Accountability and the Meaning of 'Success' In Education Systems and STEM Subjects: A Report to The Royal Society

Professor David Reynolds and
Professor Tony Kelly
University of Southampton

Email: D.Reynolds@soton.ac.uk

November 2013

Acknowledgements

The authors are grateful to the following colleagues for their assistance in the preparation of this report:

Caro Garrett
Marcus Grace
Keith Jones
Monty Paul
John Woollard

Accountability and the 'Meaning of Success' in Education Systems and STEM Subjects: A Report to the Royal Society.

Executive Summary and Recommendations

A description is given of different kinds of accountability systems and of the rise of 'performance based' and 'simultaneous supply-side / demand-side' educational policies in the United Kingdom since 1980. It is argued that these systems and policies have been useful, but that they need improvement in their operationalisation, both in terms of content and in the support given to professionals to cope with the resulting pressure. A review of the (limited) research literature on accountability, virtually all of which has come from secondary schools rather than Further Education Colleges because the traditional benchmark assessment occurs at GCSE at age 16, shows that the present accountability and evaluation systems have positive effects on student outcomes, but an analysis of contemporary policy developments in the STEM subject areas shows some problematic issues relating both to government education policy and to the adequacy of professional resourcing. Some interesting developments in STEM subjects from other countries are also discussed.

We recommend a range of policy changes currently. In terms of provision of education these are:

- Developing better formal internal school evaluation – this may act in a synergistic fashion to enhance the effectiveness of inspection systems and could be obtained by a focus on schools' own within school variation and building of capacity in the areas of middle management development, teaching and learning, data usage and use of student voice;
- Changing school Inspection to focus more on student outcomes – this could be obtained by focussing less on how outcomes were obtained in terms of whether certain desirable processes were used, and more upon whether schools had in place the systems of evaluation and indeed of 'educational thinking' that are necessary to generate high quality outcomes;
- Communicating better with professionals – given there are hints from the literature as to the lack of legitimacy that is sometimes afforded to government actions by professionals in education, a concern to communicate with professionals better (and not necessarily in the same way as to the wider electorate) may be helpful;
- Develop the 'Meso' or Middle Level of Accountability – the 'meso' level, mid way between national government and the individual school can operate to support

schools and indeed pressure them as necessary. It is likely to be effective in this given its closeness to schools geographically and organisationally, and given its likely knowledge of schools, communities, contexts and possible sources of support. Examples of this 'meso' level are the better local authorities and Academy 'Chains', where the 'meso' level is aligned with the needs of schools, is itself self-evaluating, is involved in training in schools and is involved in the generation of data systems that permit good practice to emerge at school level;

- Reconceptualising the STEM curriculum – international evidence suggests that what counts as 'success' in other societies, in terms of the nature of the science that is viewed as more desirable, may be useful for a UK audience to pay attention to, as may new methodologies of curriculum and instruction that are also being trialled in some countries.

In terms of the expectations and demands of society on the educational system, these are:

- Increasing the number of STEM-relevant outcome variables upon which schools are judged – given the likely destination of STEM school and university graduates, the 'softer' skills of collaborative working and good team working/building may be important attributes. Assessing school performance in these areas, additional to that in academic achievement, may therefore be useful;
- Enhancing Demand Side Capacity to respond to choice – the operation of parental choice of schools as a lever to improve school quality through the operation of 'market based' competition and pressure on schools will be reduced in effects if some schools are unable to expand in size because of building restrictions. In England, particularly, of the four devolved administrations, this seems to be a problem in some schools/areas, and ways of improving the situation both in terms of increased resources for school building and making it easier for extra suppliers of school places need to be found;
- Audit the Component Parts of the Demand Side – the importance given to performance data in educational systems that measure 'performance' needs to be paralleled by increased attempts to audit the systems of schools and Examination Boards that produce the data;
- Make the Performance Indicators more STEM-Relevant – the historic use of English and Mathematics at GCSE level as a benchmark of achievement neglects the subject of, and the role of, Science. This needs to change and it is interestingly in the proposals for England in which performance in eight subjects including Science will generate school positions in 'league' tables in future.

We also suggest developing both the teaching and school management professions involving changes in initial teacher education, enhancement of CPD relevant to STEM and developing within–school improvement for STEM subjects.

We conclude with a 'Commentary' on future policy areas of importance over the next 15/20 years. Specifically:

- We argue for the introduction of a new accountability system across the countries of the UK, using a range of existing, adapted and new policies;
- We argue that there are major societal and educational changes likely, to which the STEM subject community needs to adjust and respond.

An extensive Bibliography is provided.

Introduction

The American humourist Damon Runyan once argued that 'life is six to four against'! He could have been speaking about attempts to improve education systems and policies, because disappointment with the performance of national systems and with the results of individual policy reforms has been an international characteristic over the last thirty years. In spite of this, 'education' has still remained a repository for faith from governments, pressure groups, educationalists and others for the obvious reason that we have made some progress:

- There *are* highly performing countries which offer blueprints for successful education;
- There is so much internal variation within national systems that all possess, somewhere, the excellence against which practice can be benchmarked and spread within countries;
- Over time, the achievements of children in virtually all countries of the world are improving, although maybe not as fast as reformers would wish.

In the case of STEM subjects, it is also important to recognise that considerable progress has been made in recent years in establishing what 'effective' practice is, and how to improve the take-up of that practice. In Mathematics, for example, we know that teaching methods that add value to children's test scores should involve:

- High proportions of whole-class interactive teaching, in which knowledge is taken to children by teachers through active teaching models;
- Immediate help being given to students from teachers or from peers when difficulties are encountered, given that the hierarchical nature of the subject means that students cannot easily 'move on' to a new topic if they have not mastered their current work;
- We have also proven reliable programmes that can be adopted by schools to improve their performance (Slavin & Lake, 2008). (See Teddlie & Reynolds, 2000; Borich, 2006 & Muijs & Reynolds, 2011 for reviews of the research evidence).

For both Science and Mathematics we know that:

- The more that students can be encouraged to utilise their own social and community contexts from their own lives the better will be the outcomes;

- Learning will be potentiated the more that school activities can be supplemented by practical, out-of-school activities in industrial and community settings.

The questions for this review, though, are harder to answer than merely elaborating on these micro level programme and teaching strategies and effects. They are whether, at the macro-level of national educational policies, national accountability systems vary in their effectiveness upon the 'outcomes' of education in the STEM area, and whether societies differ in how they define these outcomes, so that we in the UK can build better STEM policies ourselves.

Types of Accountability Systems

Systems of accountability are largely of three kinds (Anderson, 2005):

- 'Regulatory' accountability to rules, regulations etc. (what can be called 'administrative accountability');
- Adherence to 'professional' norms, values and standards ('professional accountability');
- 'Results driven' accountability related to the effectiveness and efficiency of the educational system in generating desired outcomes ('performance' or 'outcome accountability').

The first system demands compliance with rules, regulations and laws – current procedures about child protection are an example of this, as are the current Inspection Frameworks (albeit differently administered in England, Wales, Scotland and Northern Ireland). The second system attempts to use accountability to specify the professional conduct, behaviour and processes expected from groups of professionals – the 'Association of Educational Psychologists Code of Conduct' is an example of this kind of accountability. And the third system attempts to take performance data about how well educational systems and schools individually have been doing into the society outside schools and colleges in order to give information to government, parents and the general public, which they will then use to inform their decisions about education (individually and through the compliance mechanisms of the educational 'State').

The direction of travel over the last 30 years, internationally, has been strongly of the third type - 'performance accountability' - using the 'demand' side of education as a lever of improvement through parents choosing what they see as more appropriate schools for their children, and monitoring the performance of these schools and the education system more generally. Performance

accountability is also linked to the supply side of education directly – to the national and local 'states' that run public education. This model is therefore one of simultaneous 'supply-side/demand-side' reform, and of 'supply-side/demand-side' accountability.

The Rise of the 'Supply Side/Demand Side' Accountability Paradigm

A contemporary educationalist venturing back to look in the 1950s and 1960s at the United Kingdom education system would find the landscape very different. Education was, in the phrase of the time, 'a national system locally delivered', but in truth there was very limited national involvement. Overall policies emanating from the Ministry of Education and from government attempted to steer the system into forms of organisational change, such as in Circular 10/65 that 'requested' local authorities to submit plans to make their schools 'all through' comprehensives. There was central involvement in curriculum matters through the operation of 'arm's length' quangos, like the Schools Council, and there was the operation of Her Majesty's Inspectorate, which inspected individual schools and generated national overviews of educational issues as and when required. And, of course, individual schools had governors who provided some local accountability to parents, and Local Education Authorities (LEAs) which provided a degree of local pressure, support and accountability (c.f. Lingard & Ozga, 2007; Tomlinson, 2005; Chitty, 2009; Ball, 2008). In reality, the system was one of multiple autonomies: of the educational system from national steering and recognition of national needs; of LEAs from national government; of schools from their LEAs and from parental oversight; and of subject departments and individual teachers in schools from peers.

There was of course 'regulatory accountability', but accountability was largely one of educational professionals to themselves, although the historical tendency of the teacher unions to concentrate more on the protection of their members than to operate like other professions to improve standards of professional practice meant that 'self-regulation' or 'teacher professionalism' was *not* well developed in the United Kingdom and internationally. This was the era of an educational 'club culture'.

In such a system, it was not surprising therefore that major programmes of educational reform found it hard to impact and to 'root'. Many societies across the world shared an enthusiasm for education in the 1960s, and greatly increased per capita expenditure, attempted school level organisational reforms and lengthened the school year, but the results disappointed both traditional governments and liberal reformers, as did the attempted modernisation of the

curriculum in the UK in the 1960s to one with more 'discovery' orientated (Nuffield Science and Mathematics) emphases. The widespread cry emerged subsequently that 'education didn't matter', that 'education could not compensate for society' (Bernstein, 1968) and that 'schools made no difference', conclusions supported by the very first studies in the newly emerging field of school effectiveness (e.g. Coleman et al, 1966; Jenks et al, 1972) in the United States. However, pressures for a further intensification of educational changes grew internationally in the 1970s. There was in the United Kingdom, for example:

- Attention given to the economic achievements of our European competitors, such as Germany, who had 'staying-on' rates considerably higher than our own (Comber & Keeves, 1973) and a higher participation rate in the STEM area;
- Increasing recognition of the existence of variation in school quality shown by the early and much publicised UK school effectiveness studies (Power, 1967, 1972; Reynolds, 1976);
- Some 'scandals' of particularly poor performance in some schools, such as that of the William Tyndale Primary School in London (Auld, 1976);
- Some research studies that appeared to discredit many of the educational movements that had so appealed to educational professionals in the 1960s and 1970s, such as the so-called 'progressive' or 'pupil centred' teaching methods (Bennett, 1976). Interestingly, it was precisely the high degree of professional control and autonomy that had made it possible for these innovations to be promoted, since professional enthusiasm for them had widely outrun public and governmental enthusiasm.

By the mid to late 1970s there were multiple arguments to make the educational system more productive, and that this should be done through increasing its accountability and that of its schools and professionals to the State, parents and to society in general. In a landmark intervention in 1976, the then Prime Minister, James Callaghan, called for a debate about the need for government to become involved in the 'secret garden of the curriculum', and in the 1979 General Election campaign the Conservative Party famously used the slogan, 'Edukashun isn't working' in an attempt to tap widespread popular discontent with the operation of the system.

The late 1980s saw the rising popularity of 'market based' or 'demand side' solutions to our perceived educational problems, based upon the popularity of ideas (associated with the Institute for Economic Affairs) that education should be distributed to populations in the same way as other economic goods. In the Education Act of 1980 and the Education Reform Act of 1989, schools

were required to report their assessment results to Government and to their parental consumers, and parents were appointed as of right, not as previously as custom and practice, to all school Governing Bodies. The number of school 'types' was also increased, with the introduction of City Technical Colleges for example, and initiatives like the Assisted Places scheme, to increase the capacity of the system for greater parental/consumer choice. A National Curriculum was introduced for all primary and secondary schools, designed to minimise the variation in curriculum offerings (both in the range of subjects and in the time devoted to each of them) that had been a feature of 1960s and 1970s education.

The great majority of these reforms could be characterised as operating on the 'demand side', aimed at levering up educational quality by 'pressuring' the school system and its professionals to be more effective, and by offering financial inducements so that more students brought more money to successful schools. 'Supply side' reforms aimed at improving education through State intervention/accountability were not as marked a feature of these years, although clearly government attempted to generate change here too. For example, Her Majesty's Inspectorate was incorporated into the new organisation entitled 'The Office for Standards in Education' (OFSTED) and the inspection regime also changed.

By the mid-1990s, it seemed that the operation of the mostly 'demand side' policies, particularly combined with a weakened 'supply side' through a reduction in State educational expenditure from 5.5% to 4.8% of GDP by 1997, had failed. International surveys (Harris, Keys & Fernandez, 1997) showed a poor performance in the UK by comparison for example with the countries of the Pacific Rim, and interest in the characteristics of the teaching methods in countries such as Germany (Prais, 1995), Switzerland (Burghes & Blum, 1995) and Taiwan (Reynolds & Farrell, 1996) greatly increased, a forerunner of the huge contemporary interest in 'international good practice' that marks the present decade.

The incoming New Labour Government of 1997, and indeed many other administrations internationally, went for what was increasingly called 'simultaneous demand side/supply side' reform, attempting to both make the system more accountable to the educational State, who would also be involved in improving the 'supply' of it, and to the 'demands' of parents. Labour enhanced 'demand side' accountability – with value added, more 'true' results of school performance both nationally and locally published, the generation of more 'types' of schools such as new City Academies for parents to choose from, more power

being given to (parent dominated) School Governors and also requirements for schools to have 'contracts' with their educational consumers/parents. Labour also generated major 'supply side' reforms with the often labelled 'prescriptive' introduction of national programmes to upskill teachers, such as through the Literacy and Numeracy Strategies, and through programmes of Headteacher training organised through the newly established National College for School Leadership (NCSL). Whilst Labour also set up the General Teaching Council to provide an organisational focus for the profession, and to 'police' it in terms of ensuring teachers' adherence to professional standards (although the rhetoric of 'informed professionalism' was frequently used to characterise policies, from 2002/3), little was done to improve professional 'capacity' to handle the simultaneous 'demand side/supply side' pressures. Improving teachers' CPD, for example, was not a policy focus until the mid-2000s.

Internationally, the similar simultaneous 'demand side/supply side' accountability and reform package or suite of policies were tried in many societies, as for example in the well-known at the time Chicago 'performance based' suite of educational reforms. Leithwood et al (1999) summarised nicely the characteristics of this 'performance based' reform:

- A centrally determined vision and explicit goals for student performance;
- Curriculum frameworks explicitly targeted at these goals;
- Standards for judging the degree of success of all students;
- Coherent policies that attempt to meet the standards;
- A system of resource allocation and governance that devolves to 'school' or 'State' level responsibility for improving school performance;
- A central authority that reviews information on organisational performance and distributes rewards and sanctions that have significant consequences to those organisations attempting to meet the standards.

The UK 'New Labour' supply and demand-side programmes, and their proponents like Michael Barber (2007), became something of a global model for international educational change and accountability, as Sweden with its high performing comprehensive schools had been in the 1960s and Finland is currently, but there was disappointment that more had not come from the simultaneous 'supply side/demand side' accountability mantra in evidence by the early to mid-2000s, in the UK and internationally. Attention within many societies began to shift to the improvement of the teaching profession itself, which was increasingly seen as not possessing the 'capacity' to improve standards in the new administrative landscape that educational reforms had created. And the growing popularity of international surveys of educational

achievement, organised by the OECD and entitled the Programme for International Student Assessment (PISA), meant that there was now an additional 'beginning' knowledge base about school accountability, organisation and student outcomes in a large number of countries (60 in the 2010 PISA) and a knowledge base that was systematically expanding in the areas of education that it looked at, with PISA now focussing extensively on the governance and accountability systems at national level within its participating societies.

The momentum behind an enhanced focus upon building the capacity of 'the teaching profession' as an added focus of educational reform had increased by the late 2000s because of the often repeated argument that the previous levers of 'supply side/demand side' reforms had run their course. Wave One – the international trend to self-managing schools - had largely run its course by the mid-1990s. Wave Two - the encouragement of schools to get the school effectiveness correlates into place - had probably done the same by the early 2000s. Wave Three was the 'supply/demand' side reform which itself also disappointed by the years of the late 2000s, particularly in the UK where national UK performance on the triennial PISA surveys declined through the 2000s by comparison with many other societies.

Of course, many societies were able to show rapid gains in achievement levels at particular times, but not consistently over time – as in the case of England with its National Numeracy and Literacy Strategies in the 2000s, for example. There were marked improvements in the attainment of primary school pupils (in England) associated with these – by 2007 80% of pupils achieved Level 4+ for English in the Key Stage Two Standard Attainment Tests (SATs) at age 11, up by 17% from 1997. The comparable figures for Mathematics were 77% (up by 15%) and for Science 88% (up by 19%), (Sammons, 2008). But these overall improvements had phases of very rapid growth of scores (1998 to 2002), and times when scores improved more slowly and indeed largely 'plateaued'. Achieving consistently reliable, powerful change was difficult, even with policies that 'pulled' all policy levers.

Increasingly, also, the intellectual and ideological foundations of the 'supply/demand' side reforms themselves were threatened. The need for a strong educational 'State' to both pressure and support schools became more difficult to justify in the years of retrenchment in public expenditure that followed the 2007 banking crisis internationally. The increasing micro-management of schools in many parts of the world by central governments, as in the UK, was increasingly seen as generating dependency and passivity, rather than the ownership and innovation by individual schools that a rapidly changing

society needed. Enhanced accountability – the 'mantra' educational concept of the last 30 years – was increasingly seen as having a 'downside' in the precise formulations in which it had been operationalized.

From the Conservative/Liberal Democrat Coalition government elected in 2010 also came the acknowledgement that teachers and teaching should be the key areas to impact upon, but in their case the proposed solution was to argue for a further intensification of the demand side and a reduction of Labour's supply side emphases, as the proposed solution to the disappointing performance over time in the PISA surveys of performance for England in the 2000s.

Building a New Supply-Side/Demand-Side Accountability Paradigm

For a variety of reasons, then, governments and educationalists in many parts of the world have alighted upon 'the educational professions' as the key focus for educational improvement. The 'supply side' reforms are seen as pushing down from the 'top' as it were – with both pressure and offered support. This generates the 'upwards' accountability of schools to the Central and local States. From the 'bottom' comes pressure from the 'demand side' of parents, students and other educational 'consumers' – which creates 'downwards' accountability. In the middle is the educational system and professionals that are caught in the middle of the 'sandwich' as it were, and which need interventions to help them cope.

However, there is no suggestion by us here that the existing conventional accountability systems can or should be abolished. They came into being specifically because the political system judged that the educational system being left to self-regulate and operate free from formal accountability to society, parents and the needs of the economy, was generative of both low educational standards and a slow rate of educational change in a world where all other changes are now rapid and pervasive. These systems came into being because of a specific absence of 'trust' by stakeholders of the educational profession, which was seen as having a 'club culture'.

Additionally, the factors that created this climate of opinion generally in society are still operative – the existence of national concern about professional standards across all and multiple professions, the enhanced use of user or consumer power and a more demanding clientele for state welfare, the transformative nature of IT to 'open up' communication that was formerly closed off to all but a privileged few and the weakening of support for State

welfare amongst some sections of society. As O'Neill (2013) notes, trust between citizens and professionals seems to have incurably broken down.

The rapid internationalisation of the global economy, the increased mobility of modern capital and labour, and the ease with which ideas and practices now traverse the planet in search of locations, wherever they are, to 'root', mean that governments in the UK and elsewhere are clearly bound to see a need to continue to operate a range of accountability policies, because of the perceived economic threats to their own countries if they do not.

So, if it is true that the 'Waves' of reform and indeed the existing 'supply/demand' side accountability paradigm has been rightly seen as somewhat lacking in its effectiveness, the next stage of educational reform is seen by many as potentiating the effectiveness and efficiency of the professionals to cope with what is expected of them as the accountability policies continue to operate. The logic also lies in an attempt to improve the existing accountability paradigm so that in itself it is more efficient and effective.

Research on Accountability

One further factor might incline us to want to persevere with the accountability reforms of the 1990s and 2000s internationally, and that is if the research literature confirms that they have had positive effects on student outcomes. There is very little literature that meets conventional criteria of scholarly rigour here, whereas there is much that shows opinions being put forward without regard for evidence. But we have been able to find a number of studies that are relevant.

1) Does Enhanced Accountability Generally Lead to Better Performance – International Research

Given that the range of policies about the accountability of educational systems and schools will obviously be considerably larger internationally from all the countries of the world than nationally from one country, it would seem sensible to establish the possible 'power' or lack of power of accountability systems in this area of research first.

However, there are multiple problems involved in drawing conclusions from any of this internationally based work (Fauberty, 2009). Firstly, the great majority of the studies conducted to date present performance data about the performance of varied samples of countries, but cannot adequately sort out the reasons for any possible differences between countries, This is because they

utilise 'snapshot' data on the performance of students cross sectionally at a point in time, which could be the result of the influence of national culture, national levels of socio-economic development and of course the national educational structure, rather than just being the results of educational structures and the ways in which they are organised in their accountabilities. We need 'value added' data which strips out the effects of the backgrounds of students from the 'quality' of their schools, their educational policies and their systems, with data collected at a defined start point and at a defined finish point to look at 'value added' over time after the influence of non-educational factors has been stripped out. This we don't have in any large scale studies, and only in one small scale one (Reynolds et al, 2002).

Our methodological problems are intensified in that the great majority of studies in this 'international' area collect no data systematically about educational matters such as the nature and organisation of accountability systems. The well-known IEA studies (see Reynolds et al, 2002) have extensive student-level achievement data on multiple countries and have some data on a limited range of school-level organisational factors, but do not collect data on any supra-school organisational features of schooling, nor on national/regional systems of administration, management and accountability. Issues concerning the difficulty of getting reliability in judgements about system factors cross-culturally, together with the costs of attempting to do this, are probably the explanation here.

The only international dataset that systematically does collect 'supra-school' data on national systems of education and their organisation is that of PISA, but again these data are not 'value added' so we do not know whether it is national culture or national educational structure that generates any national student outcome differences. However, PISA has in fairness in recent years begun to use 'national income per head of population' as a surrogate/proxy for 'non-educational' factors. It has also, in recent years, put a considerable effort into data collection at what can be called the 'national policy' level and more recently through the TALIS studies into professional/teacher factors and their importance.

PISA (OECD, 2010), suggests that where autonomy and accountability are, in their words, 'intelligently combined', they tend to be associated with better performance. In detail, at a country level, the higher the proportion of schools that has responsibility to define and elaborate their curricula and assessments, as in the UK, the better the performance of the entire school system, even after accounting for national income as a surrogate for student non-educational

factors. Additionally, PISA 2007 found that the publication of the student level achievement data had a statistically significant positive effect upon performance even after demographic socio economic characteristics had been accounted for. This effect was 3.5 scale points on the PISA Science scale, in a comparison of schools that published achievement data with those that did not (OECD, 2007).

It is important to note that the limited amount of international research available does not separate out the STEM subject areas from others. Although PISA measures Reading, Mathematics and Science, and although currently performance on these measures has a degree of independence (so England does much better relatively on Science for example), the analyses are of 'overall' country performance in terms of the educational system/policy/accountability factors associated with high performance.

2) Does Enhanced Accountability Generally Lead to Better Performance? Findings from National Research

Studies with research from individual countries avoid many of the problems of drawing conclusions from the very limited international datasets. Every nation has, to a greater or lesser extent, a national culture shared by all its citizens, so any internal difference or change in educational outcomes may be seen as reflecting the influence of 'structure' rather than 'culture'. However, of course, at any given point in time most nations exhibit very similar educational policies across their component regions, precisely because most country policies in education have national reach. Some societies have a degree of self-determination regarding educational policies by individual provinces (like the German Lander) or by individual countries within their nation states (such as England, Wales, Scotland and Northern Ireland), but the possible effects of substantial internal policy variation within individual nations and its effects has been rarely studied.

The United States, with its decentralised, State-based system of education, provides us with possibilities for study here, particularly since it has national systems of assessment of student achievement through the National Assessment of Educational Progress (NAEP) testing of fourth-graders in Reading and Maths utilised by all American States, and also because American States have customarily exhibited variable accountability practices and other policies. The 'No Child Left Behind' Act (NCLB), passed in 2001, requires States to have accountability systems which typically involve State-wide testing for grades 3 to 8, the disaggregated reporting of data on student performance and

the employing of sanctions when student performance is poor. State by State analysis (Hanushek & Raymond. 2005), shows that the introduction of accountability systems leads to higher achievement growth than would have been expected without accountability, controlling out the background effects of racial composition, students' family characteristics and other State-level policies. This effect is 0.2 of a (using individual State data) standard deviation on test scores. But, critically, the effect is only significant for accountability systems where there are direct consequences, such as financial incentives and disincentives for teachers, which follow on from any poor performance.

Dee and Jacob (2009, 2011) also use the NCLB policy change to study the effects of accountability and report a 0.5 pupil standard deviation impact on 4th grade Maths scores, but no impact upon Reading scores. Given that Maths is a much more 'school/teaching dependent' outcome than reading, this result would be supportive of the Hanushek thesis of accountability systems being positive in their effects.

The United Kingdom, with its four countries, also provides an opportunity to study the effects of variation in national accountability policies over time, particularly in view of the very radical departure of Wales from historical joint Wales/England policies post devolution in 1999. Wales ceased publishing national performance tables for secondary schools in 2001, and deliberately made it impossible for these to be generated by any organisations like the media that might have wanted to do so (this is before the operation of the Freedom of Information Act of course).

Burgess, Wilson and Worth (2010), report on a comparison of the GCSE results obtained by children in Wales and in England using two samples – of all secondary schools on their mean GCSE points score/proportion of pupils achieving 5 or more GCSEs at better than C grade, and of a group of schools matched on the basis of socio economic factors. Using the individual school level Key Stage Three data as a proxy for prior achievement and using England as a 'control', the 'experimental' Wales policy change of abolishing the publication of secondary school league tables markedly reduces school effectiveness in Wales, with a fall of 1.92 GCSE grades per student per year, equivalent to 0.23 of a (school level) standard deviation. The percentage of students achieving at least 5 good GCSE grades falls by 3.4 percentage points as cohorts of children go through schools that are no longer measured in their outputs nationally (these figures are calculated comparatively with the England data).

A final American study (Ladd, 1999), involves a case study of the Dallas accountability programme on student outcomes, as measured by the pass rates over time on the State criterion referenced tests, controlled for the influence of background factors. This was used to award financial incentives to school staffs to increase student learning, and showed positive and relatively large effects for Hispanic and white seventh grade children, but not for black students. Other potentially positive changes included a fall in the drop-out rate relative to that of other similar cities, looking at the whole student group together.

3) Does Enhanced Accountability Generally Lead to Better Performance? - Specific Aspects

Further specific aspects of the historic accountability paradigm include evaluation systems to measure the performance of individual schools that have been given the enhanced power to 'self-manage' within the increasingly decentralised education systems of the world. Here, as with the creation of accountability systems in general, it is difficult to evaluate the impact of policies – 'before' and 'after' studies conflate temporal changes with the policy changes being studied. Within countries, when policies for individual schools are generated nationally there is no 'control' group to compare with any 'experimental' group, no 'counter factual'.

Nevertheless, there are some studies that seem to have been well enough designed to permit some tentative conclusions:

- Accountability systems' threats of sanctions, and indeed the sanctions themselves, can raise student test scores in the short term. A number of research studies on the well-known Florida incentives scheme in the United States show raised test scores from 'under threat' schools (Figlio & Rouse, 2006). Chicago schools 'under threat' raised test scores also (Jacob, 2005). Chiang (2009), provides evidence that sanctions threats have persistent, positive effects upon student test scores even after the affected children progressed years later into middle school, using a regression discontinuity research design. Surveys (Jacob & Levitt, 2003) also show changes in school processes such as enhanced professional development and greater planning time for teachers, and consequent budgetary changes from reduce expenditures upon curriculum areas not related closely to the State testing, such as Fine Arts. The extent to which these positive changes to accountability systems reflect 'gaming' by the schools such as focussing upon past test papers, teaching to the test and exclusion of potentially low scoring pupils is unclear.

Interestingly, other research in Chicago (O'Day, 2002) found that among low performing schools placed on probation, those that historically had developed strong cultures of peer collaboration were able to exit from probationary status more rapidly than others.

- There is some evidence that the above 'accountability systems' may impact differently upon the student peer group, with hints from the United States that minimum competency testing associated with the NCLB may lead to school concentration upon marginal or borderline students to get them 'over the hurdle', with negative effects upon more able students (Reback, 2008). This has also been widely alleged in English educational policy discussions recently, related to the '5 or more good GCSE grades' historic hurdle, and to the grade boundaries in primary schools.

- There is also some evidence that external inspection of schools may also be associated with higher performance. Rosenthal (2004) found no evidence that a visit from OFSTED, the English schools inspectorate, was associated with beneficial results on student achievement. Similar results came from Shaw et al (2003), who reported that in comprehensive schools (90% of the sample of schools), inspection did not improve achievement, although in selective schools there was a slight improvement. However, on the other hand Matthews & Sammons (2005) noted that 85% of schools placed in 'special measures' were out from them within two years and OFSTED itself (2008) notes that this was because such schools were spurred into adoption of what can be called 'school effectiveness correlates'. The international evidence is even more extensive and favourable to inspection than the English evidence. DeWolf & Janssens (2007) report from an extensive international review a positive effect upon school quality and outcomes. In a Dutch study (Luginbuhl et al, 2007) student performance increased by 2 to 3% of a test score standard deviation, with gains strongest for primary school Maths performance, and with gains persisting for four years after inspection. The more 'intensive' the inspection process the higher was the gain.

  However, we should note four problematic issues concerning accountability orientated reforms. Firstly, a much quoted study from Ontario in Canada (Leithwood et al, 2002), argues that teachers who are constrained by external accountability systems may be less effective in teaching their students because their own intrinsic motivation to teach may have been adversely affected by external control. The circumstances of the study – which followed student testing programmes and the release of individually named school data to the

press – suggest that teachers may see government policies as unrelated to the improvement of teaching and learning and their own professional goals, and that the communication of the rationale, purposes and potential of accountability orientated reform is essential to maximise its effects. Governments were in fact seen as last in the list of 'reinforcers' that teachers paid attention to, after parents and students.

Secondly, the more accountability pressures are put upon schools the more that things may happen entirely unintended by policymakers (Looney, 2011). Schools being pressurised may manipulate testing conditions (Jacob, Courant & Ludwig, 2003), and teaching to the test may occur. Certain students may be selected for 'special attention'. Test items and the curricula relevant to them may be given special attention within instruction/teaching (Linn, 2005). There may be direct 'coaching' (Koretz, 2005; Popham, 2002).

Thirdly, it is possible that some of the mechanisms customarily adopted for 'accountability' purposes may have 'surface' rather than 'deep' effects. Jurges et al, (2012) used data from a German PISA study to look at the effects of central 'exit' examinations on student Maths performance, literacy in Mathematics and student attitudes. Using a value added measure to pin down the effect of these exams, they found that they improved curriculum based knowledge, but did not affect mathematical literacy in general. As a consequence students apparently performed better but in fact were less intrinsically motivated in school.

Fourthly, we need to remember that accountability systems are <u>second order</u> ways of using material about how <u>first order</u> tasks are carried out. They are not themselves the first order factors, and should not be confused with them. This recognition needs to be remembered by policymakers and politicians who appear to have become so transfixed by the second order that they miss the ways in which the continued use of the second order may diminish what is achievable by the first order.

We should conclude this section by noting that all the above research into 'national' accountability systems and into specific aspects of accountability systems does not look separately by subject at effects, and therefore cannot outline possible STEM specific accountability factors.

<u>What is Success in the STEM Area and What are the Contemporary Developments in STEM Literacies?</u>

We move on now to consider definitions of 'successful' STEM outcomes and STEM literacies within a simultaneous supply side/demand side accountability system, and at how these definitions have been changing in the UK and in other countries. We will use the subject areas of Mathematics, Physics, Chemistry, Biology and Computing as exemplars, because within the STEM area as it affects (and is affected by) accountability and educational effectiveness research, 'Technology', 'Design Technology', 'Engineering' and related vocational subjects have *not* been used in mainstream school and system effectiveness assessment. This may be due to their optional/elective status, relatively low uptake in schools, and their perceived low status and (anecdotally) low-frequency usage for HE admission purposes. From necessity then, our report focuses mainly on the 'Science' and 'Mathematics' elements within the STEM family.

1. Mathematical Literacy

The Ofsted view on defining effective Mathematics education (Ofsted, 2012) is fairly generic. It could easily be applied to any subject and if all reference to Mathematics were removed from the guidelines it would be difficult for the unfamiliar reader to infer which subject was under consideration: it has a clear focus on thinking and understanding (for example, in developing primary pupils' grasp of the effect of multiplying or dividing by ten as part of the development of their conceptual understanding of number structure); that effective whole-class teaching should be dynamic with a large element or peer (pupil-to-pupil) collaborative working; that pupils should be challenged to think for themselves in tackling new problems with a range of approaches; on the instruction side, that effective mathematics teaching should be characterised by teachers keeping explanations brief and to the point, focused on the underlying concepts; that teachers should show how the work at hand links with previous topics, should use appropriate questioning of pupils designed to encouraged reasoned responses, and where appropriate, provide pupils with an efficient standard problem-solving method; that pupils should engage directly in mathematics for a substantial portion of each lesson and as a result, have time to develop a high degree of competence in tackling challenging, varied questions and problems designed to deepen their understanding. What is perhaps less generic and more specific to Mathematics (and STEM subjects more generally) is the facility for pupils to work as a group on a mix of exploratory tasks attempting to devise their own methods as they progress from routine practice of skills to two-step questions, where the method is not immediately apparent and questions have

unusual twists that require the adaptation of the standard approach (Watson et al., 2013).

The school inspection framework is widely thought (by teachers) to be a counter-influence to group-working and risk-friendly demanding classroom approaches, one of the practical benefits of which is to provide teachers – specifically Mathematics teachers - with the opportunity to move around the classroom and assess individual pupil progress while pupils work on staged problems; the most effective teachers being able to adjust their lesson plans, and individual levels of pupil support and challenge, to reflect the progress of their charges. Perhaps most importantly in terms of risk and most specifically in terms of Mathematics/STEM subjects, for both pupils and teachers, is the ability of good Mathematics teachers and pupils to 'welcome' wrong answers as an opportunity to explore misconception and/or badly explicated instruction, and to firm-up correct understandings in the minds of pupils. Mathematics and STEM subjects more generally, are perhaps more privileged in this facility as there *is* a 'right answer' and as a consequence greater opportunity for teachers and learners alike to use error as a learning tactic. Indeed, anecdotally, the best teachers plan deliberately to elicit occasional wrong answers and tactically to progress pupils' understanding through *mis*understanding. It is a feature that is not catered for in accountability and inspection regimes.

Interestingly, the term 'mathematical literacy' does not appear in DfE policy documents. Instead the term 'numeracy' is used, defined by the National Numeracy Strategy (DfEE, 1999: p.1) as 'a key life skill' without which '*children will be disadvantaged throughout life*' and which in essence is understood by policymakers as a '*proficiency involving confidence and competence with numbers and measures … [requiring] … a repertoire of computational skills and an inclination and ability to solve number problems in a variety of contexts*' (p.4). This can be extended easily to defining 'mathematical literacy' as the ability to understand and engage with numbers and measures, to think critically about, and reason using, mathematical concepts, and to have the intellectual facility to pursue further interest if so motivated.

In 2002, OECD defined mathematical literacy for their PISA studies as '*an individual's capacity to identify and understand the role that mathematics plays in the world … and to use and engage with mathematics in ways that meet the needs of that individual's life as a constructive, concerned and reflective citizen*", but by the 2012 PISA a decade later, the definition had changed to '*an individual's capacity to formulate, employ, and interpret mathematics in a variety of contexts … [including] … using mathematical concepts … and tools to*

describe, explain, and predict phenomena … [and recognising] … the role that mathematics plays in the world and to make the well-founded judgments and decisions needed by constructive, engaged and reflective citizens." (OECD, 2012 and OECD, 2010: p.4)

Looking outside the UK and Europe to useful comparator nations, the US has also engaged with the issue of definition. In 1998, the US National Science Foundation and the US Department of Education asked the National Academy of Sciences to synthesize the research on 'pre-kindergarten through eighth-grade' mathematics learning and provide recommendations for best practice (Kilpatrick, 2001a). The report itself, *Adding It Up: Helping Children Learn Mathematics*, was published later in the year (Kilpatrick et al., 2001b), and addressed the issue of defining 'successful mathematics learning' while recognising the need 'to characterize such learning' by defining several relevant terms: mathematical literacy, numeracy, mastery, and competence. Kilpatrick (2001b) recalls that each of the terms 'captured to some extent what it means to learn mathematics successfully', but none of them seemed suitable (see Ahmed et al., 2001). They were either too unfamiliar, too limited in scope, or had acquired undesirable connotations. The committee eventually settled on '*mathematical proficiency*', defining it in terms of five interwoven strands to be developed 'in concert': (a) *conceptual understanding* - a student's comprehension of mathematical concepts, operations and relations; (b) *procedural fluency* - a student's skill in carrying out mathematical procedures flexibly, accurately, efficiently and appropriately; (c) *strategic competence* – a student's ability to formulate, represent and solve mathematical problems; (d) *adaptive reasoning* - the capacity for logical thought and for reflection on, explanation of, and justification of mathematical arguments; and (e) *productive disposition* - the student's habitual inclination to see mathematics as a useful and worthwhile subject to be learned, and a belief in the value of diligent work and in one's own efficacy as a doer of mathematics. In the US context, mathematical proficiency has been, and is, used to define learning goals for students at every age and grade; and indeed the notion of 'proficiency', as opposed to 'literacy', as intertwined strands can also be used to define what it means to be a proficient mathematics *teacher*.

In South Africa, 'mathematical literacy' is defined in terms of what it *does*, rather than what it *is*, and the country is unusual in the extent to which it focuses on the utility value of the subject:

> "*Mathematical Literacy provides learners with an awareness and understanding of the role that mathematics plays in the modern world …*

[and] … is driven by life-related applications …[enabling] learners to develop the ability and confidence to think numerically and spatially in order to interpret and critically analyse everyday situations and to solve problems." (Department of Education RSA, 2003: 9).

Mathematical literacy outcomes include the ability of learners to: use mathematical skills to identify, pose and solve problems creatively; work collaboratively in groups to enhance understanding and use a variety of individual and co-operative strategies in learning mathematics; collect, analyse and organise quantitative data; communicate appropriately by using graphs, symbols, tables and diagrams; understand the interrelatedness of systems; be able to explore the importance of mathematics for career opportunities and contribute to entrepreneurial success. Other outcomes in relation to mathematical literacy relate (unsurprisingly) to wider societal issues in that country and include the ability to: organise, interpret and manage authentic activities in mathematical ways '*that demonstrate responsibility and sensitivity to the individual and society*'; ensure that science and technology '*are applied responsibly to the environment and to the health of others*'; '*engage responsibly with quantitative arguments relating to local, national and global issues*' (Department of Education RSA, 2003: 10). And like the US, we see how easily mathematical proficiencies and literacies could be used to establish and judge what it means to be a proficient mathematics *teacher*.

2. Scientific Literacy

Looking at general issues concerning scientific literacy first, Durant (1993), defined scientific literacy as what 'the general public ought to know' about science, and Jenkins (1994: 5345) regarded it as implying '*an appreciation of the nature, aims, and general limitations of science, coupled with some understanding of the more important scientific ideas*'. The term is now synonymous in the UK with 'the public understanding of science', though 'scientific literacy' itself is more widely used in the US and 'la culture scientifique' the preferred term in France (Laugksch, 2000). However, the concept of scientific literacy does not have the same meaning or significance in (non-PISA) Russia where people are expected to be 'culturally refined' ('*kulturniy*', as opposed to '*Nekulturniy*' or more commonly '*vulgarniy*') and science is part of the general culture in which young people are educated (Fullick, 2013). In the West, scientific literacy has become widely understood in terms of the 'science for citizenship' movement, and in Russia this has been adopted as part of pro-Western initiatives regarding the development of civil society, manifested by the general interest in Russia in programmes like Nuffield's

Science in Practice and the International Council for Science (ICSU) project *Teaching Ethical Aspects of Science*, but science as directly experienced by Russian students is still more directed to practical societal and socio-political ends than it is in the UK. Historically, in the Soviet era, this manifested itself in a triumphalist way through propaganda posters and memorials to scientists and their achievements, which still can be found everywhere in towns and cities, so the notion of teaching science to help young people engage in democratic debate is an important development. In STEM subjects, especially physics and mathematics, the level of attainment in schools can be very high. Those with teaching experience in Russia and former CIS states consistently report that secondary students are typically up to three years ahead of their European and US counterparts, though (typically) that gap has closed by the time it comes to employment, due in part at least to the better 'soft skills' of European and US students in teamwork and professional settings. Russian schools are aware of this issue and are addressing it, and outreach initiatives like those of the Skolkovo Foundation are an interesting model to look at in this regard. The Skolkovo Foundation, founded in 2010 by (then) President Medvedev, operates an Innovation Center for the development and commercialisation of advanced scientific technologies. It comprises the Skolkovo Institute of Science and Technology (SkTech), established in partnership with MIT, corporate R&D centres, business incubators, private seed and venture funds, and start-up companies, as well as residential and social facilities. It is effectively a small city with over 30,000 residents and employees and some 300 companies, governed by a special law (not unlike that for Hong Kong) which gives resident companies privileged economic conditions.

In Physics, the current philosophy regarding what 'good' physics teaching looks like is widely understood within the profession as moving students away '*from mindless memorisation to understanding and appreciation*' (Wieman and Perkins, 2005), but the unintended consequence of physics teachers being encouraged by accountability metrics to train students to answer examination questions by simply 'plugging in numbers' into relevant formulae has resulted in a lack of conceptual depth among physics teachers (and perhaps among physics graduates), as a consequence of which most teacher training institutions now oblige even physics graduates to attend the 'KS3 Physics' part of PGCE courses.

Of course, what is taught in school physics has changed dramatically over recent years and there has been considerable churn in the system. In the UK, the introduction of the National Curriculum in 1989, and 'Single' and 'Double' Award Science, has resulted in a reduction in the core content of all three sciences. Further changes came following the '*Beyond 2000*' report by Millar

and Osborne (1998), which argued that the compulsory science curriculum should be designed to develop the scientific literacy of future citizens with a separate, parallel course for those who intended later to follow more advanced programmes (Garrett, 2013). This resulted in the development of the '*21st Century Science*' GCSE specification and the beginning of the '*How Science Works*' strand, which was then, in 2004, incorporated into the National Curriculum KS4 revisions. There was an accompanying expectation of scientific literacy around issues of understanding how science and scientists work in order to inform judgments about science-related issues affecting students' lives, drawing on the (1998) OECD-PISA definition of scientific literacy as the capacity to use scientific knowledge to 'help make decisions about the natural world and the changes made to it through human activity' (OECD, 1998).

Further changes in the GCSE science syllabuses in 2008 shifted the goalposts yet again. Following Ofqual concerns about the 2007 and 2008 Science GCSEs, new specifications were developed by the examination boards, rejected by Ofqual (2009, 2010) as '*not going far enough to address the examinations regulator's concerns over standards*' and sent back to them *'for more work'*. Today, in the opinion of many teachers, the resulting GCSE specifications and examinations are currently more 'fit for purpose' than was previously the case - they are certainly more difficult, with the written communication questions requiring skills at explaining / arguing a case, and they have more demanding and wider content – so it may be a step backwards if these are now made even more difficult in the latest round of GCSE revisions for introduction in 2015. As Ofqual (2013) pointed out to the Minister, if GCSEs '*are put under too much pressure, they will not reliably measure the knowledge and skills they are designed to assess*' (as was the case in 2012 with GCSE English).

Work on chemistry literacy (or 'chemical' literacy) is not as widespread as that on science literacy generally or literacy in other STEM subjects like ICT. The canon includes work by Holman (2002) and Atkins (2005), and notably by Shwartz et al. (2006), who investigated student chemistry literacy among (approx.) one thousand 10^{th}, 11^{th} and 12^{th} grade students (i.e. in the first three years of High School) in Israel. The study used a four-way taxonomy of 'chemical literacy' in terms of students' ability to recognise a concept as being scientific but with some misconceptions ('nominal' literacy), their ability correctly to describe concepts, albeit with limited understanding ('functional' literacy), their ability to develop an understanding of major concepts ('conceptual' literacy), and their ability to understand chemistry at a philosophical and societal level, where they can develop an appreciation of it in relationship to daily life ('multidimensional' literacy). The research found that

chemistry students improve their nominal and functional literacy, but not necessarily their higher levels of literacy during these years at school.

The idea behind the Shwartz taxonomy is to allow scientific literacy (and its assessment) to drive pedagogy and curriculum (see also Bybee, 1997), but obviously students can have varying levels of literacy across the scientific domain and throughout a lifetime of learning; having a high level of literacy in one particular topic and/or at any given time, and a lower level in another topic and/or at another time. As Shwartz et al. (2006) point out, scientific literacy is not something you have or do not have. Whereas scientific *illiteracy* is easily defined as the lack of ability to relate or respond to scientific questions, scientific literacy occurs to various degrees, from the 'functional' literacy that enables practical consumer competence at an everyday level to higher levels of literacy including 'civic' literacy (the ability to participate in debate concerning scientific issues) and 'cultural' literacy (the appreciation of scientific research as an intellectual activity). Shamos (1989), suggests going further by differentiating between the passive memorisation of knowledge and the active communicating of that knowledge and scientific ideas.

Turning to Biology, when the OECD (1998) definition of scientific literacy, referred to above, was broadened in 2007 so that PISA thereafter adopted a three dimensional view of its assessment, it included 'science in life and health' and 'science of the earth and the environment' (OECD, 2007). In that sense, the biological sciences, as schools would understand the term, were put centre stage, and the fact of 'scientific situations' being selected from students' everyday lives rather than from the practice of science in the laboratory (Holbrook and Rannikmae, 2009) privileged the wider public aspects of 'green' bio-issues and health (Grace et al, 2012). The Society of Biology, formerly the Institute of Biology, comes across to teachers as more 'inclusive' in its overview than the other learned societies in respect of changes to the school curriculum; for example, it supports Triple Science *for all* students, including those who are not in the highest ability groups and those from more deprived backgrounds.

According to the Society, the best biology courses challenge students, encourage a passion for biology, support and nurture the next generation of scientists and provide careers guidance to students at all levels (Society of Biology, 2013). At university level, the society accredits good bio-education degrees, believing that the best 'have the potential to educate the life science leaders and innovators of the future' and 'address the skills gaps between undergraduate study and employment in the biosciences' by driving up 'standards of learning and teaching'. Biology, of course, deals with complex systems which are predictable to varying degrees, and students in later school years are

expected to appreciate a range of approaches to analysis (Grace, 2013), such as sampling, statistical analysis to ascertain levels of confidence in data obtained from experimentation, and modelling. Perhaps as a result, assessment criteria include the requirement to demonstrate an understanding of Physics, Chemistry and Mathematics, although there is no reciprocal acknowledgement in Mathematics, Computer Science or Physics as to the importance of Biology.

3. Computing Literacy

There remains considerable discontent regarding the state of computing in schools, the reasons for studying the subject – socio-economic, vocational, pedagogic, and other (Hawkridge, 1990) - and the training of teachers in the area. Many well-motivated organisations - Computing at School, the British Computer Society and its Academy of Computing, industrial partners like Google and Microsoft which align themselves with various curriculum proposals and counter-proposals, the DfE, the Royal Society, the Royal Academy of Engineering, Joint Information Systems Committee (JISC), - and others are inputting to the debate, so that the most obvious remediation would seem to be to have a single voice. The conflation of computer science with the uses to which computers are put in communication has also flooded the arena in a fashion that has encouraged the proliferation of gadgetry as a proxy for the intellectual development of the subject. The pressure group Computing at School (CAS, 2012; 2013) properly distinguishes, as the National Curriculum will from September 2014, between 'Computing' meaning the whole of the curriculum related to the use of computers, and its 'constituent parts' - 'Computer Science' (the primary curriculum component), 'Information Technology' and 'Digital Literacy', and 'Technology Enhanced Learning' in an extra-curricular sense – but conflates in the usual fashion STEM subjects so as to give the impression of a pedagogic coherence.

> "*Computer Science is a quintessential STEM discipline, sharing attributes with Engineering, Mathematics, Science, and Technology. … [It] provides pupils with insights into other STEM disciplines, and with skills and knowledge that can be applied to the solution of problems in those disciplines. … In a world where computer-based systems have become all pervasive, those individuals and societies that are best equipped to meet this challenge will have a competitive edge. The ability to bring this combination [of computational thinking, principles, and approaches to problem solving] to bear on practical problems is central to the success of science, engineering, business and commerce in the 21st century.*" (CAS, 2012: Section 1.2, p.4. emphasis added)

This is not an approach that has served the field well. Computing (and computer science) differs in huge respects from the other STEM subjects in terms of the speed of development of its leading edge, the extent to which teacher knowledge (both subject and pedagogic) lags behind that development, the ease with which the field can be made practical in schools, and its public media perception and familiarity.

'Technological/technical literacy' is a term rarely, if ever, used - 'digital literacy' is the nearest thing – but its desirable properties are usually defined *in relation to other disciplines* or *generically*:

> "*Computer Science [shares] attributes with Engineering, Mathematics, Science, and Technology: It has its own theoretical foundations and mathematical underpinnings, and involves the application of logic and reasoning. It embraces a scientific approach to measurement and experiment. It involves the design, construction, and testing of purposeful artefacts. It requires understanding, appreciation, and application of a wide range of technologies.*" (CAS, 2012: Section 1.2, p.4. emphasis added)

And there is little sense of what accountability metrics would be suitable for gauging whether, or to what extent, students had acquired such a literacy; or in the extent to which 'political' criticality is expected about the reliability of information sources (Buckingham, 2008). The section of the National Curriculum publication on 'Attainment Targets' comprises only two lines and those might refer to any subject:

> "*By the end of each key stage, pupils are expected to know, apply and understand the matters, skills and processes specified in the relevant programme of study.*" (DfE, 2013a:3)

Gilster (1997) defined digital literacy as literacy for the digital age – the acquisition of appropriate skills and core competencies to take advantage of the digital environment – but did not exclude traditional, non-digital resources, which implies (Paul, 2013) an encouraging willingness to 'build upon', rather than 'replace with', but there is little sense of uniqueness or importance, so that the desperate need of universities, businesses, industry, public authorities and the security services to interest young people capable of reading computer science at degree level (and beyond) remains largely unmet. The STEM approach of conjoining it with other related fields of activity has not worked. Nor has it

helped that Secretary of State Gove, however well-intentioned, is overseeing a system in turmoil, announcing on 11 January 2012 that Computer Science would be part of the new English Baccalaureate, announcing the following year (7 Feb 2013) that there would be no such qualification, highlighting several times the importance of Computer Science to a prosperous economy in the intervening months, and in 2013 announcing its disapplication so that during the suspension period, schools were '*not required to teach the centrally prescribed programmes of study or use attainment targets as part of statutory assessment arrangements*' (DfE, 2013b). The disapplication of ICT was, uniquely, accompanied by a critique:

> "*ICT as a subject name carries negative connotations of a dated and unchallenging curriculum that does not serve the needs and ambitions of pupils. Changing the subject name of ICT to computing will not only improve the status of the subject but also more accurately reflect the breadth of content included in the proposed new programmes of study.*" (ibid.)

… which echoed remarks made in the Royal Society report the previous year:

> "*Assessment methods that rely on taking screenshots to document coursework (rather than take advantage of more modern submission methods) are a contributor to the negative perception of ICT.*" (Royal Society, 2012: 11)

That report, '*Shut down or restart? The way forward for computing in UK schools*', has made a contribution to the debate, particularly in defining the component parts and briefly addressing the literacy issue:

> "*Computing [is] the broad subject area; roughly equivalent to what is called ICT in schools and IT in industry. ICT [is] the school subject defined in the current National Curriculum. Computer Science [is] the rigorous academic discipline, encompassing programming languages, data structures, algorithms, etc. Information Technology [is] the use of computers, in industry, commerce, the arts and elsewhere, including aspects of IT systems architecture, human factors, project management, etc. Digital literacy [is] the general ability to use computers. This will be written in lower case to emphasize that it is a set of skills rather than a subject in its own right.*" (Royal Society, 2012: 5 emphasis added)

Peyton Jones (2011) has produced an international comparison report on second-level computer science teaching in other countries, and argues that perhaps most importantly, the role of the *informal* curriculum, the impact of which is so familiar to schools, teachers and parents in other respects, has not been widely considered in the literature (grey or otherwise) in relation to computer science and digital literacy yet it is of massive significance; perhaps more so than any other subject area (and with commensurately more serious safety issues for young learners).

However, it is important to note that if we look at the comparative education literature, it seems likely that viewed in an international perspective the UK curriculum definitions of what constitutes the STEM subjects are remarkably similar to those of other countries that the UK wishes to emulate. The extensive literature review conducted for the '*Review of the National Curriculum in England*' (Department for Education, 2012), shows substantial similarity in the breadth of content in certain 'high achieving' countries (Australia, Finland, Flemish Belgium, Hong Kong, Massachusetts, New Zealand and Singapore) compared with England, and showed 'significant commonality in how jurisdictions organise their mathematics curriculum' (p. 64). In science 'despite variation in terms of structure and approach, curricula reviewed largely cover the same ground in terms of the key domains of biology, chemistry and physics' (p. 87). It is also argued '….all curricula reviewed emphasise the importance of scientific processes and scientific enquiry at both primary and secondary, and the coverage is broadly similar across the jurisdictions analysed' (p. 88).

Conclusions: Policy Changes Recommended Currently

We have ranged across time periods, disciplines and countries in our attempt to answer the question of how 'accountable' and to whom the UK educational system should be, and what 'success' definitions should be attached to the various STEM areas. The following recommendations are grouped into those connected with STEM-specific policies, those connected with further 'supply side' reform, those connected with further 'demand side' reform, and those related to the development of the professionals in education, who sit in the middle, as it were, of the 'supply side/demand side' sandwich.

A. Contemporary Additional STEM-Specific Policies

1. The grouping of science, mathematics and computer science as school subjects under the 'STEM' umbrella is probably unhelpful. It is a useful acronym

for both policymakers and academics, who find it convenient to conflate the fields so that it is assumed to refer to everything that is prosperity-seeking and forward-looking, but in educative terms it is counterproductive in terms of the surefootedness that is so badly needed in schooling and for accountability purposes. Generally, mathematics, which in any case is antecedent to the others, is at the 'slow' end of the spectrum, with Computing (science / technology) at the 'fast' end, but the contrast with mathematics is stark in terms of how 'literacy' is defined and actualised, and how closely it is perceived to be linked to economic well-being and individual prosperity.

2. School leaders in both primary and secondary sectors are incentivised very strongly to improve results in Mathematics especially, and are very much influenced in turn by the public nature of results and how these are used in the marketplace in terms of recruitment and retention of pupils and staff. The profile of secondary school Mathematics has been increasing in this manner since 2008 when performance was first measured against statutory targets for the proportion of pupils gaining the equivalent of at least five GCSEs at grade C or better *including English and Mathematics*. Prior to that, the metric did not include these two core subjects as an obligation. At the meso level, schools learned to 'play the measurement game', just as teachers learned to 'teach to the test' at the micro level and systems learned to manage policy at the macro level in relation to PISA; and when the 'game' changed some schools found that they needed to pay more attention to Mathematics (and English). In many cases, this worked against ICT / computer science subjects which were perceived to be 'softer' options for weaker students, but the effect was not uniform across the school and college sector. For example, there was a more dampened effect in the fee-paying sector, where the priority was (and is) less the immediate gain in GCSE league table position as the perception of aspirant parents with regard to preparation for university entrance, so that the leading edge (i.e. the more academically selective fee-paying schools) turned towards IGCSE and/or took GCSE mathematics a year early, at age 15, thus disadvantaging themselves in GCSE tables but catering better to their own niche market. In this respect, we would offer a more nuanced perspective to that of the Advisory Committee on Mathematics Education (ACME) report '*Early and Multiple Entry to GCSE Mathematics*' (2011), which claims that '*the practice of early entry has a negative effect on most students' mathematical education, hindering their progression to a wide range of subjects post-16 and in Higher Education'* while at the same time claiming that the '*focus on league tables is skewing behaviour … to the extent that school leaders are placing the interests of their schools above those of the students*'. Actually the early entry of examination candidates can *dis*advantage schools in terms of their league table position since these

results are not included. In fact, in the selective fee-paying sector, early entry is common with high-achieving pupils and can be a prelude to advanced study and to the study of STEM subjects at university.

Not surprisingly, when management is encouraged (or forced) to 'play the measurement game', triage is never far away, and it has been the case that many schools serving challenging catchment areas have concentrated on those pupils likely to gain five or more GCSE A* - C grades *but who are not yet secure in English or Mathematics*, rather than those who are already safely in the 5GCSE (E & M) category or those who are already very unlikely to make the cut. The system incentivises schools to devote a disproportionate effort/resource to those at the boundary of the metric in order to maximise institutional performance in league tables, rather than maximising individual performance at the top and bottom of the ability spectrum, so while Mathematics has benefited from its 'headlining' as a subject, which in all probability will continue to be the case (c.f. Ofqual, 2013) though perhaps not in the same way, it has probably suffered from a 'performance regression' wherein students with high ability in mathematics go relatively unchallenged and at the other end, basic mathematical literacy throughout society is lowered.

3. It may be the case that the assertion by the Secretary of State and others that the current ICT curriculum is *per se* seriously flawed may be mistaken: the current ICT curriculum *does* allow computer science and computational thinking to be taught, which is what the Minister wants. While welcoming the shift to a more rigorous scientific programme, it may be that the real problems lie elsewhere. What *has* been wrong is the *way* it was taught, what schools *chose to teach* (or not) from the curriculum, and the GCSE game-playing referred to above. The negative publicity around ICT and the sometimes misunderstood 'disapplication' manoeuvre are likely to carry over into computer science post-2014 (Woollard, 2013), and a watch will be needed to ensure that: (i) the accountability structure incentivises schools to 'do the science' - control technology, sensing, programming, etc - as Minister Gove and the interested learned societies intend; (ii) that Ofsted will be equipped to inspect properly the 'hard' areas of the curriculum and not just the literacy and numeracy aspects with which its inspectors are familiar; (iii) that Computer Science is delivered by specialist teachers properly trained in university departments, as in Germany (Peyton Jones, 2011); and (iv) that an infrastructure is created to support and train teachers to meet systemic demands in relation to PISA accountability, and cope effectively with a wider range of awards in state schools including the International GCSE and International Baccalaureate in Computer Science.

4. More attention needs to be paid to issues of safety, particularly in computer science. Recent well-publicised issues around pornography, cyber stalking, bullying, sexual intimidation and student suicide do not need further explication here, but there are issues for schools and teachers when 'learning safety' is pushed away from learning communities towards law enforcement agencies.

5. It is often forgotten in the rush to action in international comparison tables, that we can learn lessons from countries outside PISA. In Russia, for example, it is telling that students are typically up to three years ahead of their European and US counterparts while in secondary school, but that anecdotally, this knowledge gap has closed by the time it comes to entry into the labour market because of the value put on, and inclusion of, the soft skills (like team-working) in the West. Yet there are potentially positive lessons to be learned too, and the impact of facilities like the Skolkovo Foundation on education across the system, whether in China, Russia or the UK, should not be underestimated. For better or worse, their likely effect on accountability structures is to skew them towards economic outcomes in response to centrally imposed societal, rather than individual, imperatives.

6. Choice in examination questions in science subjects like Physics has narrowed over recent years in order to cover more areas. This has had the effect of encouraging teachers to teach more 'to the test', which suits less-qualified, non-specialist teachers, but as Ofqual (2013) said in its letter to the Secretary of State on 11th May, '*the current system has too many perverse incentives, [which] can distort teaching [and] narrow the curriculum*'. The chemistry curriculum has undergone similar changes, although there has been less 'shifting of topics' from GCSE to AS and A level, the focus being principally on transferring the more complex concepts and skills upwards, and there is a suggestion that these will soon shift back down in the current review of the KS4 curriculum. The teacher shortage is also less acute in Chemistry as there are more Chemistry A-level students, more undergraduates, more teachers in training and more specialist teachers in post.

Triple Award Science might provide the highest attaining students with the opportunity to study for three separate GCSEs in Biology, Chemistry and Physics, but it requires greater teacher specialism than 'Single Award' Science and the supply of specialist teachers in hard science to deliver them has not increased enough, especially in underperforming schools. Therefore the effect of curriculum churn and more public accountability metrics on the recruitment and retention of specialist teaching staff in STEM subjects is an issue

regarding the extent to which high-ability students from disadvantaged areas have access to those subjects.

7. Subject Knowledge Enhancement (SKE) schemes as part of Initial Teacher Training (ITT) programmes has increased the supply of Physics, Chemistry and Mathematics teachers to the market, but there is little research on the efficacy of the initiative and little attention currently paid to it (I o P, 2012) despite the support of learned societies like the Royal Society, the Royal Society of Chemistry and involvement of the Institute of Physics, for example, with The Gatsby Charitable Foundation (2004) and the (then) Training and Development Agency in the earlier 2004 Physics Enhancement Programme. What *has* been done - most recently, by CooperGibson Research (2013) in their three-year longitudinal evaluation of mathematics, physics and chemistry SKE – finds '*clear evidence … of the success of SKE courses in preparing teacher trainees sufficiently with [the] subject knowledge they require, equipping them to specialise in teaching a subject in schools, providing an alternative route into teaching which is on a par with traditional entry teacher training and supporting the supply and quality of teachers into the profession*'. (Gibson et al., 2013: 16).

The US is similarly facing a shortage of STEM professionals amid fears that the shortage of physics teachers is 'leaving too many US students unprepared for college study in STEM disciplines' (Cornell University, 2013). As a matter of urgency, we need some research on the *demand for* (as opposed to the *supply of*) SKE-trained teachers relative to their non-SKE counterparts. The CooperGibson evaluation supports the anecdotal evidence that schools regard SKE teachers as well-motivated and well-prepared, having had six months of focused Physics or Chemistry knowledge and pedagogical input in addition to the 'normal' PGCE training, but early SKE cohorts are now moving into positions of responsibility in their schools – for example, as Heads of Department - and it will be important to monitor how that plays out. It will also be important to monitor the effect of the shift of 'core' ITT towards the 'School Direct' training route and to raise schools' awareness of the SKE 'route' in specialist teaching (Gibson et al., 2013). Current estimates from ITT providers regarding uptake and turnaround time for SD applicants suggest that it is proving very problematic and likely to be destabilising on recruitment to STEM subjects just when the system was addressing, relatively successfully, teacher shortages.

B. Contemporary Additional Supply-Side Policies Concerned with the Provision of Education

1. Develop Schools' Capacity for Internal School Evaluation

There are suggestions from the literature that systems and processes of internal school self-evaluation may potentiate the power of external systems (Nevo, 2001, 2002: Janssens & van Amelsvoort, 2008), perhaps because of the enhanced 'evaluation literacy' and 'data richness' that this may produce.

A focus upon learning from schools' own Within School Variation (WSV) may be a useful start here , particularly as there are programmes of demonstrable effectiveness (Reynolds, 2007, 2010), which build capacity in the areas of middle management development, teaching and learning, development of Standard Operating Procedures (SOPs), student voice and data usage. Hofman and colleagues (2009), provide strong empirical support for the importance of within school collaboration, and the hope would be that the synergy between enhanced internal capacity and external evaluation would be potentiated in its effects.

2. <u>Develop Inspection to Focus Upon Student Outcomes More Than Processes</u>

Current UK Inspection systems focus heavily upon evaluation of whether internal school processes are 'effective' judged by their reflection of prior stated characteristics. Yet there is now considerable evidence that 'what works' may be contextually variable in terms of the factors that generate effectiveness, in schools of different socio-economic backgrounds for example (Teddlie & Reynolds, 2000).

Rather than being 'tight' upon the precise characteristics of organisational processes, it may be more useful to be 'tight' upon the within-school systems of thinking that schools should possess, and 'loose' upon organisational characteristics required, given contextual variability in 'what works'. In other words, inspection would be to check that schools possess the (defined, same) internal capacity to generate what should be (not defined, varied) contextually appropriate organisational processes.

Such approaches involve the use of data systems, internal school benchmarking against their own best practices, continuing professional development, a focus upon education at the point of delivery as seen by students, and the development of an empirical/rational problem solving approach by school management. Blueprints for this kind of school exist currently (Reynolds, 2007, 2010).

3. Communicate Better With Professionals

There are hints from the literature that professionals may be adversely affected by external accountability systems if they do not grant legitimacy to the policies and governments associated with them, thereby reducing their effectiveness. Given that much communication of policies, and the reasons for them, has historically been aimed more at the wider electorate in the United Kingdom than at professionals (indeed New Labour used criticism of professionals in an attempt to boost its standing with voters, see Reynolds, 2010), ensuring that professionals know why they are being expected to do things as well as what the things are may be productive.

This would also apply in the case of communicating any findings from the 'Vision' initiative, given the historic presence of attitudes in schools that may be more favourable to the Humanities and Social Sciences than to STEM subjects.

4. Develop the 'Meso' or Middle Level of Accountability

The 'meso' level lies mid-way between national governments and individual schools and one of the key findings from the international PISA survey (OECD, 2010), is that the more successful educational systems utilised this 'meso' level of education. This level can operate as a form of support to schools since it is more closely located to them and is therefore more knowledgeable about their local contexts. This level could also be able to match schools with sources of support within local communities, and with 'systems' outside the educational system.

For accountability purposes, also having something more local than the national State would be able to interpret some of the detail (and vagaries) of performance indicator data in ways that are probably more knowledgeable and fairer than a national system can be.

An example of successful 'meso' level provision in a United Kingdom context would be one of the effective Local Authorities and/or one of the better of the Academy 'Chains', where the 'meso' level is aligned with school needs, is itself self-evaluating and involved in school training and professional development.

5. Re-conceptualise the STEM Curriculum Cores

The ferment in the public examination systems makes it possible to consider the possibility of curriculum reform, confident that this will in any case be

happening in Wales, Northern Ireland and England because of the break-up of the common examination system at GCSE and 'A' levels.

We noted in our review that the details of what counts as 'success' in different societies may include the following:

- Stressing the links between the various Science subjects;
- Stressing how science can be involved in the creation of a better society;
- Use of enquiry-orientated, rational empirical methodologies within individual areas of the curriculum that have across-subject effects;
- Use of the concept of 'proficiencies' (as in Mathematics in the United States) to define good learning and teaching.

There are also countries that are trialling new methodologies of curriculum and instruction that may have utility for us, such as the Chinese desire to increase the amount of practical teaching/work in Science, which is possible in their country now they have the economic resources to provide the (inevitably expensive) practical resources.

It would be a great pity to miss the opportunity of re-conceptualising the STEM curricula if curriculum reform is going to be happening anyway. It is important not to preserve the status quo simply by default.

C. Contemporary Additional Demand-Side Policies Concerned with Societal Demands of the Educational System

1. Increase the Number of Outcome Measures Schools Are Judged Upon

Schools should be judged upon a broader range of outcomes than the present range of merely academic achievement data. With Mathematics and Science particularly, given the likely destination of school and university student graduates, the skills of working collaboratively with others, being good team members and the interpersonal qualities of empathy, trust and genuineness that mutual, reciprocal working necessitates may need to be measured too, to ensure that they are developed and given the same attention as academic outcomes. Interestingly, collaborative work is to form part of the 2015 PISA testing for some volunteer countries, using in this case an Avatar as the collaborator with the individual student. These outcomes may have been historically labelled 'soft' but that does not make them unimportant.

More 'social' outcomes of education – self-esteem, perceptions of the quality/supportiveness of teachers etc – should be considered for inclusion in national UK accountability systems. Whilst the evidence is overwhelming that if one uses methods such as Structural Equation Modelling (SEM) in analysis then the direction of causality is of academic outcomes generating the social, rather than vice versa (see review in Teddlie & Reynolds, 2000), measuring the achievement of both kinds of outcomes rather than one kind is likely to potentiate the attainment of them both.

Two caveats need to be borne in mind though. Firstly, if one multiplies the number of outcome measures by which schools and systems are judged, one needs to ensure that the performance indicators are aligned with what is happening in process terms in schools, otherwise schools will be assessed upon things they are not doing. Secondly, increasing the number of outcomes being measured may have negative effects if the things being additionally measured are negatively correlated.

2. Enhance Demand Side Capacity to Respond to Choice

The operation of the 'educational markets' involved in parental choice of schools may be limited if there are blocks on the exercise of that demand. Currently, there are many schools 'maxed out' in their capacity to respond to parental demands for places (as they are seen as effective) because they do not have any spare places, whilst there are high levels of surplus places in schools that are unpopular (because they are seen as poor or due to population movement). These factors may reduce any possible impact of market based pressure and competition upon school quality. There are also many schools 'maxed out' because of increases in the size of the student age cohorts putting pressures upon school accommodation, which seems to be a particular problem in parts of England.

It is acknowledged, though, that desire to help educational markets perform better is more an English governmental concern than that of the three devolved administrations, although, paradoxically, the population of 'surplus' places is much higher in these three countries than in England.

3. Audit the Component Parts of the Demand Side

We noted earlier that the operation of 'high stakes' accountability regimes may create perverse incentives to manipulate performance (O'Neill, 2013). Rather than take individual school performance on the various performance indicators

as a given, inspection agencies could be encouraged in their Inspection Frameworks to focus on how the data has been generated as much as on simply what the numeric scores are – on entry policy, on option ranges/choices, on choice of exam boards and on policies for all the various component parts of the ability range.

4. Make the Performance Indicators More STEM Relevant

The performance metrics utilised have customarily been English and Mathematics at GCSE level, generating the benchmark of performance of 5+ GCSEs at grade 5 or better, which has been used in all the four countries of the UK as a system level indicator.

This should be changed to include a Science indicator, although we recognise that the difference between 'Triple Science' and 'Single Advanced Science' would necessitate some work in the particular construction of the indicator. Given the historic importance of the gender division in students' choice of the various STEM related subjects, it is important to develop separate performance scores for the two gender groups to expose the extent to which schools may be succeeding, or not, in equalising access to different curriculum opportunities.

D. Contemporary Policies to Develop Educational Professionals

Even a more appropriate set of 'supply side' and 'demand side' policies may of itself not be enough to further improve UK levels of educational achievement in the STEM area, and more generally, unless the middle of the sandwich – the professionals in schools – are able to respond to the changing pressures on them. Indeed, many of the suggested policy changes we have outlined here so far may make the pressure on professionals more rather than less. What kind of profession may be able to thrive in these circumstances?

The first area for attention is the basic methodology of training of the profession from entry. English and Mathematics are needed at Grade B for entry – that should include a Science pass at B too. More importantly, the entire model of training needs to become one linked to the scientific rational/empirical paradigm. At present, training can be called largely an 'apprenticeship' model, where the trainee sits at the feet of the master/mistress and receives their methods and historic stock of wisdom, yet of course such a model is particularly inappropriate in times of rapid change, as the curriculum knowledge and the methodologies passed on by the

master/mistress may be out of date and indeed irrelevant. In the STEM area this may be particularly damaging in that the speed of change of scientific and technological knowledge may be even greater than in society in general.

An alternative methodology is to graft on an 'enquiry' orientation in terms of equipping professional entrants to utilise the rational, empirical paradigm that deals in the currency of 'evidence', as in the much utilised phrase of 'evidence based practice'. Wales' new Masters degrees for all entrants to schools in their first job is orientated to this ideology and practice – teachers are encouraged to select a problem or area to concentrate upon, to scope the existing literature about the topic, to measure how the land lies in this topic in their school, to choose intervention strategies that are again evidence driven and based upon a review of the educational research and practice literature, to implement the strategy, to measure change utilising data and to go round the 'cycle' further if they need to.

There are societies that adopt methods similar to this – Cuba teaches research methods to all teachers in their professional training, encourages all teachers to conduct research in their schools, encourages schools to have research conferences in which teachers explain their findings and encourages multi-school conferences to do the same. China now wishes its advanced teachers to prepare annually, as part of their professional development, a journal quality publication about their practice and/or their subject.

It is quite extraordinarily damaging for teachers of STEM subjects to be in subject areas that have made their intellectual and practical progress precisely because they use the rational empirical model, to be themselves in an educational training system that uses the defunct methods of apprenticeship.

Other professional needs are for:

- Quality CPD offerings in both STEM and all other areas, updating and uprating curricular knowledge as it itself changes;
- School improvement to shift from its present obsession with 'school to school' transfer of good practice, which in any case rarely happens, to a 'within-school' orientation where transfer takes place by evaluating school subjects, benchmarking against best practice and therefore making each school's best practice its own standard practice.

E. Contemporary Policies to Develop the STEM Area

The STEM area currently is a much trampled field where a variety of actors, professions, government departments, professional associations, learned societies, trusts and pressure groups are operative. This is likely to result in more lateral than vertical movement/momentum. Vertical movement necessitates a STEM-wide organisation to promote STEM-wide issues.

A Commentary on Future Policies/Contexts Over the Next 15 to 20 Years

We have concluded that there needs to be some form of accountability system in the countries of the United Kingdom, for the following two reasons:

- There is an historical trend since the 1970s whereby the educational system's high degree of self-regulation seems to have been (widely agreed) as generative of low educational standards and a slow pace of change;
- The nationally and internationally structured research on accountability is supportive of the positive effects of accountability measures on student outcomes.

We also outlined the different ways in which 'success' in various STEM areas has been defined, and the present controversies and debates in the various STEM subject areas concerning curriculum coverage, assessment, professional training and educational subject leadership.

We concluded by outlining a variety of 'supply-side' policies to develop the capacity of the professionals – within the accountability system – to respond to accountability processes in a productive fashion. We also added some policies and recommendations about how the 'demand-side' of the accountability system could be made more effective.

It is now time to outline the practical detail and organisational structure of what a desired accountability system might look like, and then speculate lastly about possible future societal changes and how they might affect what is deemed to be 'successful' STEM education.

A New UK Accountability System?

The countries of the UK at present evidence different accountability systems, and attempting to recommend practice that should be adopted across all

countries may need to reflect the detail of the social, educational, cultural and economic contexts of each society, of course. But it is clear that certain country approaches may be useful across the board, given our earlier discussions and literature review:

- The use of inspection systems carried out by independent bodies such as HMI, with the frequency of inspection dependent upon schools' needs and self-evaluation (as in all four countries presently), with judgements made about a wide range of outcomes and processes;
- The use of both internal teacher-based, formative assessment and external testing and/or public examinations as the summative measures, with the formative assessment results kept within the educational system and the summative measures located in the public domain (as in all four countries presently, although in only Wales and England are the assessments put into the public domain by the State). This would have the effect of separating the data for accountability (the summative assessment) from the formative assessment utilised for decisions about the education of individual children and for teacher professional development purposes, whereas historically there have been invalid assumptions that the data collected for one purpose will do the other job too;
- Accountability to existing stakeholders being maximised with an enhanced developmental role for parents and communities through governors/school councils and through the operation of parental choice (though this latter only currently operates strongly in England because of variety of types of school to choose from);
- Outcomes being measured across a wider range of social, affective and additional academic areas;
- Approaches being developed that explicitly address the problem that disadvantaged communities may not be well equipped to maximise their 'demand-side' accountability to make their local schools accountable (which would necessitate either intensified 'State', or Inspection agency or other agency involvement to make up for their 'shortfall' of demand accountability as it were);
- The attempt to increase educational professionals' and leaders' competencies and opportunities to 'self-manage' the educational system, as trust in them returns with the programme of professional renewal that we outlined earlier in our material in Contemporary Policies.

In all these accountability areas, change would need to be incremental to ensure that there is no repeat of the 'overload' of educational change that seems to

have had such negative effects on the morale of the profession. And of course, we should not expect adoption of all aspects of a new accountability system across all UK societies – some of the 'trust' that historically has been vested in educational producers in Scotland seems to have survived the drift of the last 30 years, whereas in Wales particularly it seems to have largely vanished suggesting that system self-accountability could be relied upon more (if the trust is not misplaced) in a Scottish context than in a Welsh one.

We should note that the next few months will provide further opportunity for UK countries to assess whether their educational systems in general and their accountability systems in particular are adequate, with the publication of the most recent PISA data collection from 2012, in December 2013. The trend has been for the four UK countries to diverge more over time and in different subject areas, as no doubt the effects of their different nationally generated policies have differential effects. Wales has been a rapid decliner by comparison with the other 3 countries, whereas N. Ireland has improved over time. England and Scotland have been more stable, relative to other societies. One can expect robust nationally based discussions as well as UK –wide ones, on the nature and content of country accountability systems shortly!

A New Vision for the STEM Area

An accountability system that may be effective across the countries of the UK will exist in a world where major changes are likely in the economic, social and cultural contexts of society which have considerable implications for the process and outcomes that exist within STEM curriculum subjects.

Firstly, the revolution in information technologies means that learning will increasingly take place in a range of learning situations other than the school, involving the home, the community and non-formal educational settings, necessitating that the STEM subject community must be alive to the need to maximise the quality of both formal and non-formal education.

Secondly, the decline in the power of those parts of society that historically generated personal and social values, dispositions and attitudes – the Church, the community, maybe the family unit – means that these attributes will now increasingly have to be generated within the formal/informal education process, in turn meaning that the STEM area needs to be aware of the need to promote world views appropriate to the success of science and technology in these settings. In this context, the importance of responding to the 'non-rational' or anti-scientific' emanations that have often appeared is important, as is the

necessity of understanding and shaping the interaction between science and society, particularly in areas such as sustainability.

Thirdly, learning in the future is bound to involve an enhanced role for Computing and Communications Technology, in both the development of skills and the acquisition of scientific and technological subject knowledge. Yet the STEM area seems strangely not to be in advance of the other subject/professional communities in its attention given to technologically facilitated learning opportunities. This needs to change.

The societal changes within which the STEM area is likely to be situated in the next 15-20 years will be paralleled of course by the likely changes within the educational system in its process and desired outcomes.

Firstly, desired outcomes are likely to be increasingly of the 'metacognitive' kind, completing the transition from the old historically conventional 'knowledge based' outcomes to universal adoption of 'skills based' approaches. These outcomes can be entitled self-regulated learning, or reflective learning or 'learning to learn' skills, but are essentially all about the enhanced future need to focus upon how knowledge is accrued and generated by individual learners, and the ways in which they produce their knowledge. The movement towards adoption of these kinds of outcomes will also be potentiated by their increased adoption within the benchmark PISA surveys from OECD (where metacognitively orientated questions will greatly increase in quantity in the next 2015 testing round).

Given that the nature of scientific knowledge involves the initial acquisition of 'known to be true' bodies of knowledge and given that these maybe different to other kinds of knowledge, the implications of enhanced adoption of metacognitive skills for STEM subjects need attention.

Secondly, learning in future will increasingly be collaborative in nature, involving less individual learning and more group based learning. Partly this is because of the evidence that collaborative learning situations can actually teach the social skills that our young people need – in a group that is discussing the evidence about any subject, the students will have to develop social skills for telling their fellow students they may be wrong for example. Partly it is also because of the evidence of the power of collaborative or co-constructive situations in developing the academic and academically related outcomes. Such collaborations can of course be either 'soft' and relatively informal or formalised or 'hard' but their utility/applicability should be particularly marked

in STEM subjects because of their inherent, collaborative, practical and experimental nature.

Thirdly, education will increasingly involve the attempted development/transmission of attributes that have historically been 'left to chance', particularly related to the destinations and occupational futures of their students. The historic orientation of the educational system in many accounts did not concern itself with the need for the knowledge it transmitted to be used for wealth creation to further social advance, so there is likely to be a further emphasis upon the need for an 'enterprise' or 'entrepreneurial' culture in education which will have particular importance for STEM subjects, given that they are about wealth generation (among other things) and given that many school and university STEM graduates work directly in the so-called 'productive' industries.

These attempts historically were stymied by the absence of models of effective educational practice in these areas, which led to an emphasis upon the adoption of the attributes and characteristics personally of successful entrepreneurs and businessmen/women. But there are now – on the back of the knowledge gained from the large number of interventions in this area – replicable programmes that appear to be making progress in developing a 'technology' of intervention to ensure young people's take up of the behavioural and organisational strategies for entrepreneurship. These need to be widely adopted to generate outcomes from schools in the STEM areas that are qualitatively increasingly different from historic cohorts.

In all these areas of likely future contextual and educational development, STEM subject communities – and we return here to where we began our Report – need to be proactive in the development of strategies that when used in schools and classrooms will benefit the STEM areas, because we know that STEM subjects are particularly affected by the quality of education provided. If we are right that schools will be expected to concern themselves more with values than historically, then we need to ensure that educational professionals know how to deliver those of importance to the STEM area. If learning is to be more IT based and more collaborative, we need to ensure that it is clear what the effective procedures are in the STEM areas.

In all likelihood, organisations concerned with other school subject areas, pressure groups and organisations generally that are concerned with educational matters, will plan both interventions in any debates and programmes of

proposed educational actions to maximise the advantage for their own interests. The STEM subject communities and organisations need to do the same.

Appendix: Methods of Working

Our areas of interest were broadly twofold:

- How do different societies define what are 'successful' STEM educational outcomes, and how are the various 'literacies' of 'maths literacy', 'technological literacy' and 'scientific literacy' more generally operationalized both in the four component nations of the UK and internationally?
- How do the four UK nations, and other societies culturally, economically and socially similar to the UK, attempt to make their educational systems accountable and what is the evidence about what may be 'best practice' in this area? Specifically, how can 'trust' be built among populations and their governments concerning their educational systems?

We focussed upon an analysis of past developments in these issues, upon present thinking and projected future developments, both UK wide and internationally. We should note we were broad in scope, and looked at books, journal articles, conference papers/proceedings, material from national/international organisations such as OECD, The World Bank and UNESCO, and the 'grey literature' of reports, opinion pieces and other materials accessible through Google and other information sources.

In detail:

- We used the literature from our own research on comparative education, educational policies internationally and the effectiveness/improvement of schools and educational systems;
- We reviewed the contents of 44 of the leading international and UK educational journals to find articles on accountability and accountability related issues;
- We used keyword based database searches (such as ERIC) to identify material;
- We consulted with 'key informants' knowledgeable about the international literature, and about the STEM subjects and their subject communities;
- We searched on Google to see if it threw up further publications of interest (it didn't), but it did furnish material of use for the more inevitably speculative material on the likely trajectory of educational policies related to STEM over the next 15 to 20 years;
- For this more 'future orientated' material, we reviewed the literature we had used in our own writing on a range of areas such as IT, cognitive neuroscience, the science curriculum in schools and potential new models

of professional education/professional development, and also scoped the 'future thinking' material emanating from the school effectiveness, school improvement, transformative education and system/societal change literatures

Bibliography

Advisory Committee on Mathematics Education (ACME) (2011). *Early and Multiple Entry to GCSE Mathematics*, available online at: http://www.acme-uk.org/media/7392/early%20and%20multiple%20entry%20to%20gcse%20final.pdf

Ahmed, A., Gimenez, J., Keitel, C., Klakla, M., Kraemer, J.-M. and Porfirio, J. (2001) *Mathematical literacy in the digital era: research, teacher education and classroom practice aimed at a mathematics education for all*, Discussion paper for CIEAEM 53, Verbania, Italy.

Anderson, J. (2005). *Accountability In Education.* Paris: The International Academy of Education/UNESCO.

Atkins P.W., (2005), Skeletal chemistry, *Education in Chemistry*, 42, pp.20 & 25.

Auld, R. (1976). *The William Tyndale Junior and Infants School: Report of the Public Enquiry, conducted by Mr. Robin Auld.* London: Inner London Education Authority.

Ball, S. J. (2008). *The Education Debate.* Bristol: Policy Press.

Barber, M. (2007). *Instruction to Deliver. Tony Blair, public services and the challenge of achieving targets.* London: Politico's Publishing.

Bennett, N. (1976). *Teaching Styles and Pupil Progress.* London: Open Books.

Bernstein, B. (1968). Education cannot compensate for society. *New Society,* 387, pp. 334-347.

Borich, G. (2006). *Effective Teaching Methods* (3rd ed.). New York: Macmillan.

Buckingham, D. (2008). Defining Digital Literacy. What do Young People Need to Know About Digital Media? In C. Lankshear & M. Knoble (Eds.). *Digital Literacies: Concepts, Policies and Practices.* New York: Peter Lang.

Burgess, S., Wilson, D., & Worth, J. (2010). *A natural experiment in school accountability: the impact of school performance information on pupil progress and sorting.* CMPO Working Paper Series No. 10/246. Bristol: CMPO.

Burghes, D. & Blum, W. (1995). The Exeter Kassel Comparative Project: A Review of Year 1 and Year 2 Results, in Gatsby Foundation, *Proceedings of a Seminar on Mathematics Education.* London: Gatsby.

Bybee R. (1997) *Achieving scientific literacy: from purposes to practices* (Portsmouth, New Hampshire: Heinmann Publishing).

CAS (2012). Computer Science: a curriculum for schools. (March 2012). Available online at: http://www.computingatschool.org.uk/index.php?id=documents

CAS (2013). See website at http://www.computingatschool.org.uk/index.php?id=about-us Accessed 08/06/2013.

Chiang, H. (2009). How accountability pressure on failing schools affects student achievement. *Journal of Public Economics.* 93, pp. 1045-1057.

Chitty, C. (2009). *Education Policy in Britain.* (Second Edition). Basingstoke: Palgrave Macmillan.

Coleman, J. S., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfeld, F. & York, R. (1966). *Equality of Educational Opportunity.* Washington, DC: US Government Printing Office.

Comber, L. C. & Keeves, P. (1973). *Science Education in Nineteen Countries.* London: John Wiley.

CooperGibson Research, (2013). Longitudinal evaluation of Subject Knowledge Enhancement (SKE) courses. Available online at: http://www.coopergibson.co.uk/coopergibson_past_projects.html Accessed 30/10/2013.

Cornell University, (2013). The Crisis in Physics Education. Available online at: http://phystec.physics.cornell.edu/content/crisis-physics-education Accessed 30/10/2013.

Dee, T., & Jacob, B. (2009). *The Impact of No Child Left Behind On Student Achievement.* NBER Working Paper No. 15531. Cambridge MA: NBER.

Dee, T., & Jacob, B. (2011). The Impact of No Child Left Behind On Student Achievement. *Journal of Policy Analysis and Management,* 30 (3), pp. 418-446.

Department of Education RSA (2003) *National Curriculum Statement Grades 10-12 (General): Mathematical Literacy* (Pretoria, RSA: Department of Education).

DfE (2013a). *Computing: Programmes of study for Key Stages 1-4.* Available at: www.education.gov.uk/nationalcurriculum Accessed 08/06/2013.

DfE (2013b). *Consultation on computing and disapplication of the current national curriculum.* 3rd May 2013. Available at: www.education.gov.uk/schools/teachingandlearning/curriculum/nationalcurriculum2014/a00224578/consultation Accessed 08/06/2013.

DfE (2012). *What Can we Learn From the English, Mathematics and Science Curricula of High Performing Countries?* London: Department for Education.

DfEE (1999) *National Numeracy Strategy: Framework for Teaching Mathematics from Reception to Year 6.* London: DfEE.

DeWolf, I. F. & Janssens, F. J. G. (2007). Effects and Side Effects of Inspections and Accountability in Education: An Overview of Empirical Studies. *Oxford Review of Education,* 33(3), pp. 379-396.

Durant, J. R. (1993). What is scientific literacy? In: J. R. Durant & J. Gregory (Eds), *Science and culture in Europe* (London: Science Museum). pp. 129-37.

Fauberty, W. (2009). *School Education: Current Practices in OECD Countries and A Literature Review.* OECD Education Working Paper No. 42. Paris: OECD.

Figlio, D. & Rouse, C. (2006). Do accountability and voucher threats improve low-performing schools? *Journal of Public Economics.* 90, pp. 239-255.

Fullick, P. (2013) in interview with the authors, 11 June 2013.

Garrett, C. (2013) in interview with the authors, 12 June 2013.

Gibson, S., O'Toole, G., Dennison, M. & Oliver, L. (2013). *Evaluation of Subject Knowledge Enhancement Courses Annual Report, 2011-12.* London: DfE

Available online at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/210037/DFE-RR301A.pdf Accessed 7/7/2013.

Gilster, P. (1997). *Digital Literacy.* Chichester: John Wiley.

Grace, M., Woods-Townsend, K., Griffiths, J., Godfrey, K., Hanson, M., Galloway, I., Azaola, M. C., Harman, K., Byrne, J. & Inskip, H. (2012). Developing teenagers' views on their health and the health of their future children. *Health Education*, 112 (6), pp. 543-549.

Grace, M., Woods-Townsend, K., Griffiths, J., Godfrey, K., Hanson, M., Galloway, I., Azaola, M. C., Harman, K., Byrne, J. & Inskip, H. (2012). Developing teenagers' views on their health and the health of their future children. *Health Education,* 112(6), pp. 543-549.

Hanushek, E., & Raymond, M. (2005). Does school accountability lead to improved student performance?. *Journal of Policy Analysis and Management,* 24 (2), pp. 297-327.

Harris, S., Keys, W. & Fernandes, C. (1997). *The International Mathematics and Science Study, Second National Report, Part One.* Slough: The National Foundation for Educational Research.

Hawkridge, D. (1990). Who needs computers in education and why? *Computers and Education.* 15 (1-3), pp. 1-6.

Hofman, R. H., Kikstra, N. J. & Hofman, W. H. A. (2009). 'School Self-Evaluation and Student Achievement'. *School Effectiveness and School Improvement.* 28 (1), pp. 47-68.

Holbrook, J. and Rannikmae, M. (2009). The Meaning of Scientific Literacy. *International Journal of Environmental & Science Education*, 4(3), pp.275-88.

Holman J., (2002), What does it mean to be chemically literate? *Education in Chemistry*, 39, pp.12-14.

Jacob, B. (2005). Accountability, incentives and behaviour: the impact of high-stakes testing in the Chicago public schools. *Journal of Public Economics,* 89, pp. 761-796.

Jacob, B. A., Courant, P. N. & Ludwig, J. (2003). *Getting Inside Accountability: Lessons from Chicago. (Brookings-Wharton Papers on Urban Affairs).* Washington D.C.: Brookings Institution Press, Pp. 41-81.

Jacob, B. & Levitt, S. (2003). Rotten apples: an investigation of the prevalence and predictors of teacher cheating. *Quarterly Journal of Economics,* 118, pp. 843-877.

Janssens, F. & van Amelsvoort, (2008). School self-evaluations and school inspections in Europe: An exploratory study. *Education Evaluation,* 34, pp. 15-23

Jenkins, E.W. (1994). Scientific literacy, in: T. Husen & T. Postlethwaite (Eds), *The International Encyclopedia of Education*, Vol. 9, 2nd ed., pp. 5345-5350. Oxford: UK: Pergamon Press.

Jenks, C. S., Smith, M., Ackland, H., Bane, M. J., Cohen, D. Gintis., Heyns, B. & Michelson, S. (1972). *Inequality: A Reassessment of the Effect of the Family and Schooling in America.* New York: Basic Books.

Jürges, H., Schneider, K., Senkbeil, M., & Carstensen, C. (2012). Assessment drives learning: the effect of central exit exams on curricular knowledge and mathematical literacy. *Economics of Education Review*, 31(1), pp. 56-65.

Kilpatrick, J. (2001a) Understanding mathematical literacy: The contribution of research. *Educational Studies in Mathematics*, 47(1), pp. 101-16.

Kilpatrick, J., Swafford, J. and Findell, B. (eds.) (2001b) *Adding It Up: Helping Children Learn Mathematics* (Washington, DC: National Academy Press).

Koretz, D. (2005). *Alignment, High Stakes, and the Inflation of Test Scores*. Los Angeles: University of California, National Center for Research on Evaluation, Standards and Student Testing (CRESST).

Ladd, H. (1999). The Dallas school accountability and incentive program: an evaluation of its impacts on student outcomes. *Economics of Education Review,* 18, pp. 1-16.

Laugksch, R. (2000). Scientific literacy: A conceptual overview. *Science Education,* 84(1): 71-94.

Leithwood, K., Jantzi, D. & Mascall, B. (1999). Large scale reform: what works? Submitted as part of the 'External Evaluation of the UK National Literacy and Numeracy Strategy. Toronto, Ontario: Ontario Institute for Studies in Education.

Leithwood, K., Steinbach, R. & Jantzi, D. (2002). School leadership and teachers' motivation to implement accountability policies. *Educational Administration Quarterly.* 38 (1), pp.94-119.

Lingard, B., & Ozga, J. (2007). *The Routledge Falmer Reader in Education Policy and Practice.* Abingdon: Routledge.

Linn, R. L. (2005). 'Issues in the Design of Accountability Systems'. Los Angeles: University of California, National Center for Research on Evaluation, Standards and Student Testing (CRESST).

Looney, J. (2011). *Alignment In Complex Education Systems: Achieving Balance and Coherence.* OECD Education Working Paper No. 64. Paris: OECD.

Luginbuhl, R., Webbink, D. & deWolf, I. (2007). *Measuring the Effect of School Inspections on Primary School Performance: A Study Based on CITO Test Scores.* CPB, Den Haag, The Netherlands.

Matthews, P. & Sammons, P. (2005). Survival of the weakest: the differential improvement of schools causing concern in England. *London Review of Education,* 3(2), pp. 159-176.

Millar, R. and Osborne, J. (Eds), (1998). *Beyond 2000: Science education for the future. The report of a seminar series funded by the Nuffield Foundation* (London: King's College). Available online at: http://www.nuffieldfoundation.org/sites/default/files/Beyond%202000.pdf Accessed 14/6/2013

Muijs, D. & Reynolds, D. (2011). *Effective Teaching: Research and Practice* (3rd ed.). London: Paul Chapman.

Nevo, D. (2001). School evaluation: internal or external?. *Studies in Educational Evaluation,* 27 (2), pp. 95-106.

Nevo, D. (2002). School-based evaluation: an international perspective. *Advances in Program Evaluation,* Vol. 8. Oxford, UK: Elsevier Science Ltd.

O'Day, J. (2002). Complexity, Accountability and School Improvement. *Harvard Education Review,* 72 (3), http://gseweb.harvard.edu/~hepg/oday.html.

OECD (Organisation for Economic Cooperation and Development), (1998). *Instrument design: A framework for assessing scientific literacy, Report of Project Managers Meeting, Arnhem, The Netherlands: Programme for International Student Assessment.* Paris: OECD.

OECD (Organisation for Economic Co-operation and Development), (2002). Glossary of Statistical Terms (Paris: OECD). Available online at: http://stats.oecd.org/glossary/detail.asp?ID=5388 Accessed 6/6/2013.

OECD (Organisation for Economic Co-operation and Development), (2007). *School accountability, autonomy, choice and the level of student achievement: International evidence from PISA 2003. (OECD Education Working Paper. No. 13*). Paris: OECD.

OECD (Organisation for Economic Co-operation and Development), (2010). *PISA 2009 Results: What Students Know and Can Do*. Vols. *1-5.* Paris: OECD.

OECD (Organisation for Economic Co-operation and Development), (2010). *PISA 2012 Mathematics Framework* (Paris: OECD). Available online at: http://www.oecd.org/dataoecd/8/38/46961598.pdf Accessed 6/6/13.

OECD (Organisation for Economic Co-operation and Development), (2012). *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy* (Paris: OECD). Available online at: http://www.oecd-ilibrary.org/education/pisa-2012-assessment-and-analytical-framework/mathematics-framework_9789264190511-3-en Accessed 6/6/2013.

OFSTED (Office for Standards in Education), (2008). *Sustaining improvement: the journey from special measures.* London: Ofsted.

OFSTED (Office for Standards in Education), (2012) *Mathematics: Made to Measure* (London: Ofsted).

Ofqual, (2009). *Review of standards in Physics: GCSE 2002-2007 and A level 2001-2007.* Available online at:

http://www2.ofqual.gov.uk/files/2009-03-27-physics_standards_report.PDF Accessed 14/06/2013.

Ofqual, (2010). See announcement at: http://www2.ofqual.gov.uk/news-and-announcements/130-news-and-announcements-press-releases/311-ofqual-rejects-new-science-gcses Accessed 13/06/2013.

Ofqual, (2013). *Letter from Glenys Stacey, Chief Regulator, to the Rt. Hon. Michael Gove MO, Secretary of State for Education, on Secondary School Accountability Consultation,* 11 May. Available online at: http://ofqual.gov.uk/files/2013-05-08-letter-to-michael-gove-school-accountability.pdf

O'Neill, O. (2013). Intelligent accountability in education. *Oxford Review of Education,* 39(1), pp. 4-16.

Paul, J. R. M. (2013). Interviewed by the authors, 11th June, 2013.

Peyton Jones, S. et al. (2011) Computing at School: International comparisons. Available online at: http://www.computingatschool.org.uk/data/uploads/internationalcomparisons-v5.pdf Accessed 9/6/2013.

Popham, W. J. (2002). Right Task, Wrong Tools. *American School Board Journal,* 189 (2), pp. 18-22.

Power M. J. (1967). Delinquent schools. *New Society,* 19th October, pp. 542-543.

Power M. J. (1972). Neighbourhood school and juveniles before the courts. *British Journal of Criminology,* 12, pp. 111-132.

Prais, S. J. (1995). *Productivity, Education and Training: An International Perspective.* Cambridge: Cambridge University Press.

Reback, R. (2008). Teaching to the rating: school accountability and the distribution of student achievement. *Journal of Public Economics,* 92(5-6), pp. 1394-1415.

Reynolds, D. (1976). The delinquent school. In M. Hammersley and P. Woods (Eds.), *The Process of Schooling.* London: Routledge and Keegan Paul.

Reynolds, D. (2007). *Schools Learning from Their Best: The Within School Variation Project.* Nottingham: NCSL.

Reynolds, D. (2010). *Failure Free Education? – The Past, Present and Future of School Effectiveness and School Improvement.* London: Routledge.

Reynolds, D., Creemers, B. P. M., Stringfield, S. Teddlie, C. & Schaffer, G. (2002). *World Class Schools: International Perspectives in School Effectiveness.* London: RoutledgeFalmer.

Reynolds, D. & Farrell, S. (1996). *Worlds Apart? A Review of International Studies of Educational Achievement Involving England.* London: HMSO for OFSTED.

Rosenthall, L. (2004). Do school inspections improve school quality? OFSTED inspections and school examination results in the UK. *Economics of Education Review,* 23, pp. 143-151.

(The) Royal Society, (2012). Shut down or restart? The way forward for computing in UK schools. Executive summary.

Sammons, P. (2008). Zero tolerance of failure and New Labour approaches to school improvement in England. *Oxford Review of Education,* 34(6), pp. 651-664.

Shamos M. (1989) Views of scientific literacy in elementary schools science programs: Past, present and future. In: A. Champagne and B. Lovitts (Eds.) *The year in school science 1989: scientific literacy* (Washington DC: American Association for the Advancement of Science).

Shaw, I., Newton, D., Aitkin, M. & Darnell, R. (2003). Do Ofsted inspections of secondary schools make a difference to GCSE results?. *British Educational Research Journal,* 29 (1), pp. 62-75.

Shwartz, Y, Ben-Zvi, R. & Hofstein, A. (2006). The use of scientific literacy taxonomy for assessing the development of chemical literacy among high-school students, *Chemistry Education Research and Practice*, 7(4), 203-25.

Slavin, R. E. & Lake, C. (2008). Effective programs in elementary mathematics: A best evidence synthesis. *Review of Educational Research,* 78(3), pp. 427-515.

Society of Biology (2013). See online at:
https://www.societyofbiology.org/education
And see online at: https://www.societyofbiology.org/policy/education-policy/school-policy/triple-science-support-programme. Accessed 14/6/2013.

Teddlie, C. & Reynolds, D. (2000). *The International Handbook of School Effectiveness Research.* London: Falmer Press.

Tomlinson, S. (2005). *Education in a Post Welfare Society.* Buckingham: Open University Press.

Watson, A., Jones, K. & Pratt, C. D. (2013). Key Ideas in Teaching Mathematics: Research Based Guidance for Ages 9 to 19. Oxford: Oxford University Press.

Wieman, C. & Perkins, K. (2005). Transforming Physics Education, *Physics Today*, November 2005. Available online at:
http://70.40.207.144/sdsu_per/articles/Transforming%20Physics%20Education%20-%20Physics%20Today%20November%202005.pdf Accessed 14/6/2013.

Woollard, J. (2013) in interview with the authors, 12 June.