
All Above Average: secondary school improvement as an impossible endeavour

PHIL TAYLOR

ABSTRACT This article argues that secondary school improvement in England, when viewed as a system, has become an impossible endeavour. This arises from the conflation of improvement with effectiveness, judged by a narrow range of outcome measures and driven by demands that all schools should somehow be above average. The expectation of comparable year-on-year examination results at age 15/16 in order to maintain standards of performance persists in uneasy tension with calls for continual improvement. The examination system acts as a limiter and sorter, with students, teachers and schools competing for grades that are constrained to a normal curve. GCSEs and their equivalents increasingly serve less to allow young people to demonstrate their achievements and more for holding schools and teachers to account. This has major implications for the justifiable desire that all our youngsters should learn in a 'good' school, which adds value and narrows gaps. As some schools push forward in improving these measures, others inevitably roll back. Questions are raised regarding the sense and wisdom in maintaining the current situation, aimed at shaking off our normal-curve and above-average conditioning.

School improvement in England has become increasingly determined by measurable outcomes associated with a narrow conception of school effectiveness. An array of attainment and progress indicators is published each year in performance tables and used by the Office for Standards in Education, Children's Services and Skills (Ofsted) to make judgements about schools and drive improvement. The ongoing demand for an impossible endeavour of improved performance based on these measures is reminiscent of the mythical Greek king Sisyphus, condemned to push a huge boulder uphill forever. When the secondary education system is viewed as a whole, while individual schools can show improvement this will always be at the expense of others which roll back. Our paradoxical thinking about school improvement, constrained by the

normal curve, seeks for all to be above average, in what Barker (2008) has called a 'relentless pursuit of the unattainable'.

The main argument of this article is simple and obvious, but rarely features in public debate. The central point is that measures used to gauge improvement, which teachers work so hard to raise, cannot be allowed to increase indefinitely across the whole country. This means that schools and colleges compete for a ration of available grades on behalf of their learners, who have little or no say in how they are assessed and in what subjects. The purpose of examinations at age 15/16, ostensibly for students to demonstrate their knowledge and skills, has become the differentiation of performance across cohorts in holding schools to account. Expectations of 'continual improvement' exist in tension with 'comparable outcomes' required by the examination system. I argue that these conflicting demands cannot be reconciled and should be reconsidered, posing questions aimed at stimulating debate about what we really hope to achieve for our young learners.

Comparable Outcomes versus Continual Improvement

There are many technical debates regarding the most appropriate ways to achieve comparability of examination outcomes, defining and maintaining standards over time (see, for example, Coe, 2010; Newton, 2010; Cambridge Assessment, 2011). However, for the purposes of this article it is sufficient to conclude that, in recent years, the percentages of GCSEs (General Certificates of Secondary Education) awarded at each grade across all subjects have remained similar, within a skewed normal distribution (see Figure 1). This is the result of strenuous efforts by the Office of Qualifications and Examinations Regulation (Ofqual), its predecessor bodies and the exam boards it regulates to ensure what has become known as 'comparable outcomes'. This means that 'roughly the same proportion of students will achieve each grade as in the previous year' (Ofqual, 2012, p. 2). Also known as 'cohort referencing', it is achieved by a combination of examiner judgement in setting grade boundaries and statistical prediction from prior attainment at primary school and past performance. However, the cumulative effect of small year-on-year changes to the proportions of each grade on key threshold figures is more noticeable. For example, the percentage of A*-C grades awarded increased by 4.5% from 2008 to 2014 and, looking further back, Ofqual (2014, p. 9) has shown that since 1988 the A/A*-C grades awarded increased by over 25% (the A* grade was introduced in 1994).

Between 2005 and 2014, the improvement in the percentages of 15/16-year-olds gaining five or more GCSEs at grade C or above was mostly the result of vocational and applied qualifications being counted in addition to GCSEs themselves (see Figure 2). When these so-called equivalents (BTECs, OCR Nationals, etc.) are removed, the proportion of students achieving this measure has remained fairly stable at around 55%. Similarly, when English and mathematics are included, with equivalents still removed, the proportion

achieving five or more at grade C or above has remained close to 50%. So recent headline improvements in overall secondary school outcomes reflect the range of qualifications awarded more than threshold grade increases. This trend has now been reversed following the Wolf (Department for Education [DfE], 2011) Review of Vocational Education, and subsequent reforms to performance tables including changes to the inclusion and size of equivalent qualifications.

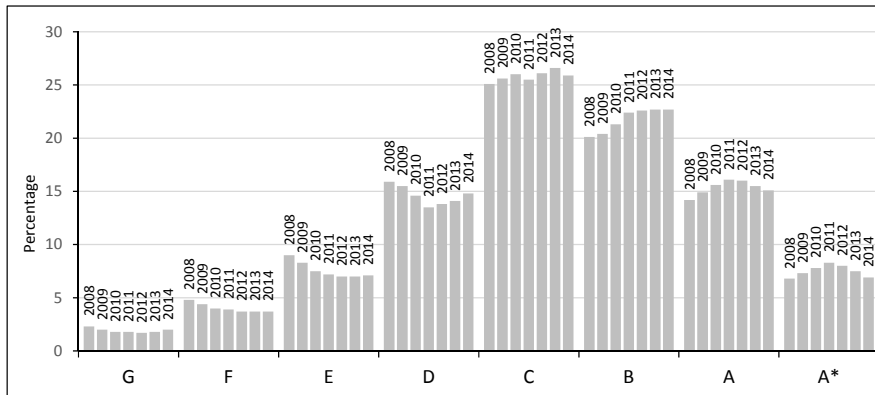


Figure 1. GCSE grade distribution (all subjects) 2008-14.
 Source: www.gov.uk/government/collections/statistics-gcses-key-stage-4

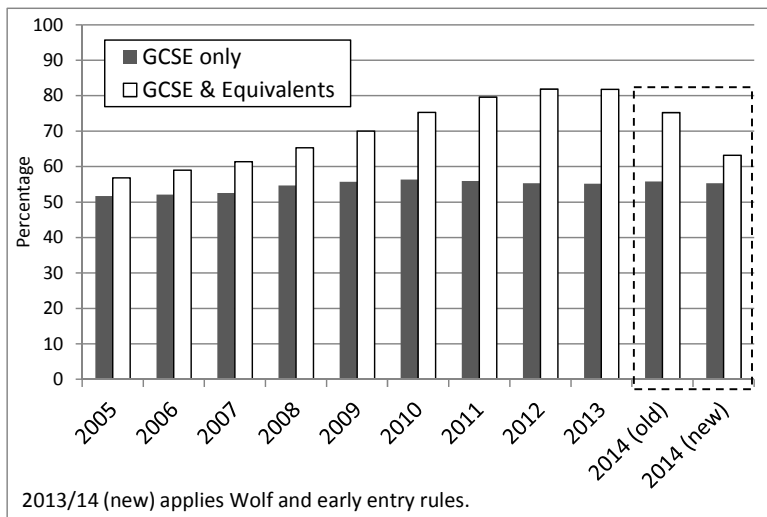


Figure 2. Percentage 5+ A*-C Grades 2005-14 (all schools in England).
 Source: www.gov.uk/government/collections/statistics-gcses-key-stage-4

Ofqual (2014, p. 10) acknowledges the pressure placed on schools for continual improvement, stating that its ‘approach aims to control grade inflation, but to

allow genuine improvements in performance to be recognised'. However, how such genuine improvements might be ascertained remains open to debate. The inherent tension in Ofqual's twin aim can be seen in recent government policy. For example, on 17 September 2012 (Parliamentary Archives, 2012a) the former Education Secretary, Michael Gove, announced changes to the examination system in England, stating: 'We believe it is time to tackle grade inflation and dumbing down, and we believe that it is time to raise aspirations and restore rigour to our examinations'.

A few months earlier, on 31 January 2012 (Parliamentary Archives, 2012b), the following exchange took place between the Chair of the Education Select Committee, Graham Stuart MP, and Michael Gove:

Graham Stuart: ... if 'good' requires pupil performance to exceed the national average, and if all schools must be good, how is this mathematically possible?

Michael Gove: By getting better all the time.

Graham Stuart: Chair: So it is possible, is it?

Michael Gove: It is possible to get better all the time.

It is difficult to imagine that Michael Gove thought that more schools could be above the national average by continually improving; however, this erroneous view underlies much school improvement endeavour. The inherent contradiction in decrying grade inflation while demanding improvement appears to escape notice. Perhaps a more plausible explanation is that the rhetoric of (regulated) comparable outcomes with (unrealisable) continual improvement exerts pressure on the education system and those working within it to perform as well as possible.

Every School a 'Good' (or Better) School

The notion that for a school to be 'good' or better requires 'above average' performance is enshrined within the criteria used by Ofsted inspectors when making judgements on effectiveness. The key measures used to ascertain a school's performance in relation to national outcomes have changed over recent years, with Ofsted's evaluation schedule reviewed three times a year. Currently, *expected progress* in English and in maths is foregrounded. These rather crude measures use the levels of attainment defined in the National Curriculum prior to September 2014, which specified level 4 as the age-related expectation at age 11 and broadly equated level 7 to GCSE grade C for age 15/16 students. Thus, the expected progress of a student reaching level 4 at age 11 is three levels in order to gain at least grade C. As a performance measure, expected progress states the percentage of students in the school who make at least three levels of progress from the end of primary school assessments to GCSE in English and also in maths. In 2013 the national averages were 71.3% and 71.9% respectively.

End of Key Stage 2 teacher assessments and tests in recent years have shown that between 85% and 90% of learners reach level 4 or above at age 11 in English and maths. If nearly all of these pupils were to make the expected progress we would see similar proportions reaching grade C or above at GCSE five years later. However, around 65% to 70% of students actually achieve grade C or above in these subjects due to the demand for comparable outcomes. This means that around 70% making the expected progress is inevitable, leading to a distribution of expected progress percentages across schools (see Figure 3), rendering what is, after all, *expected* as *impossible* for some.

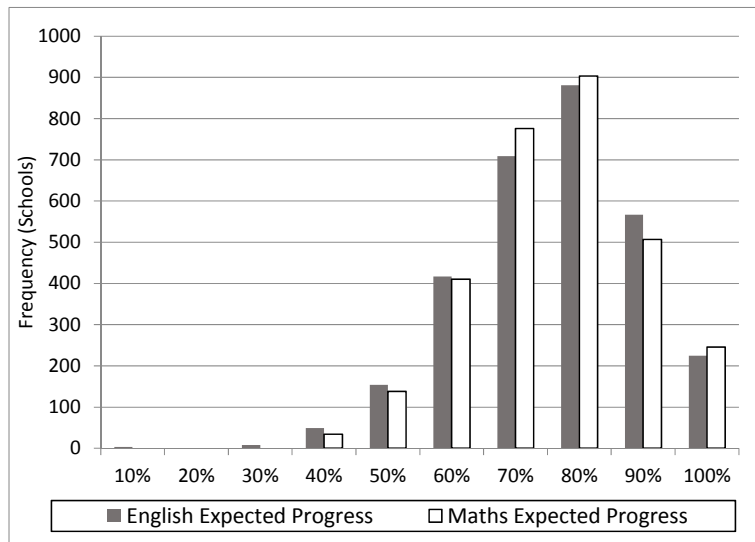


Figure 3. Distribution of Expected Progress percentages in 2013 (means: English 71.6%; maths 71.9%; 3014 secondary schools in England).
 Source: www.education.gov.uk/schools/performance/2013/

The latest (at the time of writing) Ofsted criteria for ‘good’ achievement in a school specify expected progress figures ‘close to or above national figures’, not only across the whole cohort but ‘[f]rom each different starting point’ (Ofsted, 2014a, p. 71). It should therefore come as no surprise that one-third of secondary schools were found wanting, for not yet being ‘good’, when the Chief Inspector’s annual report (Ofsted, 2014b) was published in December 2014. Of course, Ofsted uses inspection evidence as well as published data to judge school effectiveness, but its mission of ‘raising standards’ is largely determined by national averages. Another measure, which by definition perpetuates thinking in terms of cohort averages, is ‘value added’.

Value Added

Value added (VA) can be defined as the difference between actual and expected outcomes on some measure. In England, this measure has been the capped average points score (APS) derived from students' best eight GCSE grades, with English and maths double-counted in recent years. In the reformed performance tables, capped APS will become 'Attainment 8' and VA 'Progress 8' (DfE, 2013). The need to derive expected outcomes with which to compare actual results demands that a measure of prior attainment is used from which to project this expectation. Assessments of 11-year-olds at the end of primary school, reported in National Curriculum levels and again converted to points, have recently provided the necessary baseline. There is a strong correlation between these results and the median APS at age 15/16, so future expected outcomes can be mapped and predicted from historical patterns. It is worth noting a self-fulfilling prophecy at play here, because statistical projections of GCSE results from prior attainment at primary school are also used to influence comparable outcomes (as mentioned above). At the time of writing it remains to be seen how the necessary calculations will be performed in the absence of levels from the revised National Curriculum. Presumably results from future primary tests will be used in a similar way, but this change will not work through the system until 2020, when students taking the first of these new tests at age 11 reach their GCSEs.

Crucially, VA for the whole cohort is a zero-sum measure at any unit of analysis (individuals, groups, schools). In other words, there will always be around half of the individuals/groups/schools whose actual outcomes exceed those expected (predicted from prior attainment), with the other half falling short. Secondary school VA scores are centred on 1000 (by simply adding 1000 to each score) rather than zero (to avoid negative numbers). The graph of 2013 VA scores for 3014 secondary schools in rank order produces the 'snake-plot' in Figure 4, which also shows the considerable variation in 5+ A*-C, including English and maths percentages. This suggests that VA scores say as much about primary as secondary attainment – a positive VA can result from low primary and mediocre secondary outcomes. If a school improves its VA score one year, the measure itself dictates that this will be at the expense of another school or other schools where VA will fall. It is for this reason that year-on-year comparisons of value-added scores are discouraged. As an indicator of *effectiveness*, a school's VA provides a snapshot of performance in comparison to all schools nationally, but it is not a viable measure of *improvement*. Finally, school-level VA scores only say anything meaningful, if at all (see Gorard et al, 2012), for around the lower third and upper third of secondary schools. This is because errors in the calculations produce confidence limits (where there is a 95% probability of the VA score falling within the range) that vary according to school cohort size, with no statistical significance for the remaining third of schools in the middle band (see Figure 4).

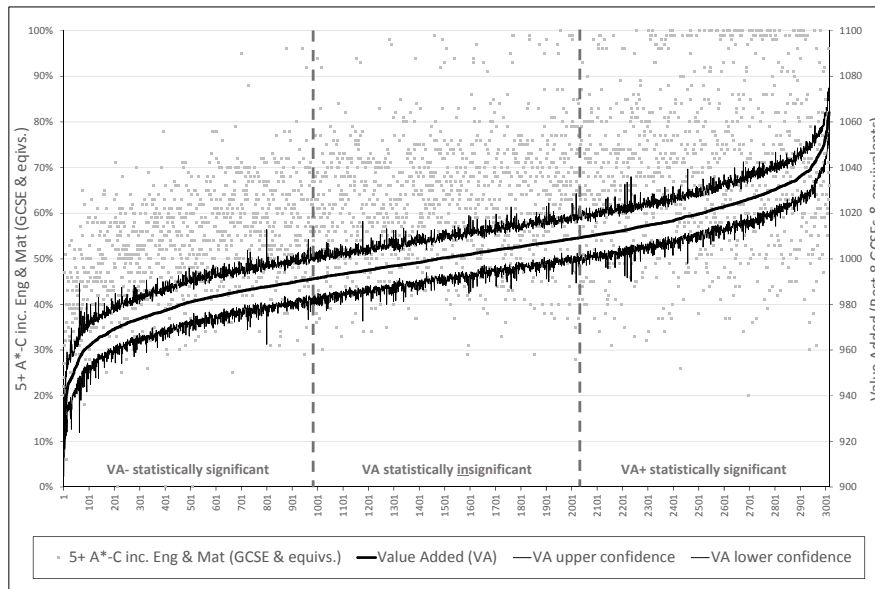


Figure 4. 5+ A*-C including English and maths (GCSE & equivalents) and value added (best 8 GCSE and equivalents) in 2013 (3014 secondary schools in England).

Source: www.education.gov.uk/schools/performance/2013

Narrowing Gaps

The gap between the percentage of 'disadvantaged' 15/16-year-olds achieving five or more GCSEs at grade C or above, including English and maths, and that of their non-disadvantaged peers has narrowed only slightly in recent years to around 27%. Currently, roughly 60% of young people each year reach this key measure (including equivalents). More specifically, this amounts to around 350,000 of each 575,000-strong cohort. Roughly 15% of each cohort are defined as disadvantaged by their take-up of free school meals within the last six years, not a particularly reliable proxy indicator (see Hobbs & Vignoles, 2010). Approximately 38% of disadvantaged students reach this benchmark measure, compared to 65% of their non-disadvantaged peers, leading to the 27% gap. The narrowing of this gap is central to current inspection criteria, with 'good' schools expected to show that 'attainment and progress of disadvantaged pupils are similar to or improving in relation to those of *other pupils nationally* and in the school' (Ofsted, 2014a, p. 71, emphasis in original).

There are two ways to close this particular gap – more disadvantaged students and/or fewer non-disadvantaged reach the benchmark. The former would of course be most desirable (no one wants to do *less* well). In 2013, to close the gap in this way, so that 65% of *all* students (both disadvantaged *and* non-disadvantaged) reached the benchmark, an additional 23,000

disadvantaged students would need to do so. But this assumes that an overall increase to 65% would be allowed by the examination system, with associated accusations of grade inflation or 'dumbing down'. If comparable outcomes mean the overall benchmark (including equivalents) remains at around 60% of all students, then the only conceivable way to close the gap is for *fewer* non-disadvantaged young people to reach it. In 2013, closing the gap this way, so that 60% of *all* students reached the benchmark, around 19,500 *fewer* non-disadvantaged students and 19,500 *more* disadvantaged students would need to do so.

Continued efforts to close the attainment gap for disadvantaged students are surely justifiable, and schools can be found where this has been achieved. However, we need to be clear about the constraints imposed by the examination system on realising this worthy aim. Also, it may not be fashionable to suggest, but it is possible that the perceived value in achieving prescribed educational benchmarks varies among students of different interests, aspirations and backgrounds. This is hinted at in a recent report of a longitudinal study, which concludes that '[y]oung people from deprived backgrounds, with greater needs, who are being bullied frequently or attending less successful schools tend to be less positive about education and to undertake more risky behaviours' (DfE, 2014, p. 21).

De-coupling School Effectiveness and Improvement and Finding New Indicators of Success

The points set out above only state the obvious, yet it is not necessarily commonly understood that for one individual, group or school to do better, i.e. *improve*, this must be at the expense of another if overall results remain broadly the same. On the contrary, calls for continual improvement, narrowed gaps and better outcomes suggest that we have either forgotten or do not realise that the examination system acts as a *limiter* and *sorter*. It is designed to differentiate attainment among young people, not to enable all to do well. So we should not be surprised when that is exactly what happens – some students, groups and schools do better than others.

With hindsight, calls to link school effectiveness and school improvement (e.g. Hopkins & Reynolds, 2001; Creemers & Reezigt, 2005) were perhaps misguided, as warned by others (e.g. Wrigley, 2004). These were two distinguishable but increasingly overlapping research and development traditions; effectiveness focusing on performance metrics as well as characteristics of success ('what works') and improvement more concerned with teacher and stakeholder led change. Despite both advocates and opponents of their merger advising against narrow definition of educational outcomes, effectiveness and improvement are combined in policy and practice through examination league tables, school inspection and top-down government initiatives. They require de-coupling, because measures currently used to gauge effectiveness cannot be improved without calling into question the integrity of

the measuring apparatus. School improvement should be based on a much stronger alignment of educational provision with the needs, interests and aspirations of the individuals, families and communities they serve.

Further, the wisdom of using the examination system to hold schools, and increasingly teachers, to account should be questioned. It is neither sensible nor tenable to expect examinations to simultaneously demonstrate recent student learning and readiness for future study, measure and quality-assure school and teacher effectiveness, and determine targets for continual improvement. If we are content with educational success measured in this way, then we are consigning up to half of our young people and schools to perceived failure. This is inevitable if outcomes are constrained to the normal distribution and success means above average. Forty-seven years ago Benjamin Bloom (1968, p. 2) suggested that we 'become "conditioned" to the normal distribution' and this seems to have remained so. He continued by stating:

There is nothing sacred about the normal curve. It is the distribution most appropriate to chance and random activity. Education is a purposeful activity and we seek to have the students learn what we have to teach. If we are effective in our instruction, the distribution of achievement should be very different from the normal curve. In fact, we may even insist that our educational efforts have been *unsuccessful* to the extent to which our distribution of achievement approximates the normal distribution. (Bloom, 1968, pp. 2-3, emphasis in original)

A renewed debate should ask *what* and *who* examinations at age 15/16 are for and whether we still need them, particularly when young people are expected to remain in education or training until age 18. If the goal of education is for every young person to find their talent and fulfil their potential, how do we assess and accredit this appropriately? Do *all* young people need to be assessed using the same or similar set of high-stakes examinations if they have differing talents, interests and aspirations? What are the ethical implications of an exam system with the main purpose of holding schools and teachers to account? Should the choice of which public assessments to take and when be for students, guided by parents and teachers, based on the learning they wish to demonstrate and the next steps in their education? How do we shake off normal-curve and above-average conditioning, genuinely raising our expectations of what young learners can achieve?

References

- Barker, B. (2008) School Reform Policy in England since 1988: relentless pursuit of the unattainable, *Journal of Education Policy*, 23(6), 669-683.
<http://dx.doi.org/10.1080/02680930802212887>

- Bloom, B. S. (1968) Learning for Mastery, *Evaluation Comment*, 1(2), 1-12. Center for the Study of Evaluation of Instructional Programs, University of California at Los Angeles.
- Cambridge Assessment (2011) *Research Matters: a Cambridge Assessment publication*, Special Issue 2: Comparability.
<http://www.cambridgeassessment.org.uk/Images/109991-research-matters-special-issue-2-comparability.pdf> (accessed 3 January 2015).
- Coe, R. (2010) Understanding Comparability of Examination Standards, *Research Papers in Education*, 25(3), 271-284.
<http://dx.doi.org/10.1080/02671522.2010.498143>
- Creemers, B.P.M. & Reezigt, G.J. (2005) Linking School Effectiveness and School Improvement: the background and outline of the project, *School Effectiveness and School Improvement*, 16(4), 359-371.
<http://dx.doi.org/10.1080/09243450500234484>
- Department for Education (2011) *Review of Vocational Education: the Wolf report*.
<https://www.gov.uk/government/publications/review-of-vocational-education-the-wolf-report> (accessed 3 January 2015).
- Department for Education (2013) *Reforming the Accountability System for Secondary Schools*.
<https://www.gov.uk/government/consultations/secondary-school-accountability-consultation> (accessed 3 January 2015).
- Department for Education (2014) *Longitudinal Study of Young People in England: cohort 2, wave 1*.
https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/374649/RR388_-_Longitudinal_study_of_young_people_in_England_cohort_2__wave_1.pdf (accessed 3 January 2015).
- Gorard, S., Hordosy, R. & Siddiqui, N. (2012) How Unstable are 'School Effects' Assessed by a Value-Added Technique?, *International Education Studies*, 6(1), 1-9.
<http://dx.doi.org/10.5539/ies.v6n1p1>
- Hobbs, G. & Vignoles, A. (2010) Is Children's Free School Meal 'Eligibility' a Good Proxy for Family Income?, *British Educational Research Journal*, 36(4), 673-690.
<http://dx.doi.org/10.1080/01411920903083111>
- Hopkins, D. & Reynolds, D. (2001) The Past, Present and Future of School Improvement: towards the Third Age, *British Educational Research Journal*, 27(4), 459-475. <http://dx.doi.org/10.1080/01411920120071461>
- Newton, P.E. (2010) Contrasting Conceptions of Comparability, *Research Papers in Education*, 25(3), 285-292. <http://dx.doi.org/10.1080/02671522.2010.498144>
- Office for Standards in Education, Children's Services and Skills (2014a) *Inspecting Schools: handbook for inspectors*.
<https://www.gov.uk/government/publications/School-Inspection-handbook> (accessed 3 January 2015).
- Office for Standards in Education, Children's Services and Skills (Ofsted) (2014b) *Ofsted Annual Report 2013/14: schools report*.
<https://www.gov.uk/government/publications/Ofsted-Annual-Report-201314-Schools-report> (accessed 3 January 2015).

- Office of Qualifications and Examinations Regulation (Ofqual) (2012) GCSEs and A levels in Summer 2012: our approach to setting and maintaining standards.
http://dera.ioe.ac.uk/15941/1/2012-05-09-maintaining-standards-in-summer-2012.pdf_Itemid%3D217 (accessed 3 January 2015).
- Office of Qualifications and Examinations Regulation (Ofqual) (2014) Setting GCSE, AS and A level grade standards in Summer 2014.
<http://dera.ioe.ac.uk/20136/1/2014-04-11-setting-gcse-as-and-a-level-grade-standards-in-summer-2014-pr...%20%281%29.docx> (accessed 3 January 2015).
- Parliamentary Archives (2012a) HC/1786i.
<http://www.publications.parliament.uk/pa/cm201012/cmselect/cmeduc/uc1786-i/uc178601.htm> (accessed 3 January 2015).
- Parliamentary Archives (2012b) Column 653.
<http://www.publications.parliament.uk/pa/cm201213/cmhansrd/cm120917/debtext/120917-0001.htm> (accessed 3 January 2015).
- Wrigley, T. (2004) 'School Effectiveness': the problem of reductionism, *British Educational Research Journal*, 30(2), 227-244.
<http://dx.doi.org/10.1080/0141192042000195272>
-

PHIL TAYLOR, School of Education, Birmingham City University.
Correspondence: phil.taylor@bcu.ac.uk

