

## School Accountability in England: Past, Present and Future.

Daniel Acquah

### Abstract

This review focuses on accountability in school-based education in England. Drawing on the work of West, Mattei and Roberts (2011) it outlines the dominant forms of accountability in England, namely 'market' and 'hierarchical' accountability, and explains how both forms make extensive use of examination data as the mechanism through which schools are held to account. It goes on to explore the historical and policy context that led to these forms of accountability becoming dominant, outlining the pervasive influence of 'new public management'. Both national and international evidence suggests that publishing examination data publically can have a positive impact on test scores, although there is less evidence to suggest that competition between schools is a driver of school improvement. The effectiveness of accountability measures is contingent upon the measures' statistical properties: existing evidence reviewed suggests there are a number of issues that need to be addressed. Consequences of the accountability system from the perspective of an awarding organisation are then considered, focusing on issues of early entry and strategic marking. The final section examines the impact of the dominant forms of accountability and suggests ways that the negative consequences of high stakes accountability might be ameliorated while still maintaining the positive impacts. This will involve reforming the particular metrics used in relation to test scores but also taking a broader perspective on accountability, reflecting the diversity of activities for which schools are responsible.

*Keywords: accountability, value added, education market, parental choice, strategic behaviour.*

### 1. Aims and scope of the review

The Department for Education has announced that there is to be a consultation on changes to the accountability system in England. In this paper, a number of themes are addressed, but the principal goal is to facilitate 'policy memory' – an understanding of where mistakes have been made in the past and what previous good practice could be incorporated into the current reform effort (Hodgson & Spours, 2003). In section 2, the work of West, Mattei and Roberts (2011) is drawn upon in order to describe the dominant forms of accountability in England, namely 'market' and 'hierarchical' accountability. Section 3 then goes on to chart the historical and political context that led to these two forms becoming dominant, describing the influence of 'new public management' and ending with a description of the accountability measures currently in use. This is followed by a review of national and international evidence concerning the effectiveness of market and hierarchical forms of accountability presented in section 5. The effectiveness of accountability measures is contingent upon the measures' statistical properties and so section 6 considers issues around the use of raw scores, value added and contextual value added measures. The paper then considers some of the consequences of the accountability system from the perspective of an awarding organisation. The final section examines the impact of the dominant forms of accountability and suggests ways that the negative consequences might be ameliorated while still maintaining the positive impacts. This discussion considers two

possibilities: progress that may come about by reforming and improving the individual metrics used and advances that may involve taking a broader perspective on accountability, reflecting the diversity of activities for which schools are responsible beyond purely academic outcomes.

## **2. What do we mean by 'Accountability'?**

The term 'accountability' is used in relation to a minimum expectation or standard regarding the effectiveness of a particular activity. An accountability system may be applied to something broad such as medical services, or restricted to specific initiatives such as truancy reduction (Stobart, 2007). The rationale for accountability in relation to the education system resides in the fact that it is a publicly-funded and universal state service. Education is therefore in the public interest and so the education system must be accountable both at the national and local level (House of Commons Children, Schools and Families Committee, 2009). The 2010 Schools White Paper (Department for Education, 2010) argued that schools should be accountable for achieving a minimum level of performance because tax-payers have a right to expect that their money will be used effectively.

Accountability is a multifaceted concept. In recent work, West and colleagues (Mattei, 2012a; West et al., 2011) present an account designed to capture the diversity and complexity of the forms of accountability. The forms of accountability are conceptualised as professional, hierarchical, market, contract, legal, network, and participative. The Appendix describes each form in greater detail. The types of accountability are distinguished in terms of who is accountable to whom, what they are accountable for, the various types of potential sanctions and the likelihood and severity of such sanctions.

In this paper, the focus will be on hierarchical and market accountability and West et al's conceptualisations of these two forms will form the theoretical underpinning for the rest of the review. The existing accountability regime in England is heavily focused on these two forms and they have major sanctions associated with them and thus have the potential to impact significantly upon schools. Both types also have a strong focus on test performance, stressing the same indicators in relation to attainment and these indicators are likely to be a major focus of the Government's consultation.

### **Hierarchical accountability**

Schools are held accountable through hierarchical structures for a variety of aspects of their performance. For example, schools and their governing bodies are accountable to their local authority and to Ofsted for their national test and examination results. Schools are also accountable to local authorities and the Department for Education for how they spend resources and can be challenged by auditors within the local authority. Sanctions in relation to hierarchical accountability take a number of forms. A negative Ofsted inspection can have serious consequences for the viability of a school. The Secretary of State for Education can direct a local authority to consider a warning notice, when the standards of a school are deemed to be unacceptably low. Once a warning notice has been issued the Secretary of State is also able to appoint additional governors or replace a governing body with an interim executive board. Further reputational sanctions, such as publicly 'naming and shaming' schools and replacing management teams are also associated with the hierarchical accountability regime.

### **Market accountability**

Market accountability has been promoted by policies initiated by both prior Conservative and Labour governments. These policies sought to make available a range of information by which consumers (parents) could hold English schools accountable in the market place. Information is available to parents in a variety of forms. The first is the results of national tests taken at 11, 16 and 18. As well as being published by the government, these are also widely reported and

commented upon in the media. The second source of information is the reports of school inspections by Ofsted. The current government has made a commitment to provide comprehensive information available to parents about every school (Department for Education, 2010). Collectively, these different sources of information impact upon a school's reputation and inform consumer choice (West et al., 2011). Sanctions in relation to market accountability centre on the possibility of parental exit: school closure is possible if consumer demand declines significantly. A more likely outcome is a reduction in funding: as funding is primarily based on pupil numbers, if a school becomes less popular and the numbers decrease, its budget will decrease. Nevertheless, it is debateable how likely these sanctions are in reality. West et al. point out that there is very limited scope for schools to enter and exit, citing a study showing that between 2003 and 2004 only 1.5 per cent of secondary schools entered or exited the market. Similarly, the potential for successful schools to expand is also restricted, due to size constraints. Nevertheless, as we shall see presently, recent policies have attempted to make it easier for schools to enter and exit the 'market'.

### **3. What is the historical and political context of the current accountability regime?**

The Conservative government in the 1980s and 1990s introduced a number of policies with the intention of creating a market in education (Hursh, 2005a; Mattei, 2012a). These policies are couched in the discourse of 'new public management' (Verhoest & Mattei, 2010), which seeks to use private-sector and market incentives in state education (Mattei, 2012a). The underlying philosophy of this approach is encapsulated by Pring (2012):

*Essential to the 'effective school' is agreement on precise targets. There need to be performance indicators, reliably measurable, so that we know whether those targets have been hit. Regular audits need to be carried out to check that the necessary inputs for attaining the outputs have been adopted and delivered by the education workforce (who are referred to as delivering the curriculum). This language of 'targets', 'performance indicators', 'audits', 'delivery', 'workforce', 'inputs' and 'outputs'... is drawn very much from the business world, and so it is a small step to see the learners or their parents as the customers or clients, choosing a particular commodity within a system of education which is increasingly seen as a market and in which there are schools competing with each other for custom and in which parents can exercise choice based on the audits of the respective schools (pp.747-748, emphasis in the original).*

The origins of this market model market reside in a series of policies enshrined in the 1988 Education Reform Act (UK Parliament, 1988). The first reform introduced a system of funding in which the 'money followed the pupil'. Local authorities were required to allocate money to schools on a weighted *per capita* basis. This meant that if a given school failed to attract as many pupils as in the previous year, then it would receive less money. As we have seen, this is a key sanction associated with *market accountability*. Secondary schools in England and Wales were also given the power to opt out of local authority control, with the ability to spend the money from central government as they wished. The 1988 Act also introduced the National Curriculum and Key Stage testing, with tests taken at age 7, 11 and 16, which would later become a key measure to which schools would be held accountable.

In an attempt to empower 'consumers,' reforms in the Education Act 1992 transformed the role and composition of governing bodies of schools, giving parents a major role. The rationale was to make headteachers hierarchically accountable to parents. Schools were required to provide

national test and examination results to the government, and these were then made available to parents and published widely in the press as league tables<sup>1</sup> (West et al., 2011; West, 2010). The arrival of league tables proved to be a significant development in school accountability, and there has been much debate about their impact on schools (e.g. Mattei, 2012b). The Office for Standards in Education (Ofsted) was also established in 1992, with the responsibility for inspecting the quality of education in schools (House of Commons Children, Schools and Families Committee, 2009).

It was not until 1993 that the results of independent schools were published alongside those of state maintained schools (West & Pennell, 2000). At this point, the school league tables were based on 'raw' performance measures, such as the percentage of pupils gaining five or more GCSE passes at grade A\*-C. In 1995, the government was persuaded by research evidence and approved the move to a 'value-added' system, with prior achievement of pupils upon entering secondary school used to make adjustments from different intake achievements (Leckie & Goldstein, 2009). These value added rankings were used between 2002 and 2005. In 2006, 'contextual value added' systems were adopted. In addition to adjusting for individual student prior achievements, these models also attempt to adjust for factors such as the prior achievements of a student's peers, eligibility for free school meals, and lack of spoken English at home (Goldstein & Leckie, 2008). Also at this time, the government recognised that each school-effect estimate should have a confidence interval attached to represent the degree of statistical uncertainty. The purpose of the confidence intervals was to inform judgements about differences between schools or between any one school and the population average (Leckie & Goldstein, 2009).

In 2009, the then Labour government set out proposals for a School Report Card, which was to supersede the performance tables (Department for Children, Schools and Families, 2009). The purpose of the Report Card was to report on outcomes across the breadth of performance, including pupil attainment, but also incorporating factors such as: student wellbeing; a school's success in reducing the impact of disadvantage; and parents' and pupils' views of the school and the support received. The Report Cards were to be based on the model in use in New York schools and there was a consultation on the Report Cards and their content, alongside research into their use in other jurisdictions (Husbands, Shreeve & Jones, 2008; Maughan, Cooper & Benefield, 2009). Nevertheless, the plans were abandoned due to lack of cross party support ahead of the 2010 general election (as reported in the media e.g. Shepherd, 2010).

That election brought a change of government, and accountability was a prominent feature of the Conservative Liberal-Democrat Coalition's plans for education (Department for Education, 2010). Five key reforms related to accountability were outlined. The first reform was to provide a greater variety of information to the public about each school and how it performs. This included making publically available all the information underpinning government statistical publications. The second reform involved 'sharpening' the school performance tables. As part of this the 'contextual value added' measure was removed. The argument was that the measure leads to the expectation of different levels of progress from different groups of young people on the basis of their ethnic background or family circumstances. The government argued that this was wrong in principle because it entrenches low aspirations. Accordingly, the value added measure currently in use controls for prior attainment only.

A further way in which the performance tables were reformed was the inclusion of the 'English Baccalaureate'. This performance measure was introduced to encourage students to pursue a

---

<sup>1</sup> The Government does not rank schools according to performance and accordingly refers to the tables it produces as 'performance tables'. It is newspapers who rank schools and these rankings are commonly referred to as 'league tables'.

broad academic education until sixteen-years-old (Department for Education, 2010). It is achieved by students who obtain 'good' ('C' grade or above) GCSE/iGCSE passes in English, Mathematics, the Sciences, a modern or ancient Foreign Language and either History or Geography. The Government has been keen to point out that whilst the English Baccalaureate is a performance measure, it is *not* an accountability measure. This point trades on an ambiguity in exactly which form of 'accountability' is intended. The government has not put in place school level targets for the number of pupils that should achieve the English Baccalaureate and so it is not a *hierarchical* accountability measure. Nevertheless, it clearly is intended as a further source of evidence to inform parental choice in the market place, and it has been widely reported in league tables in the press and so it is a *market* accountability measure.

The third reform was to create a more sophisticated minimum expectation for schools. Schools were to be classified as 'below the floor' if fewer than 35 % of pupils achieve 5 A\*-C grade GCSEs (including English and mathematics). This has since risen to 40 %. A school is also considered to be 'underperforming' if a below average percentage of pupils at the end of Key Stage 4 are making their expected level of progress in English (national median for 2011 = 74 %) and Maths (national median for 2011 = 66 %). Such schools can be forced to become academies. The other key hierarchical accountability measure is the level of progress made by students between key stages 2 and 4, the so called 'value added measure'. The fourth reform focused on developing a proportionate approach to inspection on the part of Ofsted. Schools judged to be 'outstanding' will only be inspected where there is evidence of decline or widening attainment gaps. In contrast, schools judged 'inadequate' receive monitoring visits each term to assess improvement. The final reform was to ensure that school governing bodies have the skills to hold their school to account, in part by offering high quality training through the National Governors' Association.

In summary, the education reforms started under the Conservatives and continued under Labour were intended to create a quasi-market in education (West et al., 2011). The current Conservative-Liberal Democrat Coalition Government is continuing with this agenda. It has rapidly extended the 'academies' programme. For example, new academies are being created as part of the Government's 'free' schools programme, which enables parents and other groups to set up state-funded schools, thus increasing choice and diversity and making it easier for schools to enter and exit the 'market'. Nevertheless, whilst market accountability is thus a prominent part of the Government's plans for the education system, this market operates in the shadow of a hierarchy (Mattei, 2012a): the results from national tests such as the GCSE are available to parents to inform school, but these indicators are also the focus of targets by the Treasury, the Department of Education and Ofsted. The paper now moves on to consider whether these two forms of accountability are related to student performance.

#### **4. What does national and international evidence reveal about the effectiveness of accountability practices?**

##### **Country profiles of assessment practices and purposes**

To distil the results and patterns of accountability arrangements across jurisdictions, PISA 2009 (reported in OECD, 2010) presented a latent profile analysis. This analysis divided the OECD jurisdictions into four groups that shared similar profiles based on two features. The first of these features was whether achievement data were used for various benchmarking and information purposes. The second was whether achievement data were used to make decisions that affect the school. The classification of jurisdictions based on the latent profile analysis showed that England is classified, along with the majority of jurisdictions into a group that use achievement data for benchmarking and information purposes and forming decisions that affect

the school. The crucial question is whether these uses have a positive impact on student performance.

### **School choice and student performance**

Competition among schools, as part of market accountability, is intended to provide incentives for schools to innovate and create effective learning environments. Nevertheless, the cross country correlations of PISA do not show a relationship between the degree of competition and student performance: the proportion of schools that compete with other schools for student numbers seems unrelated to the school system's overall student performance.

However, there is evidence that the extent to which schools compete with each other for students can be related to equity. Existing research has shown that school choice and, by extension, school competition, is related to greater levels of segregation in the school system (e.g. Bunar, 2010; Whitty, Power & Halpin, 1998). The PISA data supports this research: a greater prevalence of school competition is related to a stronger relationship between a school's average socio-economic background and the school's average student performance. PISA thus apparently offers no support for the Government's attempt to extend the 'market' approach in education, if the goal is to improve system performance and increase equity.

### **Assessment and accountability policies and student performance**

Across the OECD, jurisdictions that use standards based external examinations<sup>2</sup>, such as the GCSE, tend to perform higher, even when accounting for national income: students in school systems that use such tests perform, on average, 16 percentage points higher than students in school systems that do not use these examinations.

PISA 2009 also examined whether student achievement data is posted publicly, communicated to parents, used to make decisions regarding the allocation of resources, or tracked by administrative authorities. According to the cross country analysis, there was no measurable relationship between the use of assessment data for accountability purposes and the performance of school systems. However, looking at this relationship *within* jurisdictions, the pattern was mixed, but was positive for some measures. For example, in Germany, Japan, Norway, Switzerland and the UK, having achievement data tracked by administrative authorities was positively related to student performance. A more consistent pattern emerged in the relationship between making achievement data available and school performance: schools whose principals report student achievement data publicly perform better than schools whose achievement data is not made publicly available in sixteen jurisdictions. However, the interpretation of this finding is complicated by the fact that in most of these jurisdictions the schools that post achievement data publicly also tend to be socio-economically advantaged. When socio-economic background was controlled the positive relation held only in six jurisdictions<sup>3</sup>. This complex pattern of findings demonstrates the difficulty in using these cross country analyses to support individual accountability policies in a particular country.

A further important finding was that, whilst the use of standardised tests tended to be unrelated to school performance, it did appear to be related to levels of *equity* within school systems. Specifically, school systems that had high proportions of students in schools that use

---

<sup>2</sup> In this context 'external assessment' refers to standardised examinations that are designed and marked outside individual schools and normally take the form of a written test. Moreover, external assessment is usually conducted in supervised conditions which ensure that the work being assessed has actually been done by the student.

<sup>3</sup> Turkey, Columbia, Hong Kong-China, Kyrgyzstan, Romania, Chinese Taipei.

standardised tests tended to show a lower impact of socio-economic background on learning outcomes *between* schools and a higher impact of socio-economic background on learning outcomes *within* schools. The explanation offered for this in the PISA report is that standardised tests provide schools with instruments with which to compare themselves with other schools. It is possible that this allows schools to observe the inequalities among schools, acting as a first step towards redressing them.

The findings from these cross country analyses are consistent with two studies carried out in the UK (Burgess, Wilson, & Worth, 2011; Mattei, 2012b), which took advantage of a naturally occurring experiment. Between 1992 and 2001, secondary school performance tables were published annually in both England and Wales. Following devolution, the Welsh Government decided to abolish the publication of these tables (Reynolds, 2008). Using a difference-in-difference analysis, Burgess et al. tested the hypothesis that school effectiveness in Wales after the abolition of performance tables was lower than that of schools in England. The authors report that the abolition markedly reduced school effectiveness in Wales – the effect was equivalent to a fall of 1.92 GCSE grades per student per year, with the key performance measure of 5 A\*-C falling by 3.4 percentage points per school. Nevertheless, the impact of other changes that might partly explain the effect, such as tougher regulation of standards by Ofqual, should also be considered.

The possibility that the effect is attributable to 'gaming' of the performance measure by English schools would appear to be refuted by the fact that the national trends in the PISA tests mirrored the GCSE findings. Nevertheless, sampling issues with the PISA data for the UK also require consideration and the extent to which they might weaken the weight placed on the PISA evidence. Significantly, the impact on GCSE scores was not consistent across all schools: the top quartile of schools (according to student's prior attainment, poverty status or league table position) was not affected, suggesting that abolishing league tables in Wales not only reduced average performance but also raised educational inequality. Burgess et al. (2011) concluded that publishing tests results – in a locally comparative format – appears to be an extremely cost-effective policy for raising attainment and reducing inequalities in attainment. This conclusion is consistent with the PISA cross country analysis discussed above. William (2010) provides a review of similar studies carried out in the US which also document positive benefits on achievement.

Whilst there is evidence that the publication of test results has had a positive impact on test performance, the mechanism by which this has occurred is less clear. The Department for Education has argued that there is strong evidence that using formal external assessment as the basis of accountability has significant benefits, citing the PISA research which is argued to demonstrate that test based accountability has a positive impact on how children perform, with particular benefits for disadvantaged and minority groups. This conclusion is problematic, because it is not clear if the publication of test results is sufficient to produce the effects mentioned or if Government targets, such as the floor measure also play a role. So care is needed in going from these results to specific policies such as increasing the floor targets.

## 5. What are the technical issues with current accountability measures?

Whether using accountability measures can be used to improve educational outcomes depends partly on the measures statistical properties (Goldstein & Thomas, 1996; Harris, 2011)<sup>4</sup>. Several authors have raised issues concerning the current measures and these can be divided into three broad areas: public understanding of the measures, statistical issues around regression to the mean, and issues with using the measures for school choice purposes. Understanding these issues is essential to appreciate the limitations of the measures and may serve as the impetus for improving the functionality of the measures as part of the reform effort.

### Public understanding of the measures

Allen and Burgess (2011) argue that performance measures should be evaluated according to their functionality, relevance and comprehensibility. One of the strengths of the 5 A\*-C measure is that it is straightforward and easy to understand how it is derived. However, as Allen and Burgess suggest, parents might conclude that this measure tells them the probability that their own child would have of securing five good passes if they attended that particular school. This interpretation is not necessarily legitimate, because pupils may not be randomly allocated to schools. A related issue is the fact that the published results related to students who entered school five years previously: school quality may have improved or deteriorated by the time the current cohort sits their GCSEs.

Interpretation of the value added measure is much more problematic. Value added models are described in detail by Ray (2006), but in essence, the score that appears in the performance tables is a school level residual extracted from a multilevel regression (Allen & Burgess, 2011). This score is centred on 1000 to avoid negative values: schools expressed concern that a negative value may be misinterpreted as meaning that pupils are going *backwards* in terms of their progress. Allen and Burgess note that the measure has no natural interpretation or translation into a metric that parents are interested in, such as GCSE grades.

Leckie and Goldstein (2011) raise a number of further issues concerning interpretation of the value added score. The first is that interpretation of the score and their associated intervals are difficult for users to interpret, which is compounded by the lack of clear explanation by the government. This issue is even more serious when the scores appear in league tables in the newspapers: they often appear with no explanation whatsoever. A related issue is that no attempt is made to communicate to users, the units in which the value added scores are measured, nor the concept of a 95 % confidence interval or how they should be used. This makes it difficult for parents and other users to interpret the magnitude of the differences between scores. Leckie and Goldstein have explored a number of different ways of presenting value added and associated confidence intervals in graphical formats. Given the issues with current presentation of value added, the use of Leckie and Goldstein's method as a means of improving user understanding warrants investigation as a matter of urgency.

A further issue with the value added score is that, whilst it attempts to control for individual ability to provide an overall measure of school effectiveness, it assumes that this is the same for all students in the school (Allen & Burgess, 2011; Wilson & Piebalga, 2008). However, Thomas (1998) points out that overall statistics of pupil performance cannot give a complete picture of how effective a school is at raising and maintaining the achievement of its pupils.

---

<sup>4</sup> The situation is probably more complex than this. If Ofsted needs to make a judgement about school quality, there is a need for quantitative measures with reliable statistical properties. In contrast, one could argue that the impact of school league tables on performance is less tied to the reliability of the statistics and more about the school's attempt to improve their position.

Research has shown that some schools that appear to be effective in terms of the overall value added measure may not be so effective in terms of individuals departments or for different groups of pupils (Goldstein et al., 1993; Thomas, Sammons, Mortimore & Smees, 1997a, 1997b) – producing a single value added score for each school glosses over this nuance alongside differences that may exist between teachers and the stability of these estimates over time (Harris, 2011). The Government's intention to provide the value added measure for disadvantaged pupils and others as well as low middle and high achievers (Department for Education, 2012a) would therefore seem to be a step in the right direction, although the specific details of how this will be calculated and reported have yet to be published.

### **Regression to the mean**

The interpretation of school league tables is also complicated by the statistical phenomenon known as 'regression to the mean'. As Smith and Smith (2005) have explained, observed test scores are an imperfect measure of ability: high scores are typically an over-estimate of ability whereas low scores are typically an underestimate. This in turn, causes high and low scores to regress to the mean in subsequent tests. If this is not taken into account, changes in test scores over time may be misinterpreted as changes in achievement rather than fluctuations in achievement scores. Goldstein and Spiegelhalter (1996) apply this to the interpretation of league tables, noting that over-interpretation of a set of rankings, especially where there are large uncertainty intervals, can lead both to unfairness and to unwarranted conclusions about changes in ranks. Apparent 'improvement' for low ranking institutions or 'deterioration' for high ranking institutions may simply be a reflection of 'regression to the mean'.

Ready (2012) suggests that the crucial question is whether the observed regression is larger or smaller than that predicted by purely statistical considerations. He introduced a method that is based on the premise that if those students who were above average slip less and those who were below average improve more than predicted by regression to the mean, then a school is adding value. A further issue for school accountability systems that include value added components, argues Ready, is whether student achievement is associated with subsequent achievement gains. Using longitudinal state assessment data, he was able to show that assessments that exhibit positive or negative correlations between initial status and progress produce results that vary widely across analytic techniques used to determine value added. This means that depending on the strength of the child-level relationships between status and growth, various modelling strategies produce dissimilar or even contradictory estimates of school performance. Ready concludes that those responsible for school accountability systems need an awareness of these issues as well as the fact those different modelling strategies address different questions are likely to address to produce different results. This is particularly important in light of the Government's decision to remove certain contextual variables from the value added calculations. There needs to be a careful consideration of the issues Ready has highlighted and an understanding of appropriate uses of the new value added model used by the Department for Education.

### **Issues with using the measures to inform school choice**

Further issues with league tables, and value added specifically, concern its use in enabling parents to choose between schools. One issue, discussed by Goldstein and Leckie is that, from the point of school choice, school level factors, such as average prior achievement of students or the average social composition of the pupil, should not be controlled for in the value added model (Goldstein & Leckie, 2008; Leckie & Goldstein, 2009): If a school level factor, is associated with success, this is of interest to parents and should not be controlled for.

A further limitation of the tables for the purposes of informing school choice is that the most recent published information is based on the current performance of the cohort who entered school several years beforehand. Whilst this is perfectly legitimate for the purposes of

hierarchical accountability, for the purposes of parental choice, what is of interest is the *future* performance of the current cohort. One reason this lag is an issue is because institutions can change quickly. Goldstein and Spiegelhalter (1996) pointed out that one way this can happen is if potential students actually take decisions on the basis of previous results. In cases where students decide to choose institutions on the basis of previous results, those institutions may well end up with different value added scores, meaning that the current cohort of students making decisions between schools will be doing so on the basis of outdated and incorrect information.

These considerations motivated Leckie and Goldstein (2009) to examine whether the functionality of league tables were improved by predicting future performance of the current cohort. They were able to demonstrate that such was the level of uncertainty in predicting future performance, only a handful of schools' future performances could be separated from the overall mean and from each other with a sufficient degree of precision. This led the authors to conclude that publishing school league tables to inform school choice is a 'somewhat meaningless exercise'. Allen and Burgess (2010a, 2011) also reported low functionality of value added in relation to school choice. They propose that the average grade over the best eight subjects may be a better metric, but propose that field experiments are needed to trial different contents and formats of school performance information.

## **6. What are the consequences of the accountability system?**

Alongside concerns relating to the statistical limitations of current measures, a now voluminous literature has discussed the negative consequences of the accountability measures on teaching and learning (e.g. Harlen & Deakin-Crick, 2002; Mansell, 2007; Stobart, 2007). The main issues are summarised below, before moving onto a discussion of the unique perspective offered by an awarding organisation on three specific issues.

The 5 A\*-C measure has been criticised for leading to a concentration of resources around 'C' grade, and reducing the educational gains of lower achieving students (Burgess, Propper, Slater & Wilson, 2005; Wilson & Piebalga, 2008; see Reback, 2008 for a similar analysis from the US). The immense pressure of teachers to improve results has led to behaviours up to and including cheating (Stobart, 2007; see Jacob & Levitt, 2003 for a US perspective). Other authors have described 'teaching to the test' – a focus on the areas of the curriculum subject to targets at the exclusion of others (Goldstein, 2004; A. Smith, 2004). Coe (2008) and West (2010) both point out that the GCSE league tables create incentives for schools to enter students for 'easier' subjects and Stobart (2007) has described the regulatory 'cat and mouse' that ensues.

Recent data from an awarding organisation offers fresh insights into the impact of accountability on teaching, learning and market behaviour. The use of performance tables for accountability purposes has been identified as one of the factors influencing early entry policies at GCSE. Early entry refers to situations in which students complete a qualification before intended (Taylor, 2012a). In one of the most comprehensive quantitative analyses to date, Taylor (2012b) examined trends in early entry between 2007 and 2011 using data from all the UK awarding bodies. There was a difference in entry patterns between subjects and Taylor argued that this was likely to be a reflection of the perceived importance of different subjects in relation to the 5 A\*-C measure as early entry was found to be most prevalent in English and Mathematics. There was also evidence that early entry is more prevalent in schools with a higher percentage of students eligible for free school meals and in those schools with a lower percentage of students achieving 5 A\*-C grades. This is consistent with strategic behaviour to maximise the chances of students achieving a particular grade, such as the 'C' grade.

A further source of relevant evidence is AQA's contribution to Ofqual's report on GCSE English (Stockford et al., 2012). This evidence suggested that there was an increase in the volatility of

the results experienced by schools and colleges and this might partly be accounted for by strategic marking of the teacher assessed controlled assessment. Based on the previous performance of candidates in written examinations, it was possible to calculate the mark required for each candidate to achieve a grade C (or another target grade) on the controlled assessment units for which they were yet to be graded. There was a sharp discontinuity, with large numbers of candidates receiving the mark required to achieve a C grade. Furthermore, there was strong anecdotal evidence that the pressures felt by teachers in terms of school accountability measures have impacted on these intervention strategies and added to the level of volatility of results. The point is not to pass judgement on these behaviours, but rather to document some of the specific impacts observed by awarding organisations.

An additional effect was observed nationally in the grading of GCSE Mathematics in summer 2012 (Stockford, 2012a, 2012b). In contrast to GCSE English, where strategic behaviour was evident in the allocation of marks, in GCSE Mathematics there was evidence of highly strategic *entry* behaviour. The modular nature of the qualification enabled students to enter modules early on in their course. This meant that a detailed picture of students' achievement was available in a high proportion of the specification prior to final certification. This information allowed teachers and students to make estimates of the likelihood of achieving a grade C in that specification overall. As a result, a large number of candidates, mainly those achieving lower grades (below grade C) abandoned the specification in which these early units had been sat and instead opted to sit a linear specification. In a large number of cases, this involved switching to a specification offered by another awarding body. As an indication of the extent of this behaviour, around 160,000 unique candidates sat Mathematics units in the 2010-2011 academic year. However, only around 60 % of these candidates remained to complete the qualification in summer 2012. From the point of view of the awarding organisations, this created technical challenges to the maintenance of standards. In addition, the switching of specifications at a relatively late stage is likely to have impacted upon teaching and learning.

Although it would be premature to conclude that accountability caused the behaviour in each of these cases, it is unlikely to be coincidental that such strategic behaviour occurred in the two highest profile subjects in the context of school accountability.

## 7. Discussion

As part of an informed public debate about school accountability, it is necessary to question the priorities and the values that underlie accountability policies and practices: education is not ethically neutral (Pring et al., 2009). Reflecting on the policy history, the existing regime in England is heavily focused on hierarchical and market accountability and the reforms of the current Coalition Government have served to strengthen the dominance of these forms of accountability, by increasing school 'floor' targets and making it easier for schools to enter and exit the 'market'.

The Government's use of targets as part of hierarchical accountability as well as the attempt to create a quasi market as a further means of holding schools to account, are premised on the ethos of 'new public management' (Verhoest & Mattei, 2010). Several scholars of education have questioned whether the private-sector model is an appropriate one for state education (Green, 2010; Hursh, 2005a, 2005b, 2013; Power, 2012; Pring, 2012; Pring et al., 2009). Pring et al. argue that the application of performance management to education distorts the nature of teaching as a profession: the distinctive expertise of the teacher lies in knowing how to advance the understanding and skills of the learner beyond their current levels of competence, *not* in 'delivering the curriculum'. Pring (2012) extends this critique, arguing that the notion of *performance*, which is central to 'new public management', fails to treat young people as *persons*. Instead, they are 'means' to some 'end' or purpose such as a school's place in the

school league tables, the successful attainment of targets, or the supply of skilled workers. These purposes may be in conflict with that which is of value to the learners themselves.

Hursh (2013; see Lipman, 2013 for a analogous analysis) has provided one of the most developed critiques of these neo-liberal policies. At the risk of greatly over simplifying, his central point is that whilst these policies are couched in a discourse of promoting equality, the real goal is the privatisation of state education. The analysis Hursh provides has striking parallels with the current context in England. One issue concerns the way progress measures are used to evaluate schools. As was described earlier, schools are considered to be 'underperforming' if a below average percentage of pupils at the end of Key Stage 4 are making their expected level of progress in English and Mathematics. By using *average* levels of progress, *absolute* levels of progress become less important. Schools are, in effect, forced into a competition with each other, with large numbers facing 'failure' or being classified as below the floor and therefore being forced to convert to academy status. The point is not to pass judgement on the desirability of these policies, but rather to suggest that the rationale and implications merit wider public consultation. Such a debate could usefully include discussion of the evidence that school competition improves educational outcomes, as the PISA and national evidence on this is equivocal. Nevertheless, it is open to proponents of a market driven approach to argue that certain features of the market, such as the ability of schools to enter and exit easily, are insufficiently developed at present to produce the desired outcomes (e.g. Allen & Burgess, 2010b): recent Government policies, designed to extend the scope of the market may address this issue.

There is a contradiction manifest in the evidence reviewed so far: current forms of accountability in England have a variety of negative consequences, yet contemporaneously can have positive effects on student achievement such as an increase in equity. On balance then, it would seem appropriate to explore whether the negative effects of high stakes accountability might be ameliorated while still maintaining the positive impacts (William, 2010). Progress in this area will be facilitated by a greater understanding of the mechanism through which accountability exerts its impact. PISA has suggested publishing test results provide schools with instruments with which to compare themselves with other schools and that observing the inequalities among schools may be the first step to redressing them. Burgess et al (2011) elaborate on this 'principal-agent model' arguing that the publication of school performance tables helps to monitor the output of the school and maintains schools' focus on test scores. Whether the publication of results in a locally comparative format is sufficient *in itself* or whether this serves to raise the stakes of government targets is unclear at present. For example, there is the notion that high profile targets may act as missions around which employees may coalesce (Dewatripont, Jewitt & Tirole, 1999). Detailed qualitative and mixed methods research (Diamond & Spillane, 2004; Perryman, Ball, Maguire & Braun, 2011) has begun to unpick how high stakes accountability drives school behaviour. As this work progresses, hopefully it will become clearer how the public availability of test scores can drive school behaviour in *both* a positive and negative fashion.

A further issue that needs to be tackled concerns the tensions inherent in using a single assessment, such as the GCSE, for multiple purposes, including certificating student attainment and school accountability (Newton, 2007; Pring et al., 2009). It is an opportune moment to debate these issues in England, because there are plans to reform the national qualification sat by 16 year olds alongside the consultation on reforms to the accountability system. Originally, the purpose GCSE was to certificate student achievement and only later did it come to be used for school accountability purposes. If the primary purpose of the new qualification is school accountability, then this needs to feed into the design process. In correspondence with the Department of Education, Ofqual (Stacey, 2012) has pointed out that accountability measures require highly reliable tests with minimal susceptibility to marking challenge: something that is notoriously difficult to achieve with the unstructured essay questions currently used in written

examination in subjects such as English. Piloting the new qualifications therefore seems highly desirable, especially as one of the lessons from GCSE English is that any piloting would need to be delivered in the same context as that in which the qualification would continue to be delivered. This is because of the difficulty in anticipating user behaviour in any kind of non-live pilot. However, even live pilots cannot be expected to reveal all the potential issues that may occur in a high stakes accountability system.

Some of the issues identified may be addressed by reforming details of the particular measures currently in use. The average grade over the best eight subjects should be considered as an alternative to the 5 A\*-C measure. Arguably, this measure will encourage schools to focus on all students, rather than those on the C/D borderline. Statistical issues with current measures also require urgent attention as part of the reform effort. Although the provision of more information about schools is presented by policy makers as self-evidently beneficial to parental choice, there remain significant issues with the presentation of the current measures. Existing research has identified a number of issues with the value added measure. All the same, the measure continues to be widely reported in school league tables in the media. Given the low functionality and comprehensibility of the measure this is unfortunate: unless these issues can be resolved, it may be more appropriate to abandon the measure for school choice purposes. At the very least, parents need a clear explanation of the level of uncertainty and the issues around the temporal lag of the data on which the measure is based. Nevertheless, value added may still have a role in hierarchical accountability, such as providing one measure on which Ofsted may gather information on a school's performance. Even in this context, issues with comprehensibility would need to be addressed as even those in the education field struggle to understand the measure (House of Commons Children, Schools and Families Committee, 2009).

Several authors have argued against what they perceive as a narrow focus on test scores, proposing instead a broader focus on accountability and these proposals also require consideration. Sahlberg (2007, 2010) has developed the notion 'intelligent accountability' to describe a model combining internal accountability such as school processes, self-evaluations, critical reflection and school-community interaction, with external accountability built on monitoring, sample-based student assessment and thematic evaluations appropriate to the stage of development of each individual school. One of his central points is that cooperation and networking rather than competition is necessary to create the conditions for young people to develop into 'well- educated and prepared people who possess the knowledge and skills to work in an innovation rich world' (Sahlberg, 2010, p. 61). This is consistent with the argument put forward by West et al. (2011) who argue for a broader focus the child's school experience (educational, social and psychological). This is not only about test results, but also about children's wellbeing and a cohesive society. This broader focus may lead to other forms of accountability becoming prominent. A case in point concerns the UK Government's decision to raise the participation age (Department for Education, 2012b) a policy couched in the discourse of social justice (Woodin, McCulloch & Cowan, 2012). At the piloting stage of this policy, networks of schools worked together in order to explore ways of increasing participation rates. Dialogue of this nature is a key feature of *network* accountability; in which individuals cooperate for shared purposes.

## Conclusion

There is a danger that critics of the current accountability system are portrayed as yearning to return to a golden age in which teachers were free to do as they wished in their own classroom: this is unfortunate and curtails productive discussion. There seems to be broad consensus that some form of public accountability is desirable (House of Commons Children, Schools and Families Committee, 2009; Hursh, 2005b; Pring et al., 2009) – it is the mechanism through which this accountability should take place that requires detailed discussion. It is hoped

that this paper can contribute to this debate at both the theoretical level and at the 'coal face'. By reflecting on the theoretical underpinning of the current accountability system this paper has sought to clarify its goals and rationale. On the micro level this paper has explored some of the limitations of current accountability measures and has presented some possible ways for making these measures less susceptible to subversion and more accurate and fit for purpose.

1 March 2013

## References

- Allen, R., & Burgess, S. (2010a). Evaluating the provision of school performance information for school choice. *The Centre for Market and Public Organisation*, 10, 241.
- Allen, R., & Burgess, S. (2010b). *The future of competition and accountability in education* (2020 Public Services Trust at the RSA).
- Allen, R., & Burgess, S. (2011). Can School League Tables Help Parents Choose Schools?\*. *Fiscal Studies*, 32(2), 245–261.
- Bunar, N. (2010). Choosing for quality or inequality: current perspectives on the implementation of school choice policy in Sweden. *Journal of education policy*, 25(1), 1–18.
- Burgess, S., Propper, C., Slater, H., & Wilson, D. (2005). Who wins and who loses from school accountability? The distribution of educational gain in English secondary schools. *CEPR Discussion Paper No. 5248*.
- Burgess, S., Wilson, D., & Worth, J. (2011). *A natural experiment in school accountability: the impact of school performance information on pupil progress and sorting* (CMPO Research Report). Bristol, UK: University of Bristol, Centre for Market and Public Organization.
- Coe, R. (2008). Comparability of GCSE examinations in different subjects: an application of the Rasch model. *Oxford Review of Education*, 34(5), 609–636.
- Department for Children, Schools and Families. (2009). *Your child, your schools, our future: building a 21st century schools system*. London, UK: The Stationary Office Ltd.
- Department for Education. (2010). *The importance of teaching: The schools White Paper*. London: DfE.
- Department for Education. (2012a). *School and college performance tables: Statement on intent - 2012*. London, UK: Department for Education.
- Department for Education. (2012b). *Raising the participation age (RPA) regulations: Government response to consultation and plans for implementation*. London, UK: Department for Education.
- Dewatripont, M., Jewitt, I., & Tirole, J. (1999). The economics of career concerns, part II: Application to missions and accountability of government agencies. *The Review of Economic Studies*, 66(1), 199–217.
- Diamond, J., & Spillane, J. (2004). High-stakes accountability in urban elementary schools: challenging or reproducing inequality? *The Teachers College Record*, 106(6), 1145–1176.
- Goldstein, H. (2004). Education for all: the globalization of learning targets. *Comparative Education*, 40(1), 7–14.
- Goldstein, H., & Leckie, G. (2008). School league tables: what can they really tell us? *Significance*, 5(2), 67–69.
- Goldstein, H., Rasbash, J., Yang, M., Woodhouse, G., Pan, H., Nuttall, D., & Thomas, S. (1993). A Multilevel Analysis of School Examination Results. *Oxford Review of Education*, 19(4), 425–433.
- Goldstein, H., & Spiegelhalter, D. J. (1996). League tables and their limitations: statistical issues in comparisons of institutional performance. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 385–443.

- Goldstein, H., & Thomas, S. (1996). Using examination results as indicators of school and college performance. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 149–163.
- Green, J. (2010). *Education, Professionalism and the Quest for Accountability: Hitting the Target But Missing the Point* (Vol. 26). Taylor & Francis US.
- Harlen, W., & Deakin-Crick, R. (2002). *A systematic review of the impact of summative assessment and tests on students' motivation for learning*. London: EPPI-Centre, Social Science Research Unit, Institute of Education.
- Harris, D. N. (2011). Value-Added Measures and the Future of Educational Accountability. *Science*, 333(6044), 826–827.
- Hodgson, A., & Spours, K. (2003). *Beyond A levels : Curriculum 2000 and the reform of 14-19 qualifications*. London ; Sterling VA: Kogan Page.
- House of Commons Children, Schools and Families Committee. (2009). *School Accountability*. London, UK: Stationary Office.
- Hursh, D. (2005a). Neo-liberalism, Markets and Accountability: transforming education and undermining democracy in the United States and England. *Policy Futures in Education*, 3(1), 3–15.
- Hursh, D. (2005b). The growth of high-stakes testing in the USA: accountability, markets and the decline in educational equality. *British Educational Research Journal*, 31(5), 605–622.
- Hursh, D. (2013). Raising the stakes: high-stakes testing and the attack on public education in New York. *Journal of Education Policy*, 1–15.
- Husbands, C., Shreeve, A., & Jones, N. R. (2008). *Accountability and Children's Outcomes in High-performing Education Systems: Analytical Maps of Approaches to Measuring Children's Education, Health and Well-being Outcomes in High-performing Educational Systems*. EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.
- Jacob, B. A., & Levitt, S. D. (2003). *Rotten apples: An investigation of the prevalence and predictors of teacher cheating*. National Bureau of Economic Research.
- Leckie, G., & Goldstein, H. (2009). The limitations of using school league tables to inform school choice. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 172(4), 835–851.
- Leckie, G., & Goldstein, H. (2011). Understanding Uncertainty in School League Tables\*. *Fiscal Studies*, 32(2), 207–224.
- Lipman, P. (2013). Economic crisis, accountability, and the state's coercive assault on public education in the USA. *Journal of Education Policy*, 1–17.
- Mansell, W. (2007). *Education by numbers : the tyranny of testing*. London: Politico's.
- Mattei, P. (2012a). Market accountability in schools: policy reforms in England, Germany, France and Italy. *Oxford Review of Education*, 38(3), 247–266.
- Mattei, P. (2012b). Raising educational standards: national testing of pupils in the United Kingdom, 1988–2009. *Policy Studies*, 33(3), 231–247.
- Maughan, S., Cooper, L., & Benefield, P. (2009). *School Report Cards: Review of the use of school report cards internationally*. Slough: NFER.

- Newton, P. E. (2007). Clarifying the purposes of educational assessment. *Assessment in Education: Principles, Policy and Practice*, 14(2), 149–170.
- OECD. (2010). *PISA 2009 Results: What Makes a School Successful? - Resources, Policies and Practice*. Paris: OECD.
- Perryman, J., Ball, S., Maguire, M., & Braun, A. (2011). Life in the Pressure Cooker – School League Tables and English and Mathematics Teachers' Responses to Accountability in a Results-Driven Era. *British Journal of Educational Studies*, 59(2), 179–195.
- Power, M. (2012). Education, professionalism and the quest for accountability: hitting the target but missing the point. *British Journal of Sociology of Education*, 33(4), 621–628.
- Pring, R. (2012). Putting persons back into education. *Oxford Review of Education*, 38(6), 747–760.
- Pring, R., Hayward, G., Hodgson, A., Johnson, J., Keep, E., Oancea, A., ... Wilde, S. (2009). *Education for all: The future of education and training for 14-19 year olds*. London, UK: Routledge.
- Ray, A. (2006). *School value added measures in England*.
- Ready, D. D. (2012). Associations Between Student Achievement and Student Learning: Implications for Value-Added School Accountability Models. *Educational Policy*. Retrieved from <http://epx.sagepub.com/content/early/2012/05/04/0895904811429289.abstract>
- Reback, R. (2008). Teaching to the rating: School accountability and the distribution of student achievement. *Journal of Public Economics*, 92(5), 1394–1415.
- Reynolds, D. (2008). New Labour, education and Wales: the devolution decade. *Oxford review of education*, 34(6), 753–765.
- Sahlberg, P. (2007). Education policies for raising student learning: the Finnish approach. *Journal of Education Policy*, 22(2), 147–171.
- Sahlberg, P. (2010). Rethinking accountability in a knowledge society. *Journal of Educational Change*, 11(1), 45–61.
- Shepherd, J. (2010, April 7). Ed Balls forced to drop key reforms. *The Guardian*. London, UK.
- Smith, A. (2004). Making mathematics count. *Report Commissioned by the UK Government*.
- Smith, G., & Smith, J. (2005). Regression to the Mean in Average Test Scores. *Educational Assessment*, 10(4), 377–399.
- Stacey, G. (2012, November 19). Letter to Michael Gove: English Baccalaureate Certificates. Retrieved from <http://www.ofqual.gov.uk/files/2012-11-19-letter-glenys-stacey-to-secretary-of-state.pdf>
- Stobart, G. (2007). *Testing Times: The Uses and Abuses of Assessment*. London: Routledge.
- Stockford, I. (2012a). *AQA award of GCSE Mathematics B (4365) (Linear specification) - summer 2012*. Manchester: Assessment and Qualifications Alliance.
- Stockford, I. (2012b). *AQA award of GCSE Mathematics (4360) - summer 2012*. Manchester: Assessment and Qualifications Alliance.
- Stockford, I., Meadows, M., Wheadon, C., Pinot de Moira, A., Eason, S., Taylor, M., ... Faulkner, K. (2012). *AQA GCSE English and GCSE English Language January 2011 to*

*June 2012: Report to Ofqual.* Manchester: Assessment and Qualifications Alliance.  
Retrieved from <http://www.ofqual.gov.uk/files/2012-11-02-gcse-english-final-report-annex-a-aqa-report.pdf>

- Taylor, R. (2012a). *Chapter 2 – Early entry to GCSE.* Manchester: Assessment and Qualifications Alliance.
- Taylor, R. (2012b). *Chapter 3 – Analysis of the trends in early entry at GCSE.* Manchester: Assessment and Qualifications Alliance.
- Thomas, S. (1998). Value-added measures of school effectiveness in the United Kingdom. *Prospects*, 28(1), 91–108.
- Thomas, S., Sammons, P., Mortimore, P., & Smees, R. (1997a). Stability and Consistency in Secondary Schools' Effects on Students' GCSE Outcomes over Three Years\*. *School effectiveness and school improvement*, 8(2), 169–197.
- Thomas, S., Sammons, P., Mortimore, P., & Smees, R. (1997b). Differential secondary school effectiveness: comparing the performance of different pupil groups. *British Educational Research Journal*, 23(4), 451–469.
- UK Parliament. (1988). *Education Reform Act 1988.* London, Her Majesty's Stationery Office.
- Verhoest, K., & Mattei, P. (2010). Guest editorial for special issue on welfare governance reforms and its effects. *Public Management Review*, 12(2), 163–171.
- West, A. (2010). High stakes testing, accountability, incentives and consequences in English schools. *Policy & Politics*, 38(1), 23–39.
- West, A., Mattei, P., & Roberts, J. (2011). Accountability and Sanctions in English Schools. *British Journal of Educational Studies*, 59(1), 41–62.
- West, A., & Pennell, H. (2000). Publishing School Examination Results in England: Incentives and consequences. *Educational Studies*, 26(4), 423–436.
- Whitty, G., Power, S., & Halpin, D. (1998). *Devolution and choice in education : the school, the state, and the market.* Buckingham [England]; Bristol, PA, USA: Open University Press.
- Wiliam, D. (2010). Standardized Testing and School Accountability. *Educational Psychologist*, 45(2), 107–122.
- Wilson, D., & Piebalga, A. (2008). Performance measures, ranking and parental choice: an analysis of the English school league tables. *International Public Management Journal*, 11(3), 344–366.
- Woodin, T., McCulloch, G., & Cowan, S. (2012). Raising the participation age in historical perspective: policy learning from the past? *British Educational Research Journal*, 1–19.

**Appendix      Seven forms of accountability**

<b>Type</b>	<b>Who is accountable</b>	<b>Accountable to</b>	<b>For what (example)</b>	<b>Mechanisms (example)</b>	<b>Sanctions (example)</b>
<b>Professional</b>	Teachers	Professional peers, General Teaching Council	Teaching, planning, assessment, moral conduct	Conformity to standards and code of conduct, peer review, self-regulation	Possibly via GTC, peer review
<b>Hierarchical</b>	Headteacher and school governing body	Local authority DfE Ofsted	National test and exam results  Exclusions data  Attendance data  Community involvement  Financial management	Test and examination data  Targets  Ofsted  Financial report	Possibly via Ofsted, LA  Reputational (naming and shaming)  Possible dismissal  Replacement of governors  School closure
<b>Market</b>	Headteacher and school governing body	Parents	National test and exam results  Quality of education	Test and examination data  Ofsted reports  Parent judgement; parent networks	Possibly, via parents  Parent exit  Falling roll (financial; reputational; theoretical school closure)
<b>Contractual</b>	Academy sponsor	Secretary of State	National test and exam results  Exclusions and  Attendance data  Financial management	Test and examination data  Ofsted reports  Legally binding funding agreement	Possibly, via the courts  Withdrawal of contract, threat of closure
<b>Legal</b>	Teachers	Parents and children	Special educational	Legislation, regulations	Possibly, via the courts (financial, criminal)

	Governing body	Teachers	needs	Schools Adjudicator	conviction)
<b>Network</b>	Headteacher and school governing body	Other bodies	Behaviour	Dialogue, internal reporting	Unlikely
		Other schools	Attainment	Peer review	Exclusions from benefits of network
<b>Participative</b>	Headteacher and school governing body	Parents (as individuals and group), pupils, community, staff, local authority	Child's educational progress Holistic educational outcomes Community outcomes	Voice dialogue Responsiveness	Few

Note: This Table is an edited version of one appearing in West, Mattei and Roberts (2011). As the focus of this paper is school accountability, I have excluded details of the way in which local authorities and the Secretary and State are accountable.