa school rankings* December 06, 16 Financial Times <https://www.ft.com/content/1ea10484-bb14-11e6-8b45-b8b81dd5d080>. Accessed March 19 20, 16:15.
- Gorard, S. (1997). *Market forces, choice and diversity in education: the early impact Sociological Research Online* <https://doi.org/10.5153/sro.111>. Accessed February 12, 20 15:50.
- Guardian. (2013a). *Michael Gove unveils GCSE reforms*, June 11, 13, Guardian.
- Guardian. (2013b). *Graham Stuart: 'You can't drive education like a sports car'*, March 25, 13, Guardian.
- Guardian. (2014). *Reading Lessons; why synthetic phonics doesn't work*, March 04, 14, Guardian.
- Hall, K., & Ozerk, K. (2008). *Primary Curriculum and assessment: England and other countries Cambridge Primary Review*. Cambridge: University of Cambridge.

- House of Commons Library. (2019). *English Baccalaureate* <https://researchbriefings.parliament.uk/ResearchBriefing/Summary/SN06045>. Accessed March 12, 20 10:20.
- House of Lords. (2006). *Science teaching in schools: Report with evidence; 10th report of session House of Lords Science and Technology Committee*. London: The Stationary Office.
- Jay, M. McGrath-Lone, L., Gilbert, R. (2019). Data Resource; the National Pupil Database (NPD). *International Journal of Population Data Science*, 4(1) <https://ijpds.org/article/view/1101>. Accessed March 01, 20 16:00.
- Jerrim, J. (2011). *England's "plummeting" PISA test scores between 2000 and 2009: Is the performance of our secondary school pupils really in relative decline?* DoQSS Working Papers 11–09, Department of Quantitative Social Science, UCL Institute of Education, London.
- Jerrim, J., Micklewright, J., Heine, J. H., Salzer, C., & McKeown, C. (2018). PISA 2015: How big is the 'mode effect' and what has been done about it? *Oxford Review of Education*, 44(4), 476–493.
- Kirschner, P., Sweller, J., & Clark, R. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of Constructivist. *Discovery, Problem-Based, Experiential and Inquiry-Based Teaching, Educational Psychologist*, 41(2), 75–86.
- Kucirkova, N., & Littleton, K. (2016). *The digital reading habits of children*. Manchester: Arts Council England.
- Macdougall, A., & Lupton, R. (2018). *The 'London Effect': Literature review*. Manchester: University of Manchester & The Rowntree Foundation.
- Machin, S., McNally, S., Viarengo. (2016). Changing how literacy is taught: evidence on synthetic phonics. *American Economic Journal*, 10 (2), 217–241.
- Mansell, W. (2007). *Education by numbers: the tyranny of testing*, Politico's Publishing, Michigan.
- McGrane, et al. (2017). *Progress in International Reading Literacy Survey (PIRLS): National Report for England December 2017 DFE-RR770*. London: Department for Education.
- Micklewright, J., & Schnepf, S. V. (2006). *Response bias in England in PISA 2000 and 2003 Department for Education Research Report RR771*. London: Department for Education.
- National Audit Office. (2018). *Converting maintained schools to academies PN: 15/18*. London: National Audit Office Press Office.
- Newton, P. (2008). *Monitoring national attainment standards—a collection of working papers*, Office of Qualifications and Examinations Regulation, Coventry.
- NFER. (2018). *NFER educational briefings—Key insights for England from PIRLS, TIMSS and PISA*, National Foundation for Education Research, Slough.
- NFER. (2019). *Teacher labour market in England—Annual Report 2019*, National Foundation for Educational Research, Slough.
- Oates, T. (2014). The 'qualifications sledgehammer': why assessment-led reform has dominated the education landscape. In: Heller-Sahlgren, G. (ed) *Tests worth teaching to*, Centre for Market Reform of Education, London.
- OECD. (2003). *Attracting, developing and retaining effective teachers: Country background report for Finland*. Paris: Organisation for Economic Co-operation and Development.
- Ofsted. (2002). *The National Literacy Strategy: the first four years 1998–2002 Office for Standards in Education*, London.
- Ofqual. (2017). *Modelling GCSE grade 9 to 1 outcomes in 2017: A technical explanation*, Office of Qualifications and Examinations Regulation, Coventry.
- Ofqual. (2018). *GCSE 9 to 1 grades: a brief guide for parents 02 03 18 Blog* <https://ofqual.blog.gov.uk/2018/03/02/gcse-9-to-1-grades-a-brief-guide-for-parents/accessed> February 21, 20.
- Ofsted. (2015). *Key Stage 3: the wasted years?* Office for Standards in Education, London.
- ONS. (2019). *Population estimates for the UK, England, and Wales, Scotland and Northern Ireland: Mid-2018*. Newport: Office for National Statistics.
- Priestly, M., & Minty, S. (2013). Curriculum for excellence; 'A brilliant idea, but....' *Scottish Educational Review*, 45(1), 39–52.
- Raffe, D. (2000). *'Home international' comparisons of post-16 education and training*. CES Briefing: University of Edinburgh, Edinburgh.

- Reynolds, D., & Farrell, S. (1996). *Worlds apart?: A review of international surveys of education achievement involving England*. London: Stationary Office Books.
- Sammons, P., et al. (2008). *The influence of school and teaching quality on children's progress in primary school DCSF-RR028*. Schools and Families, London: Department for Children.
- Schmidt, W. (2018). Why there is no “holy grail” in education. Lecture Cambridge 30 05 18 available at <https://cambridgeassessment.org.uk/news/why-theresno-holy-grail-in-education/>. Cambridge: Cambridge Assessment.
- Schools Week. (2019). *PISA 2018: Maths up, no improvement in reading or science*, December 03, 19, Schools Week <https://schoolsweek.co.uk/pisa-2018-maths-up-no-improvement-in-reading-or-science/>. Accessed January 23, 20 14:10.
- Sizmur, J. et al. (2019) *Achievement of 15-year-olds in England: PISA 2018 Results DFE-RR961* Department for Education, London.
- Spichtig, A., Hiebert, E., Vorstius, C., Pascoe, J., Pearson, D., & Radach, R. (2016). The decline of comprehension-based silent reading efficiency in the United States: A comparison of current data with performance in 1960. *Reading Research Quarterly*, 51(2), 239–259.
- Sturman, L., Burge, B., Cook, R., & Weaving, H. (2012). *TIMSS 2011: Mathematics and science achievement in England*. Slough: National Foundation for Educational Research.
- Teach First undated Our work and its impact, Teach First, London.
- TES. (2018a). *England held back by rote-learning, warns PISA boss*, March 23, 18, TES Global Ltd.
- TES. (2018b). *English national pupil database one of the most impressive in the world*, January 22, 18, TES Global Ltd.
- TES. (2019a). *Pisa results must be a relief for the government*, December 03, 19, TES Global Ltd <https://www.tes.com/news/pisa-results-must-be-relief-government>. Accessed January 12, 20 13:40.
- TES. (2019b). *Scotland's mistakes will shape Welsh education reform*, February 18, 2019, Times Global Ltd.
- Tveit, K. (1991). The development of popular literacy in the Nordic Countries. *a Comparative Historical Study, Scandanavian Journal of Educational Research*, 35(4), 241–252.
- Willingham, D. (2017). *The Reading Mind: A cognitive approach to understanding how the mind reads*. San Francisco: Jossey-Bass.
- Whitty, G. (2014). *Recent developments in teacher training and their consequences for the University Project*. London: University College London.

Tim Oates is Group Director of Assessment Research and Development at Cambridge Assessment, focusing on national and international research on assessment and measurement. In 2010 he published ‘Could do better’ which laid down the principles for the review of the National Curriculum in England. He was chair of the Expert Panel for Review of the National Curriculum in England. Emerging from this review, subsequent research on the quality and function of textbooks and other resources has been taken up around the world and discussed at two international summits on learning resources. He chairs various curriculum groups for the Department for Education in England. He has published widely on assessment and curriculum issues, and routinely provides briefings and advice to UK and other governments. He has worked with OECD on curriculum matters and is leading a new UNICET project on a curriculum framework for displaced children. He is Fellow of Churchill College Cambridge and in 2015 received a CBE for services to education.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.