Measuring what matters: secondary school accountability indicators that benefit all

Chris Paterson

CENTREFORUM
SUPPORTED BY PEARSON
Measuring what matters: secondary school accountability indicators that benefit all

Chris Paterson
About the author

Chris Paterson is Senior Researcher at CentreForum. He specialises in social policy and education, with a particular focus on social mobility. He was a contributor to the recent publication ‘The Tail – How England’s schools fail one child in five’. He was previously a solicitor at city law firm Slaughter and May.

Acknowledgements

CentreForum would like to thank Pearson for its support. The author would particularly like to thank Emma Whale and Bob Osborne for their invaluable help. Special thanks to experts Gill Wyness and Adam Corlett for their tireless efforts in analysing the data and for putting up with endless questions and pestering.

The author would also like to thank the large number of stakeholders and education experts who generously took the time to participate in the interviews and discussion sessions that informed the project and without which it wouldn’t have been possible. Particular credit is also due to the individuals in our test case local authority who provided the qualitative input that informs the latter part of the report.

Thanks to both Sean O’Brien and Sean McDaniel who both provided vital assistance to the project. Major thanks also to Tom Frostick, Russell Eagling and James Kempton.

The views expressed and any errors are however of course the author’s.

ISBN: 978-1-909274-05-1

Copyright August 2013 CentreForum
All rights reserved

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of CentreForum, except for your own personal and non-commercial use. The moral rights of the author are asserted.
Contents

Foreword – Rod Bristow 4
Foreword – Graham Stuart MP 6
Executive summary 8
Introduction 16

Part 1: Headline accountability measures in principle
1 Secondary school accountability in England
2 Overview of the failings of the established headline measure 31
3 Design principles - what should a good headline measure look like? 33
4 The proposed alternative headline performance measures 39

Part 2: Headline accountability measures in practice
– coherence with overarching objectives
5 Accountability measures and the performance of all 45
6 Accountability measures and closing the gap 56
7 Accountability measures and school performance 63
8 Conclusion and recommendations: Putting progress at the heart of the system 73

Part 3: Headline accountability measures in practice
– curriculum incentives
9 Introduction 79
10 National breakdown of APS 8 measure 82
11 Test-case local authority APS 8 breakdown 89
12 Test-case local authority – progress measure breakdown 96
13 Conclusion 104

Annex 1: Data and modelling 105
Annex 2: Current headline measure vs proposed English and maths threshold measure 108
Annex 3: Progress 8 measure vs current value added measure 109
Foreword
Rod Bristow

Over the last three years, reform and change have been the watchwords of the English education system, with a new government setting a challenging agenda that seeks to raise standards at the same time as creating a more diverse and autonomous school system.

In the conversation about reform, accountability and performance measures are never more than a few words away. Often, league tables have been blamed for driving “perverse” behaviour and encouraging “gaming”. They are alternately presented as the cause of low standards, and the reason why we can’t act to address problems – integral to insuring us against the very worst practice, but also in their crudeness acting as a barrier to achieving the very best for learners.

Pearson has always shared the view that accountability matters. However, we strongly believe that it can be much more than a necessary evil - rather a positive force for good.

The strength of this paper, and earlier work supported by Pearson from Family Lives, is that it sets out much more ambitious goals for accountability, conceiving it not as a means of stopping bad things happening, but incentivising the very best in education.

We should be ambitious enough to create an accountability measure that reflects, as closely as possible, the ingredients that we know come together to make a good school. We should be judging schools by the quality of the learning experience and the start in life that they give to all their learners.

While the proposals set out in this paper are grounded in rich data and analysis, they come together to create a performance measure
which speaks to me as a parent, and as someone who has been deeply involved in education for 25 years. This measure values how far a learner has come, not just where they end up. It recognises the importance of a broad, balanced curriculum - whilst still reminding us (as I have relentlessly told my own children) that without English and maths, everything else is so very much harder.

For these reasons, I am sure that this piece of work will resonate with the wider education community, with parents and with students. I am convinced that its contribution will be important and much valued as debate and policy continues to evolve.

**Rod Bristow**
President, Pearson
Measuring what matters

Foreword
Graham Stuart MP

Incentives matter. Like all individuals and institutions, teachers and schools respond to the priorities established by the system that holds them to account. It is accordingly vital to ensure that those incentives are aligned correctly to deliver the outcomes we desire.

Regrettably, this is not presently the case when it comes to our school league tables. There is widespread evidence that the current headline accountability measure for secondary schools - which ranks schools by the percentage of their pupils who achieve 5 A*-C grades, including English and maths - encourages teachers to focus excessively on pupils at the crucial C/D borderline, at the expense of everyone else. That is where the system focuses. Accordingly, that is where schools focus also.

The government is currently considering how best to reform the school accountability system. In this important report, CentreForum models and tests the potential alternative headline measures. It demonstrates that the move towards judging schools based on the progress they enable every pupil to make could be a significant leap forwards.

However, ministers still want to retain a measure based on the percentage of pupils achieving a C grade in English and maths. It is easy to understand why. English and maths are crucial to a satisfactory education.

Yet continuing to emphasise this by means of another threshold measure would restrict the improvement to a tentative shuffle sideways. The solution would replicate the problem it is trying to resolve.
Accordingly, I am very supportive of the proposal developed by CentreForum. This would remove the threshold measure, but instead give double weighting to the grades pupils achieve in English and maths within a new progress measure. This has the potential to offer the best of both worlds.

If implemented, the incentives acting in the system would match the two admirable objectives that have been set for it: improving outcomes for all and closing the shameful gap between children from disadvantaged backgrounds and the rest.

Accountability change may not be the most high profile of the coalition government’s education reforms but, if this realignment of the incentives can be achieved, it will be the most important.

Graham Stuart MP
Chairman of the Education Select Committee
Executive summary

Secondary school accountability and the ‘good school’

At the heart of any system of secondary school accountability lies a decision about what a ‘good school’ is and does. This is because the real goal of such a system is to judge schools against – and therefore encourage school behaviour towards – this chosen ideal. For government, the accountability framework should thus allow it to monitor and drive up standards for the benefit of both pupils and society as a whole. For parents, it should facilitate informed choice about where best to send their children.

The precise definition of the ‘good school’ is perhaps inevitably a values based judgement subject to competing philosophies. In accountability terms, however, the prevailing definition is heavily shaped by the government of the day. As such, the coalition has identified two core goals for the education system and for the schools within it:

1. to secure the best possible outcomes for all children; and
2. to ‘close the gap’ between pupils from disadvantaged backgrounds and the rest.¹

There is a good degree of cross-party consensus around these overarching objectives.

Beneath this lies the government’s subsidiary aim around which there is greater contention: that these outcomes should be within high value qualifications as part of a broad and balanced curriculum, with a particular emphasis on achieving an ‘academic core’ of subjects for all pupils.²

The accountability framework put in place should therefore seek to align with these qualitative objectives. That is, it should monitor

¹ Michael Gove, Education Select Committee oral evidence, 31 January 2012
school performance against them and drive school behaviour in their direction.

**Headline performance measures**

In England, a central aspect of this accountability framework is the chosen headline measure of secondary school performance. This provides the basis for two profound drivers of school behaviour: minimum ‘floor’ performance targets below which intervention may be triggered and media league tables that rank (and thus partly define) school ‘quality’. As such, these measures are so powerful precisely because they say something about what a good school does.

In the context of educational reform and a desire to improve outcomes, this power presents an opportunity – in effect, what gets measured will generally get done. However, it also carries a corresponding risk if the incentives driven by a chosen headline measure do not point towards the goals of the system in which it operates.

It is now well recognised that this is precisely the problem with the established headline performance measure – the percentage of pupils achieving 5 A*-C grades at GCSE (or equivalent), including English and maths. There is a damaging disjoint between the behaviours it rewards and encourages and the objectives set for a ‘good school’.

This is particularly problematic given the current thrust of reform towards combining higher levels of school autonomy with high levels of accountability. International comparison suggests that this framework can lead to improved outcomes but only if the two are intelligently combined. If schools are to be given greater freedom to improve and indeed compete, the markers by which ‘success’ is defined (and the behaviours they encourage) become even more important.

The aim of this report is therefore to aid the adoption of effective and equitable alternative headline performance measures. The government has recently consulted on proposals to replace the established measure and strike a balance between three new headline measures (see Figure 1). Given the unintended consequences of their predecessor, it is important that

---

4 ‘Secondary School Accountability Consultation’, 2013. Full details of the measures are outlined in Chapter 4.
these proposed measures are modelled and thoroughly tested. This report will use 2012 school data to do precisely that. Specifically, it will do so in relation to the objectives identified for a ‘good school’.

**Figure 1 – Summary of proposed alternative headline performance measures**

The consultation outlines proposals for three alternative headline measures:

1. A threshold measure capturing the percentage of pupils achieving A*-C GCSE grades in the core subjects of English and maths.

2. A points based attainment measure in which every grade for every pupil in up to their best eight qualifying subjects is given a score and the combined total for each pupil is averaged to give an overall school score (**APS 8**). The combination of qualification entries that can contribute to this measure is specifically structured to encourage particular curriculum choices.

3. A progress measure that isolates the actual ‘value added’ by a school. In effect, this judges a school by comparing the outcomes it secures for its pupils against the national average performance for an intake with precisely the same level of prior attainment.

The first and third measures will be the most important, forming the dual aspects of a reformed floor target. The main role of the APS 8 measure is to facilitate the third measure – in effect, it provides the output against which pupil progress is to be gauged.

**Findings and recommendations**

1. **Coherence between the proposed measures and the overarching objectives**

   The report first analyses the alternative measures in light of the two core objectives: improving performance of *all* pupils and closing the gap.

   It finds that the reforms have the *potential* to be genuinely
transformative. They contain the prospect of removing the damaging incoherence between the defining goals of the system and the key drivers within it. **In and amongst its more high profile battles, if this can be achieved, it will be perhaps the most positive educational reform of the coalition government.** However, such a positive outcome is *not guaranteed* with - if all three measures are retained – much depending on the actual interaction between them.

**Key findings:**

5: The established headline measure generates an incentive structure that encourages a focus on *some* pupils rather than *all*. Specifically, it actively drives attention *away* from both the most able and from pupils entering secondary school most at risk – those in the underperforming educational ‘tail’.

6: Around 40 per cent of disadvantaged pupils fall into this ‘tail’. The existing measure therefore not only fails to provide an impetus to close the gap, it in fact operates to *widen* it.

7: The proposed new English and maths threshold measure retains precisely the same problematic incentive structure.

6: **In contrast, by rewarding improvements in performance across the board, the proposed progress measure provides a *direct and equal* incentive to raise outcomes for *all*.**

6: It therefore shines a light on those “poor, unseen children” recently identified by Ofsted and hidden by both threshold measures. It will thus work *with* rather than *against* efforts to ‘close the gap’ driven by the pupil premium and by Ofsted itself.

6: By controlling for the impact of prior attainment, the proposed progress measure provides a genuine indication of *school* performance and thus a judgement not of what some of its pupils *get*, but of how much all of its pupils *learn.*

In relation to both core objectives therefore, the proposed progress measure is, by its very nature, a balanced indicator. That is, the performance it reflects and the behaviours it will drive are directly aligned with these central goals. As such, by definition, it cannot

---

5 These can be read alongside the analysis presented in Figure 2 and 3 at the end of the summary.
provide a counterbalance to a measure pulling off in one direction.

Thus, the more prominence given to the proposed new threshold indicator, the more its flawed incentive structure will drive attention towards some and not all. By contrast, the more prominence achieved by the progress measure, the greater the coherence between the drivers in the system and the admirable goals set for it.

Key recommendations:

- The proposed new English and maths threshold indicator should be replaced with an alternative measure that can provide the desirable additional focus on these core subjects without the problems identified above. This should be achieved by double weighting pupil performance in English and maths within the underlying APS 8 measure against which progress is gauged.

- Significant work is required to make the new progress measure accessible to parents and to encourage a culture shift towards judging (and choosing) schools on this basis. This will require active engagement on the part of government. Ultimately, school league tables should be driven by the new progress indicator.

2. Curriculum incentives under the proposed measures

Using the schools in a representative local authority area, the report then examines the relationship between the proposed measures and the specific curriculum related objectives of the present government. It analyses the strength of the curriculum incentives they drive and thus seeks to enable a meaningful debate as to the resulting implications for pupils and schools alike.

This analysis indicates that the specific model of the ‘good school’ the proposed measures point towards is one that not only secures the best possible outcomes for all pupils, it also secures them across a broad and at least partially structured range of subjects. Both the APS 8 measure and the progress indicator it facilitates will therefore be ‘performance’ measures in a broad sense in that they will be sensitive to both the subjects pupils take and the grades they achieve. Schools with higher numbers of pupil entries qualifying under the APS 8 measure and therefore contributing to the school score will generally perform better. As such, schools making early marked improvements are likely to
provide evidence of shifting curriculum decisions as well as enhanced raw pupil outcomes.

Key findings:

: On 2012 data, around 60 per cent of the difference in outcomes between deciles of schools under the APS 8 measure can be explained by differences in the number of qualifying pupil entries and around 40 per cent by differences in pupil performance within their entries. In effect, the curriculum currently being offered by some schools fits considerably better with the measure than that presently being – quite legitimately – offered by others.

: Of this variation arising from different patterns of entries, around 70 per cent is attributable solely to differences between schools in relation to uptake of ‘core academic’ subjects (i.e. EBacc).

: At a pupil level, three quarters of the large gap between scores for disadvantaged pupils and their more affluent counterparts in the EBacc ‘section’ of the APS 8 measure is the result solely of the tendency to enter fewer of the former group for these subjects.

The intended incentive for schools to encourage uptake of ‘an academic core of subjects’ is therefore likely to be a reasonably strong one. This will be, to an extent, an extension of a journey many schools have already begun. Testing also reveals, however, the clear potential for schools to achieve good scores under the measures based primarily on strong pupil performance in high value vocational and creative subjects.

Key recommendations:

: The ‘correct’ position with regard to curriculum incentives is perhaps inevitably contentious. However, it is only by establishing the operation of the measures in practice that a meaningful debate can be had. Having done so, the government should consult on the appropriate balance in the context of the decision about the points-per-grade scale to be used under the new measures.
Figure 2 – Reflection of a good school?8

This graph divides pupils into deciles of prior attainment on entry (1 as the lowest). It then illustrates the differing pattern of contributions made by these pupils to an average school score under each of the alternative measures.

Both the existing and proposed threshold measures provide a judgement about a school that heavily reflects the performance of some pupils (i.e. those with higher levels of prior attainment) more than others. Given the disproportionate presence of disadvantaged pupils in the underperforming ‘tail’, this judgement also significantly under-represents the performance of those from disadvantaged backgrounds. By contrast, the proposed progress measure provides a judgement about a school that reflects the performance of all pupils equally.

8 Details of full contributions analysis can be found in Chapter 5.
Figure 3 – Driving good school behaviour? 9

This graph represents the different incentive structures created by the alternative headline performance measures.

It again divides pupils into deciles of prior attainment on entry. It then isolates the statistical pattern of immediate incentives operating in a school in relation to each measure. In effect, the pupils represented are those where a school can most readily improve (or protect) its institutional score with the least outlay of effort – that is, where it can get the biggest return on its allocation of resources.

This illustrates the ‘race to the middle’ incentive structure of both threshold measures that actively drives attention away from both the underperforming ‘tail’ (with a disproportionate number of disadvantaged pupils) and the most able. By contrast, the proposed progress measure provides an equal reward to a school for any improvement in the performance of any pupil – i.e. it is entirely neutral or balanced as to the allocation of resources it encourages schools to make. It shines a light on all, including those in the ‘tail’ and those from disadvantaged backgrounds.

9 Details of the full analysis – based on the impact of hypothetical alterations to pupil grades – can be found in Chapter 5.
Introduction

“The government should not underestimate the extent to which the accountability system incentivises schools to act in certain ways with regard to exams. Sometimes these may be in students’ interests; sometimes, however, they are not.”

Education Select Committee, 2012

“Poor, unseen children can be found in mediocre schools the length and breadth of our country. Good schools ensure that no child remains unseen. They shine the spotlight on these children and bring them out of the shadows.”

Sir Michael Wilshaw, 2013

Headline measures of secondary school accountability

Education is too important to be delivered without scrutiny. Parents, the government, further and higher education institutions and employers all have an interest in and a right to know about the success of our schools. Fundamentally, however, systems of school accountability exist to protect and enhance the life prospects of pupils themselves.

In this context, the role and use of headline measures of secondary school performance are perhaps inevitably controversial. Almost by definition, they are reductive: an attempt to distil the complex myriad of things that a school ‘does’ – and how well it does them
Measuring what matters into a limited number of outputs or indeed perhaps a single number. This is of course, in one sense, an impossibility. There will always be much that goes on behind the school gates that no quantitative measure – however well designed – can capture. It is important therefore that such measures form only one limited aspect of a broader framework of intelligent accountability (with the regulator – Ofsted – at its heart).

However, it is also clear that in England, headline measures – along with the high stakes floor targets and media league tables that such indicators facilitate – are now an inescapable feature of the educational landscape. Parents need (and want) access to clear and effective ‘snapshot’ markers of school performance to aid choice. Government too requires accessible levers and benchmarks to both flag up struggling schools and drive standards right throughout an increasingly autonomous and decentralised system.

As such, any chosen headline measure will continue to have a profound impact. As a recent report for BIS puts it: “schools’ educational accountabilities are largely measured by and therefore defined by their position in comparative league tables” which, in turn, “strongly influence parental choice of schools and schools’ reputation”. Thus:

“In such a climate it must be recognised that accountability measures will directly influence the behaviour of schools, particularly related to the curriculum and qualification pathways they develop for their students.”

An insight into the impact of the established ‘5 A*-C including English and maths’ measure is evident from the attempt in Wales to abolish performance tables based upon it. The result has been a significant drop in school outcomes against the indicator - an average of 3.4 percentage points in each school. In fact, there is evidence that this prevailing measure has come to have a major influence on almost every aspect of contemporary schooling: from the subjects and qualifications schools teach and the pedagogy they employ to the resources they allocate and even the children they prefer to admit.

15 ASCL response to DfE ‘Secondary School Accountability Consultation’, 1 May 2013
None of this is to question or malign the motivations of individual teachers – as Graham Stuart puts it, “If you create the framework, you create the incentives; don’t blame the people in the system if they follow the incentives.” Nor is it to suggest that the accountability framework is the only factor influencing decisions. It is, however, to recognise the significant impact that the operating incentive structure will have on the outcomes that follow: simply put: “if you do not understand the power of incentives, you will not understand the behaviour in the system”.

Project aim

In the context of the government’s welcome decision to reform the headline accountability measure of secondary school performance, it is therefore imperative to understand and address these issues head on. This project marks an attempt to do just this and thus aid the adoption of effective and equitable alternative headline performance measures.

It is now a well recognised public policy phenomenon that with such core targets, what gets measured generally gets done. The resulting susceptibility to ‘gaming’ behaviours must be guarded against. However, in the context of driving educational improvement and reform, this behaviour-changing power also marks a potential strength, providing such measures are appropriately designed.

This requires (and indeed, presents an opportunity for) a return to first principles: we must first establish what it is that we want ‘done’ and then assess the suitability (or otherwise) of potential alternative measures in direct relation to this.

What will a good headline measure do?

The starting point for this project is that the accountability framework put in place – and the drivers within it - should align with the key objectives set for the system. As such, as an important aspect of this framework, a well designed headline measure should have two overlapping but distinct attributes:

1. it will judge schools against – and thus reflect - what a good school does;
2. it will drive school behaviour in the direction of what the government and society want a good school to do.

---

17 Education Committee meeting, ‘The Responsibilities of the Secretary of State’, 31 January 2013.
18 Ibid.
For too long the education system has been shaped by a headline measure of Key Stage 4 (KS4) performance that does neither of these: the percentage of pupils in a school achieving 5 A*-C grades at GCSE (or equivalent) including English and maths (5 A*-C EM). As such, the perverse incentives driven by this measure have come to have significant negative effects on the system as a whole and for certain groups of children in particular (perhaps most notably, those in the underperforming educational ‘tail’).  

**Project methodology**

This project has therefore taken place in two parts.

The first phase, drawing on the recognised failings of the current measure, involved an extensive series of interviews with key stakeholders. These included prominent secondary head teachers, representatives from leading academy chains, teaching unions, Ofqual, Ofsted, the CBI, academics, school governor organisations and other educationalists. These discussions were carried out prior to the launch of the government consultation with a view to identifying and influencing the overarching design principles that should underpin alternative headline measures of KS4 accountability.

In February 2013, the government launched its ‘Secondary School Accountability Consultation’ containing its proposed alternative headline accountability measures. The detail of these measures is set out in Chapter 4. Their potential implications, however, can only be partially grasped in the abstract. A central problem with the current measure has been the failure to thoroughly anticipate the unintended behaviours it would reward and drive *in practice*. Thus, as ASCL note in their consultation response:

“It is essential that all new accountability measures are thoroughly considered, modelled and tested to ensure different but equally perverse incentives are not introduced.”

The second phase of the project has therefore involved precisely this: modelling and testing the proposed measures in relation to actual school results for 2012.

---

Definition of a ‘good’ or ‘effective’ school

The key question however remains: what is the criteria against which the measures should be analysed? The complex issues of values in education, the purposes of schooling, the quality of students’ educational experiences and of what constitutes a ‘good school’ rightly remain the subject of much argument and are unlikely to be resolved easily.\(^{21}\) Similarly, as identified in the academic field of School Effectiveness Research, “effectiveness is not a neutral term” – it always requires choices among competing values which are perhaps inevitably the subject of political debate.\(^{22}\)

In accountability terms, however, the prevailing definition is heavily shaped by the government of the day. Thus, Michael Gove has given a clear and concise outline of the two central goals the government wants the system – and the schools within it – to achieve:

“*There are two things that we need to do in education:*

1. *raise standards overall for all children; and*

2. *close the gap between those children who come from poorer backgrounds and those who are fortunate enough to grow up in more comfortable backgrounds.*\(^{23}\)

These priorities constitute “the two principal drivers of education reform.”\(^{24}\) Indeed, in terms of these broad objectives, there is a relatively strong degree of cross-party consensus.\(^{25}\)

The first aim of this report is therefore to address the crucial issue of the coherence between the potential alternative headline measures and these defining objectives. Do the proposed drivers actually align with these key goals?

Beneath this, the coalition also has a subsidiary aim around which there is greater contention: that the outcomes achieved should be within high value qualifications as part of a broad and balanced curriculum, with a particular emphasis on achieving an ‘academic core’ of subjects for all pupils.\(^{26}\) The second aim of the report is therefore to examine the curriculum centred aspect of the proposed measures and the strength of the related incentives they will actually

\(^{21}\) White and Barber (eds.) *Perspectives on School Effectiveness and School Improvement*, Institute of Education (1997).


\(^{23}\) Education Select Committee oral evidence, 31 January 2012.

\(^{24}\) Ibid.

\(^{25}\) Andy Burnham in ‘Labour league tables would scrap ‘five good GCSEs’ measure’, TES, 15 July 2011.

create in practice. This can, in turn, facilitate a meaningful debate as to the resulting implications for pupils and schools alike.

**Report structure**

The report is therefore split into three parts, which can be read either together or independently.

**Part 1 – Headline accountability measures in principle**

- The role of headline performance measures within the context of the wider accountability framework and the thrust of current reforms is discussed (Chapter 1).
- A brief overview of the failings of the established headline measure is provided (Chapter 2).
- Drawing on the stakeholder interviews, key design principles for alternative headline measures are identified (Chapter 3).
- The proposed measures contained in the consultation are then outlined in detail with reference to these design principles (Chapter 4).

**Part 2 – Headline accountability measures in practice: coherence between the proposed measures and the overarching objectives**

- Using 2012 school data, the coherence between the alternative measures (old and new) and the first core goal – improving the performance of all – is analysed (Chapter 5).
- The coherence between the alternative measures (old and new) and the second core goal – closing the gap – is then analysed (Chapter 6).
- The resulting extent to which the measures enable a reliable means of monitoring and comparing school performance is then examined (Chapter 7).
- Conclusions and recommendations are made (Chapter 8).

**Part 3 – Headline accountability measures in practice: curriculum incentives and the proposed measures**

- Using 2012 school data, the detailed operation of the proposed measures is analysed on a national level, with a particular focus on the curriculum related incentives they drive (Chapters 9 and 10).
This is then broken down to the level of the individual institution with an in depth analysis of the operation of the proposed measures and their implications for the schools in a representative test-case local authority (Chapters 11 and 12). This process is supported by qualitative analysis based on discussions with the relevant individuals within the local authority with detailed knowledge of the schools in question.

Conclusions and recommendations are made (Chapter 13).
Part 1: Headline accountability measures in principle
1 Secondary school accountability in England

“PISA results suggest that, when autonomy and accountability are intelligently combined, they tend to be associated with better student performance.”

OECD, 2011

The purpose of school accountability systems

The basic premise of the academic study of school effectiveness is that “schools matter, that schools do have major effects upon children’s development and that, to put it simply, schools do make a difference”. The same premise lies at the heart of any system of school accountability.

Accordingly, such systems exist not only to serve the legitimate desire of parents, government and other stakeholders for information about school performance but also, fundamentally, to safeguard and enhance the educational prospects of pupils themselves. Indeed, there is clear evidence to suggest that accountability boosts overall performance. In comparing different education systems across different countries, the OECD finds that external accountability has a major positive impact on how well children do, with particular benefits for disadvantaged and minority groups whose performance is systematically underrated by internal assessment.

---

28 This chapter draws on work published previously by CentreForum in ‘School choice and accountability: putting parents in charge’ (2011) which addresses some of these areas in more detail.
School accountability in England

Such was the opacity of the maintained school system in England during the post-war decades that it came to be known as ‘the secret garden’: a place to which only teachers were granted access. Decisions on virtually all aspects of education rested with individual schools themselves, with no formal criteria against which institutional performance could be measured (and thus meaningful comparisons made).

This position only began to change dramatically following the 1988 Education Reform Act - introducing regular national testing and performance reporting (giving rise in turn to the publication of media league tables) – and the creation of a national inspectorate (Ofsted) in 1992. These two acts therefore put in place the institutional architecture of the accountability framework we have today and meant that, for the first time, standards could be systematically measured and failure identified.

With the inadequacy of large numbers of schools subsequently exposed by the unflattering light of media scrutiny, politicians had no option but to act. And none did so with greater zeal than Tony Blair, who campaigned in the 1997 general election on a “zero tolerance” platform, promising to “wage war” against failing schools. The age of high-stakes teaching and testing had firmly arrived. Along with it, however, came a much noted tendency towards micro-management from the centre.

Thus, in just 20 years, England’s maintained school system had moved from a model of ‘high autonomy, low accountability’ to the precise opposite. This model, however, is now changing again.

The coalition is engaged in a process of far-reaching supply side reform through mass ‘academisation’ and granting schools greater freedoms across the board. As former Chief Inspector of Schools, Christine Gilbert puts it:

“The school system in England is currently experiencing the most significant period of change for a generation. Schools and school leaders are being offered unprecedented levels of autonomy. With greater freedom comes the expectation that schools and school leaders will be the primary drivers of systemic improvement.”

31 The term ‘secret garden’ was first used by Lord Eccles, Minister of Education, in 1960.
Decentralisation, diversity, choice and indeed competition are the watchwords. Thus, for Michael Gove:

“I think it is a good thing if schools feel a sense of friendly rivalry towards one another, and I have noticed that in areas like Hackney, where you have had academies seeking to compete against one another in order to perform well, that has driven standards up and meant that you have had schools that were previously underperforming and losing pupils become high-performing and oversubscribed”.

As such, the direction of travel is towards a ‘high autonomy and high accountability’ model (see Figure 4). International comparison provides some strong support for the efficacy of such a structure, as identified by the OECD. However, as noted at the head of the chapter, this is only the case if the two are intelligently combined. As Gilbert again puts it:

“The key question is this: how should the current accountability system evolve to support a more autonomous and self-improving system?”

Figure 4 – The changing dynamic between secondary school autonomy and accountability

<table>
<thead>
<tr>
<th>Period</th>
<th>School autonomy</th>
<th>School accountability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post-war era</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>1988 – 2010</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Coalition (direction of reforms)</td>
<td>High</td>
<td>High</td>
</tr>
</tbody>
</table>

Current accountability framework

Secondary schools thus now sit within a multi-faceted accountability framework, flowing both upwards to government (so called ‘administrative accountability’) and downwards to parents (‘market accountability’).

Administrative accountability operates primarily through three key

---

35 Education Select Committee oral evidence, 31 January 2012.
institutions: Ofsted (via the inspection regime), the Department of Education itself (primarily through performance measures, tables and targets) and local authorities (now to a lesser extent). Each of these retain the power to intervene when they judge schools to be failing (sometimes referred to as ‘consequential accountability’).  

These interventions range from the provision of additional resources and support to more drastic measures, with the closure of a school the ultimate sanction.

Market accountability in turn operates primarily through the lever of parental choice. By making use of a variety of sources of information about the attributes and performance of potential schools, parents exercise an accountability pressure by voting with their (or indeed, their children’s) feet. This is intended to become an increasingly prominent factor in light of the decentralising thrust of current reforms.

This external framework is also complemented by schools own important internal accountability regimes, driven primarily by the professional responsibilities of teachers themselves and by school governors (and increasingly, in the case of academies, by sponsors). See Figure 5 below.

**Position and role of headline accountability measures**

Headline performance measures – the focus of this paper – therefore constitute one important aspect of this accountability framework.

One of the explicit thrusts running through the coalition reforms is an express intention to rebalance the accountability framework in the direction of market accountability:

“We will make direct accountability more meaningful, making much more information about schools available in standardised formats to enable parents and others to assess and compare their performance. And, through freeing up the system, we will increase parents’ ability to make meaningful choices about where to send their children to school.”

---

39 Although not possible to explore adequately in a report of this length, striking the right balance between these different aspects within a layered accountability system is a crucial issue requiring further consideration (and one frequently raised in our stakeholder interviews).
Measuring what matters

Michael Gove is right that increased transparency will provide “an infinite number of ways in which the data can be sliced and diced”\(^ {41}\). However, the firm reality is that any enhancement of the centrality of parental choice will in turn place greater importance on the chosen headline performance measure(s) – that is, the chosen shorthand encapsulation of what constitutes a ‘good school’. This is particularly so if the opportunity for informed choice is to be a meaningful reality for all parents, not just those most likely to engage in the detailed ‘slicing and dicing’ on the DfE website.

Indeed, parental demand is unequivocal. 96 per cent want information on school performance and 87 per cent want to be able to assess this performance in comparison with other schools.\(^ {42}\) Similarly, while keen to see information across a range

\(^{41}\) Education Select Committee oral evidence, 31 January 2012.

of measures, parents are also strongly in favour of having an overall, accessible rating by which to judge schools. Attempts to make a wider range of indicators readily available to parents (such as the recently launched Ofsted ‘data dashboard’) can only be a good thing. However, the continuing centrality of any designated headline measure(s) – not only through the league tables it drives and the position on school banners it occupies but also as a clear and concise focal point – should not be underestimated.

Notably, this is also the case from an administrative accountability perspective, again particularly so in an increasingly decentralised system. Headline accountability measures provide the potential for rising floor targets and thus the lever for driving minimum standards. However, government reliance on such measures extends far beyond this. As a recent TES editorial puts it:

“It is not just publicity that gives the main GCSE measure its huge significance, but also the importance that the government and Ofsted attach to it when assessing schools.”

In effect, the headline measure is used as a mechanism for monitoring and driving up standards right throughout the system. It provides, in Michaels Gove’s words, an “anchor measure”. If the world of ‘free floating’ secondary schools in which the centre has relinquished many of its other direct levers does fully materialise, this will only become more important.

As such, while offering potentially distinct things to both parents and government, headline performance measures occupy an important cross-over space between market and administrative accountability.

The significant impact in Wales of the attempt to remove the teeth from the established measure supports the finding that, if appropriately designed, such indicators have a “positive role to play in improving the quality of education”. This brings us back, therefore, to the crucial issue of design:

44 William Stewart, ‘Minister bids to end the ‘gaming’ of league tables’, TES, 28 January 2012.
45 Education Select Committee oral evidence, 31 January 2012.
“The way these school performance data are presented to both policymakers and the public is a crucially important part of the school choice system.”

If secondary schools are indeed to compete as envisaged, the sharper the incentive becomes to exhibit excellence in relation to the prevailing marker of a ‘good school’. It thus, in turn, becomes even more important to get this prevailing marker (and, crucially, the incentives it drives) right. In an attempt to do so, it is illustrative to begin with a measure that is not.

2 Overview of the failings of the established headline measure

“England’s school system has a major incentive problem. For too long we have judged secondary schools by the proportion of children getting five C grades or better. Schools doing badly are investigated, and sometimes closed down. Those doing well rise up the league tables – and become popular choices. This incentive system is as flawed as any seen in finance.”

David Laws, FT, 2013

The full extent of the disjoint between the existing headline measure and the objectives set for the system will be brought out in the data analysis in the following chapters. However, as a brief overview, three key problems can be identified at the outset which provide the context for the discussion of potential replacements. The first two relate to the discussion of the overarching objectives in Part 2; the third to the curriculum specific objectives discussed in Part 3.

1. The measure encourages a focus on some pupils rather than all

By capturing only the percentage of pupils who get over the D/C threshold and lacking any sensitivity to grade changes anywhere else on the spectrum, the measure perhaps inevitably drives attention towards pupils around this borderline. It “judges schools by how well they did with some children, not all children”. This in turn leads to excessive pressure on this boundary and strategic behaviour around it (including patterns of double and early entry).

---

2. It fails to provide a genuine measure of school performance

By simply identifying the pupils that reach a minimum threshold, the measure captures far more about the nature of the intake a school ‘inherits’ than the contribution it makes to their development. That is, it is a marker that fails to make any distinction between school and pupil performance and is thus driven primarily by the latter. In the context of a system of school accountability – in which a reliable means of monitoring and comparing the performance of institutions across the board is required - this is problematic. Schools with lower attaining intakes are at a significant disadvantage, while those with more advanced intakes may be able to ‘coast’ without detection.

3. It drives a tension between the interests of the institution and the interests of individual pupils with regard to qualification choice.

By calculating a school’s score across only five of the qualifications taken by its pupils, the measure encourages a narrow (rather than a broad and balanced) curriculum. Furthermore, within this, it has also often encouraged schools to prioritise those subjects and qualifications that will present the school in the best light (i.e. those deemed easiest to score well in), even if these may not be best suited to the needs and interests of pupils.
3 Design principles - what should a good headline measure look like?

“If we get it right, a new headline performance measure would align the political imperative to measure how schools are doing with the professional vocation of teachers to make a difference for every child.”

CentreForum interviewee, 2013

The recognition of the flaws with the established measure set the context for our interviews with key stakeholders that took place prior of the launch of the government consultation. These were carried out with a view to identifying and influencing the design principles that should underpin effective and equitable alternative headline measures of KS4 accountability. This chapter outlines and discusses the findings from this process.

Broad design principles

Impact and role of headline measure(s)

A clear thrust emerging from stakeholders was that any chosen headline measure(s) will swiftly become a heavy driver of school behaviour. Indeed, it was repeatedly stated by interviewees that the impact of such measures is hard to overestimate. For some, this is an intrinsic flaw with the present high-stakes system that needs to be more successfully mitigated by a general rebalancing of the component aspects of the accountability framework. However, even here, the reality was accepted that – within the prevailing set-up – it is imperative that new headline measures are intelligently designed and tested in relation to both what they reflect and encourage.

In particular, a dominant theme was the need for new measures to ensure that a school is judged against and directed towards improving the outcomes for all of its pupils. This issue was identified
as being of particular importance to prevent those at the bottom being ‘cut adrift’ or ‘written off’.

Headline measures of school performance (along with the performance tables derived from them) fill an important cross-over role between the spheres of administrative and market accountability. An important recurring point was the need to keep the potential distinction between these two audiences in mind when designing the measures and to ensure suitability in relation to both (if at all possible). In effect, the balance required is for a measure / measures that present and encourage a broader conception of ‘what a good school does’ while remaining capable of being made readily accessible to parents to aid school choice.

**Specific design principles**

**Progress**

A second dominant theme was that any headline measure based solely on attainment would be fundamentally flawed in that it would run the inevitable risk of capturing far more about the nature of a school’s intake than what it does with it. This is problematic both from an administrative and a market accountability perspective. An appropriate headline measure of school accountability should be equally adept at identifying those institutions performing well with challenging intakes and at exposing those ‘coasting’ on the back of a strong cohort.50

Thus, as one of our interviewees put it, “the holy grail is a progress measure”:

“I’m keen to see a progress measure which is giving schools proper credit for the good that they do. It would best answer the key parental question ‘how much will this school help my child?’”

These sentiments are directly aligned with the findings in the field of School Effectiveness Research (SER), which identifies the major flaw in using raw exam results to make judgements about school performance: it takes no account of the differences in the make-up of individual schools and of the communities they serve.51 Instead:

---
50 This latter point was particularly emphasised by our interviewees.
51 A number of School Effectiveness researchers have demonstrated the need to make adequate control for prior attainment and other intake characteristics in comparing school performance (Nuttall, 1990; Goldstein et al, 1993; Mcpherson, 1993; Scheerens, 1992; Mortimore, 1991; Mortimore, Sammons & Thomas, 1994; Sammons, 1996; Sammons, 2007).
“Natural justice demands that schools are held accountable only for those things they can influence (for good or ill) and not for all the existing differences between their intakes.”

As one of our interviewees put it, “if you are really looking to compare schools, a progress measure is by far the fairest”. Thus, the working definition of an effective school within SER is simply one that “adds extra value to its students’ outcomes, in comparison with other schools serving similar intakes.”

Such a measure – if appropriately designed - would chart the straight distance travelled from a baseline of prior attainment (i.e. Key Stage 2). It would not, importantly, involve any contextual element. Nor need it inherently disadvantage schools with high attaining intakes (as it would compare their performance only with similar schools). Equally importantly, it would not be a mask or an excuse for low performance – quite the opposite if demanding levels of progress are set and expected. It would, however, provide a genuine measure of school performance.

However, a key point raised throughout our interviews was the need to ensure that such a progress measure – being inherently more complex than a basic threshold attainment measure – is made accessible and readily understandable to parents. If this can be achieved, such a measure has the potential to transform the way parents think about the schools they choose:

“I occasionally have conversations with parents who say ‘I’ve put my child down for every grammar school they can conceivably get into’. I think people only say this because of the image they have of a grammar school over a state school [based on headline exam results]. Giving people better information about the progress a school makes would help to break this kind of stereotyping and social segregation down”.

**Attainment**

A measure of attainment will clearly always remain central to any headline accountability measure, be it as a stand-alone or as the output marker against which progress is gauged. However, the key question becomes attainment *in and across what?*

---

54 CentreForum interviewee, 2013
**Breadth, but not too much**

There was a strong sense among our interviewees that a headline measure only capturing performance across five subjects at KS4 is in danger of being excessively narrow. However, it was also felt that the practice of schools “trying to show off how great they are by entering kids for 13 GCSEs” should not be encouraged.\(^{55}\) There was general agreement that around 8 GCSE subjects (or correctly weighted equivalents) was appropriate and that, therefore, the central attainment measure should be based on pupil performance under a form of capped ‘Best 8’ measure.

**Points based rather than a threshold measure**

As identified, the central point emphasised by stakeholders was the need to move to a measure that encourages schools to enhance the outcomes for *all* children. This will require either a removal or a considerable dilution of the strong threshold effects present in the current headline measure.

There was a strong sentiment in our interviews that this could best be achieved by moving towards an attainment measure based on a points per grade system rather than a percentage meeting a particular minimum standard. Under such a measure, each grade advancement on the spectrum would be allocated an increasing points score. This would create a cumulative total for each pupil which can in turn be averaged across the cohort to establish an overall score for a relevant school.

Under such a system, every single grade for every pupil would contribute to the institutional score, thus making it “less susceptible to gaming than the threshold measure”.\(^{56}\) In turn, enhancing the grade of any pupil anywhere on the spectrum would have a corresponding impact on that score. A school would therefore be rewarded for (and incentivised to focus on) both improving the outcomes for those in ‘the tail’ (be it from an F to an E) and for stretching those at the top (be it from an A to an A*).

**Retaining some form of ‘line in the sand’?**

There was a (relatively even) split between our interviewees as to whether, within such a points based system, any additional credit should be awarded to schools in relation to pupils achieving C

---

55 CentreForum interviewee, 2013
grades (perhaps in particular subjects).

For some, the very purpose of moving to a points score was precisely to remove the incentive around this threshold and, as such, the grade point system should be strictly linear (i.e. based on equal jumps in the points allocated to each grade). For others, however, the recognition that, on an individual pupil level, reaching this minimum benchmark does retain additional significance (in terms of post-16 progression and employer recognition) should be factored into the framework of institutional accountability. As such, they advocated the possibility of giving some limited degree of additional weighting within the points measure to the jump from a D to a C grade (i.e. a non linear scale).

However, there was very little support in the latter camp for achieving this additional weighting by retaining any form of straightforward threshold measure. Indeed, those advocating the non linear points increase around the D/C borderline were acutely sensitive to the need to limit any such reward to prevent simply reintroducing the problems with the established measure. The clear concern was that, as one interviewee put it, “as long as we have any form of threshold measure we will have these perverse incentives”.

Subject choice and weighting

There was a full spectrum of opinion from the stakeholders we interviewed as to the extent to which the attainment measure should be used to drive particular curriculum decisions.

There was strong agreement that English and maths should be compulsory components in some form. There was also strong (although not unanimous) support for giving them additional weighting within any measure to reflect the fact that, as one interviewee put it, they are “the most important subjects for later life chances, particularly for those from poorer backgrounds and at the lower end of the ability range”.

Beyond this, however, there was no clear consensus. For some, pupils should be allowed to select any combination of other GCSE or high value vocational subjects as they see fit. For others, science (as either a double award or three individual subjects) should also be compulsory. For others again, EBacc subjects generally were of additional significance and their take-up should be somehow incentivised under any new measure. Indeed, some interviewees

57 CentreForum interviewee, 2013
58 CentreForum interviewee, 2013
discussed the possibility of attributing a full range of different weightings to different subjects and qualifications to reflect perceived differences in value on the labour market. The general opinion, however, was that there should remain scope for pupils to choose some vocational or creative subjects in accordance with their interests.

**Conclusion**

The overriding preference therefore was for a central attainment measure based on a capped ‘Best 8’ point score within which English and maths are compulsory (and potentially given additional weighting). Within this there is also the possibility of attaching some *limited* degree of weighting to the D/C grade point boundary jump to recognise the significance of this for the individual learner. Similarly, there is the possibility of introducing a mechanism to encourage the take-up of EBacc subjects within the Best 8 (but without squeezing out the possibility of pupils selecting other subjects of particular interest).

**Other possible measures?**

A further issue was whether any perceived ‘softer’ indicators – such as pupil attendance, staff retention, parent satisfaction surveys and pupil exclusion rates – could and should be factored into any *headline* accountability measures. The feeling among stakeholders was that these are particularly important issues to parents and have a bearing on school choice. However, it was also generally felt that assessment of such issues was better suited to the qualitative domain of Ofsted rather than the being factored into the quantitative remit of headline performance measures. This was frequently put in the context of the need for such measures to remain clear and accessible to parents.

However, one area where there was significant interest was in relation to the development of viable destinations measures. Although it was generally agreed that the indicators of this type that presently exist are not, as yet, sufficiently sophisticated and robust, there was a strong sense that they could in future come to play an important role. As a marker of school success against genuine pupil *outcomes* rather than using exam *outputs* as proxies, appropriate destination measures have the potential to tie judgements about schools more closely to the desirable societal outcomes they are intended to facilitate.
4 The proposed alternative headline performance measures

“The accountability system must work in tandem with, rather than against, teachers’ aim to help all their pupils acquire the skills and qualifications they need to succeed in future.”

Secondary School Accountability Consultation, 2013

In February, the government announced its intention to abandon the current headline measure in light of its ‘perverse incentives’ that ‘distort teaching and narrow the curriculum’. In its place, the consultation outlined alternative proposals for “the headline accountability measures we intend to use to hold schools to account”.

In keeping with the overarching objectives, the stated driving thrust behind the reforms is the recognition that “the accountability system should recognise the achievements of all pupils”. The consultation outlined proposals to achieve this by striking a balance between three principal accountability measures. Each of these is detailed below in the context of the design principles identified in the previous chapter.

**Measure 1: English and maths threshold**

The first measure – intended to reflect the particular importance of English and maths within the curriculum – remains a threshold indicator based on the percentage of pupils in a school achieving a C grade (or better) in both of these subjects. This measure would form one half of a reformed floor target to benchmark minimum acceptable performance (below which intervention may be triggered).

---

60 Ibid
61 Ibid., 10.1.
62 Or the equivalent threshold grade under the new grading system once confirmed.
The measure is intended to encourage “extra focus on pupils who struggle in English and mathematics” in recognition that, in these core subjects, “secondary schools should place particular importance on making sure all their pupils leave with high value qualifications”. The retention of a decisive ‘line in the sand’ aspect is designed to reflect the particular value of achieving the ‘C pass’ in these subjects to the prospects of the individual learner.

This measure therefore reflects well the strong theme from our stakeholder interviews that English and maths – as the most important subjects for progression at this stage – should play a central (and accentuated) role within any headline accountability measures. However, achieving this by retaining a ‘cliff-edge’ threshold indicator raises issues in terms of the optimal design principles identified above (a point explored further in Part 2).

**Measure 2: ‘average points score 8’ (APS 8)**

The second key measure is a capped total points score calculated across each pupil’s performance in (up to) their best 8 subjects and then averaged to provide a score for the institution (hence ‘average points score 8’). In itself, this measure will not be an aspect of the reformed floor target but it provides the basis for the ‘Progress 8’ measure that will.

Under the APS 8 measure, each grade on the scale is awarded a progressively higher point score. As such, every grade achieved by every pupil in each subject (within the specifications of the measure) will count towards the overall school score. This measure is therefore well in tune with the clear preference among stakeholders for a broad (but not ‘too broad’) measure of attainment that reflects the performance of all pupils and directs schools accordingly (without perverse threshold effects).

However, in line with the subsidiary objective identified at the outset, the proposed measure is also designed with the express intention of driving particular curriculum decisions. As outlined in Figure 6 below, the subject combination across which pupil contributions are to be calculated is specifically structured. For ease of reference, the measure can be thought of as a total of eight separate slots that can each only be filled by selecting qualifications designated as belonging to one of three different baskets.

---

63 DfE, Accountability Consultation, 5.3.
64 The detail of this point score system will not be clear until decisions on the grading of reformed GCSEs have been confirmed.
The first two slots must be filled with a relevant English and maths qualification (which occupy the first basket). Slots 3, 4 and 5 can be filled only by subjects from a second basket containing all remaining EBacc subjects (i.e. sciences, humanities and languages). The final three slots can then be filled with subjects falling into the final basket, comprising approved high value vocational qualifications (in light of the Wolf review), further EBacc subjects as yet unselected by the pupil and any other GCSE subjects (including creative subjects).

Pupil performance in any qualifications that do not fall into this structure simply do not count towards the accountability measure score generated for the school.

**Figure 6 – Structure of APS 8 measure**

<table>
<thead>
<tr>
<th>Basket</th>
<th>Slot</th>
<th>Subject</th>
<th>Detail</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>English</td>
<td>Compulsory core subjects</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>Maths</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>EBacc 1</td>
<td>Remaining EBacc subjects</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>EBacc 2</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>EBacc 3</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>Other 1</td>
<td>Approved vocational qualifications, further EBacc subjects, other GCSE subjects</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>Other 2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>Other 3</td>
<td></td>
</tr>
</tbody>
</table>

As virtually all pupils will automatically fill the first two compulsory slots, the key area of interaction will be between baskets two and three. For example, a scenario might arise where a pupil is only taking sufficient EBacc subjects from the second basket to fill one of the corresponding slots and yet is taking five qualifications designated to the third basket. In this instance, only the three highest scoring of these final basket qualifications would be allowed to count (i.e. the remaining two scores would be discarded for the
purposes of the institutional score). Thus, by failing to fill the slots allocated to basket 2, this pupil’s total contribution to the school accountability measure would only in fact be calculated based on their performance across six qualifying entries (two compulsory, one ‘EBacc’, three ‘Other’).

The rationale for this structuring is clear from the consultation document:

“In most cases, we think an academic core of subjects should be studied up to age 16... This approach incentivises schools to offer an academic core of subjects to their pupils, by reserving five slots for these qualifications. Including three further qualifications in the measure will reward schools that also offer a broad and balanced curriculum [so] pupils can follow their interests.”65

As outlined above, there was a full (and fairly evenly distributed) spread of opinions among stakeholders as to the extent to which the headline accountability measures should be used to actively drive particular curriculum priorities. Indeed, to an extent, the desired balance here is perhaps one of the ‘values in education’ questions that are inescapably the subject of (small p) political debate.

However, it is important to establish and understand in practice the strength of the curriculum related incentives contained within this new measure. This can in turn facilitate a meaningful debate going forward as to the merits of the particular balance that is struck. This analysis is set out in Part 3.

**Measure 3: ‘Progress 8’**

The final proposed headline measure is a progress measure (which, as currently planned, will form the second aspect of the dual floor target).

The Progress 8 measure will be generated using pupil performance under the APS 8 measure at KS4 as the key output variable. It will isolate the value added by a school by charting the ‘distance travelled’ by its pupils from their position on entry (KS2) to the level they reach at KS4 as defined by this APS 8 indicator. The degree of progress made is then benchmarked against the national average for pupils with precisely the same levels of prior attainment (to see whether the school is adding more or less value than this average).

---

65 DfE, ‘Accountability Consultation’, 5.7.
Importantly, school progress outcomes will therefore also be directly shaped by the curriculum related implications of the underlying APS 8 measure – the use of a different output measure could give a significantly different picture of the ‘distance travelled’.

As the consultation puts it, such a measure “helps to make judgements about schools fair” as it “give[s] schools credit for helping all their pupils, whatever their starting point” – high performing schools with challenging intakes will be rewarded, coasting schools exposed.66 The overwhelming preference from our stakeholder interviews was that such a measure should be designed and given prominence (and indeed, pre-eminence).

The importance of testing

Thus, viewed as a whole from the perspective of the overarching design principles, the proposed alternative headline measures present a significant improvement. However, the key lesson from the outgoing measure is that it is application in practice (including any potential unintended consequences) that matters. Indeed, much of the necessary detailed debate on the new measures is difficult to have in the abstract. It is crucial to understand how the measures will in fact operate in relation to real schools. Thus, as David Laws himself puts it, “for each of these things we need to test what behaviour it is likely to drive and will it drive good behaviour”.67 The following chapters mark an attempt to do precisely this using 2012 school data.

---

Part 2: Headline accountability measures in practice – coherence with overarching objectives
5 Accountability measures and the performance of all

“Most people know that incentives matter. Appropriately designed incentives spur people to work well. Badly designed incentives lead to bad things happening.”

David Laws, 2013

As established, the coalition has set two core objectives for the education system and the schools operating within it:

- to secure the best possible outcomes for all children; and
- to ‘close the gap’ between pupils from disadvantaged backgrounds and the rest.

The aim of this part is to use 2012 school data to analyse how well the alternative headline accountability measures – as key drivers within the system – succeed in aligning with these defining goals. To do so, it will assess the extent to which the measures (both established and new) do two distinct things:

- judge school success against (i.e. reflect) these objectives; and
- drive school behaviour towards these objectives.

This chapter will use this framework to analyse the coherence with the goal of securing the best outcomes for all. The next chapter does the same in relation to the goal of closing the gap.

The incentives discussed operate at a school level. However, to ensure sufficient sample size, the analysis is based on all pupils in our representative test-case local authority area. In effect, it is treated as one ‘super school’ with over 5,000 pupils (with an FSM level and current headline measure performance closely equivalent to the national average).

---

69 For reasons discussed below, the specific local authority used has been anonymised. The picture it presents is closely representative of that at a national level and it is therefore presented in this way as an approximation of the average school.
**Established measure – 5 A*-C EM**

The analysis in this section is based around variations on the two graphs presented below. Both divide pupils into deciles of prior attainment on entry (with 1 as the lowest).

Figure 7 illustrates the respective contribution made by the pupils in each of these deciles to the overall score achieved under the 5 A*-C EM measure. This illustrates the extent to which the judgement the measure makes about a school reflects the performance of some pupils more than others.

Figure 8 isolates the pattern of immediate incentives being driven by the measure. The analysis carried out plots the combined impact of a hypothetical increase and decrease of the grades of all pupils by one grade in each relevant subject (i.e. the fluctuation that can be made / prevented with the least effort). The resulting graph shows the percentage of pupils in each decile whose altered performance would actually have an impact on the overall school score.

Thus, if a school is to follow the incentives driven by the measure it would be encouraged to focus attention on these pupils as they are potentially most likely to impact on the outcome generated for the institution. In effect, the pupils represented are those by which the school can most readily enhance or ‘protect’ its overall score – that is, where it can get the *biggest return on its allocation of resources*.70

**Overall impact**

Taken together, these graphs provide a picture of the glaring disjoint between the established measure and the first core objective. The powerful impact of prior attainment in relation to the threshold benchmark gives a clear illustration of the extent to which it “judges schools by how well they did with some children, not all children”.71

Particularly striking, however, is the overall incentive pattern created. Combining the total pupils represented within each of the deciles in Figure 8 (i.e. those whose altered performance affects the school score) indicates that under the current measure, the immediate incentives drive attention towards less than half of the cohort. That is, changing the grades of *more than half* (51%) of pupils either up or down one (in all relevant subjects) has no impact whatsoever on the institutional outcome.

---

70 On the working assumption that each grade advancement is of roughly equal difficulty.
Figure 7 – Reflection of a good school? Percentage contributions to 5 A*-C EM measure

Figure 8 – Driving good school behaviour? Percentage of pupils affecting school score when changing grades by +/- 1
Thus, the very nature of the measure risks undermining the basic notion that ‘every child matters’.

**Impact on ‘The Tail’**

As identified in the recent CentreForum publication, *The Tail*, a key challenge for our schools is the need to enhance the performance of the lowest achieving 20 per cent of pupils. These children routinely leave school lacking basic literacy and numeracy skills, facing significant barriers to full participation in society. Indeed, for David Laws, this group represents “the biggest challenge in English education”, while for Lord Adonis, “educating the tail is the key priority for our schools”.

As Figure 7 indicates, however, the performance of pupils in the bottom two deciles of prior attainment is significantly underrepresented in the judgement provided by the existing headline measure. Indeed, this group contributes only 4% to the overall school score. That is, the measure fails to capture the performance of 80% of the pupils in the tail at KS2 – they are truly, as Ofsted puts it, *unseen*.

Moreover, as Figure 8 highlights, the measure actually actively drives focus away from these struggling pupils. Even if the school were to raise every grade these children achieve by one, this would have no impact whatsoever on the institutional score in almost four out of five cases. This contrasts directly with the picture for prior attainment deciles 5 and 6, where the relevant increase or decrease would impact on the institutional score in relation to 80% of pupils. This indicates just how strong the incentive is for a school looking to improve its headline score to focus attention and resources towards the middle.

Unsurprisingly, therefore, the performance of lower attaining children appears to suffer significantly as a result:

> “Nationally the average grade achieved by this tail was a low E and the worst school averaged only a grade G. This suggests a very variable level of interest in achieving high progress for children whose results are unlikely to contribute to headline school measures.”

---

This point is made particularly powerfully in Machin and Silva’s analysis of the impact of academisation on the results of early converting schools.\textsuperscript{77} This revealed that, while post-conversion performance rose overall, there was no improvement whatsoever in the performance of the tail. Given the pressure on these schools to show a quick and clear upturn in performance against the headline indicator, it seems very likely that the incentives identified above were a central factor in this.

Thus, for those in the tail, the current measure has the potential to operate as an ‘anchor’ in a very different way, keeping them firmly down in the “shadows” identified by Sir Michael Wilshaw.\textsuperscript{78}

\textit{Impact on high attainers}

Although those in the highest deciles of prior attainment contribute heavily to the overall score, this is largely because achieving the minimum ‘C’ grade is relatively easy given their advanced starting point. As is evident from Figure 8, however, the established headline measure also provides a limited impetus to stretch high ability pupils to reach their full potential.

Thus, altering the grades of almost 80 per cent of the pupils in the top two prior attainment deciles has no impact on the overall school score. Once again, therefore, performance suffers: in 2012, 65\% of the most able students on entry (those achieving level 5 at KS2) – 65,000 pupils – failed to obtain A* or A grades in English and maths. More than a quarter (27\%), failed to get better than a C.\textsuperscript{79}

There is therefore a danger that, as noted by the Sutton Trust, the measure “leav\[es\] the highly able as a peripheral issue”.\textsuperscript{80}

\textit{Impact of threshold pressure}

A further result of this ‘race to the middle’ incentive structure is that the all importance of the D/C grade boundary – and of the pupils around it – can result in excessive strategic or ‘gaming’ behaviours. As Ofqual highlights:

\begin{quote}
\textit{“The trend [is] of running Years 10 and 11 as a tactical operation to secure certain grades and combinations of grades. This}
\end{quote}

\begin{footnotes}
78 ‘Unseen Children’ speech, Church House, 20 June 2013.
79 Ofsted, “The most able students: Are they doing as well as they should in our non-selective secondary schools?”, June 2013.
80 Ibid
\end{footnotes}
has come to be seen as ‘what good schools do’ despite the awareness of many GCSE teachers and parents that the concept of broad and deep learning can get lost along the way.”

This in turn puts heavy stress on qualifications which can ultimately “buckle under the pressures of accountability”.  

It has led to increasing patterns of double entry, with schools looking to maximise their score by putting pupils in for multiple exams in the same (generally core) subjects. Similarly, the lack of sensitivity as to the performance of high attaining pupils has led to the common practice of early entry with a view to ‘banking’ the C grade. Here again, the interests of the institution conflict with and take precedence over those of the individual learner (who would be better served by entering the qualification later in the year and aiming for an A or B grade).

Proposed alternative headline measures

The explicit rationale behind the proposed reforms is, as David Laws puts it:

“to replace targets which distort, such as the present 5 A*-C targets, with more intelligent accountability which gives incentives to stretch all pupils.”

The actual coherence between the proposed measures and this first core objective can therefore be assessed by repeating the analysis carried out above.

Figure 9 illustrates the differing contributions made by pupils in different prior attainment deciles to the institutional score generated under each measure. For ease of comparison, these contributions have been simplified and presented in terms of their relative pattern (i.e. normalised). Figure 10 again isolates the pattern of immediate incentives driven by each of the measures.

---

81 Ofqual, ‘GCSE English 2012’, p.2  
82 Ibid.  
83 Ofsted, ‘The most able students’, 2013.  
84 David Laws speech delivered to ATL Annual Conference 2010, 29th March 2010.
Figure 9 – Reflection of a good school? Normalised contributions to accountability measure

Figure 10 – Driving good school behaviour? Percentage of pupils affecting score when changing grades by +/- 1
Overall Impact

Taken together, these graphs provide a clear picture of just how much of an improvement the measures have the potential to be. However, they also reveal that this will depend heavily – if all three are retained – on the interaction between the measures themselves. The individual indicators have very different potential impacts and thus the overall incentive structure created will be directly shaped by the degree of prominence that each comes to hold.

Progress 8

As highlighted in Figure 9, under the proposed Progress 8 measure, the contribution made by each decile of prior attainment is directly proportional. That is, the resulting judgment on a school it provides is based equally on the performance of every single pupil within the institution. Similarly, as Figure 10 identifies, altering the performance of every pupil has a direct and equal impact on the institutional score. Thus, the school benefits from any improvement it can make in the performance of any pupil – not only does every child matter, every child matters equally. The measure therefore drives behaviour accordingly - that is, it is neutral or balanced as to the allocation of resources it encourages schools to make.

APS 8

As the combined representation in Figure 10 suggests, the position in terms of driving behaviour is the same in relation to the APS 8 measure (as modelled). As a points based measure with equal grade score jumps, the institutional total is equally affected by any grade improvement (or fall).85

However, as a measure of raw attainment with higher scores for higher performing pupils, the contribution to the overall score made by the different deciles is still significantly skewed towards those with higher prior attainment. Thus, while the points system ensures that the school score does reflect the performance of all pupils, it reflects that of some more than others (with resulting implications for the judgement it provides about schools with greater and lesser numbers of these pupils).

Notably, if a non-linear grade point scale was to be used (potentially with a view to introducing a more subtle recognition of the additional value of the ‘C pass’ to the individual learner), the position in relation

---

85 Ignoring for these purposes any impact of the anomaly of the enhanced jump from a U (0) to a G (16).
to the incentive structure would also shift towards these higher attainers. In any such staggered points system, improvements towards the upper end of the scale would become more beneficial to an institution and would accordingly create a resource related incentive under this measure. This would not, however, change the operation or incentive structure of the Progress 8 measure which the APS 8 measure facilitates.

\textit{A*-C EM}

By contrast – and unsurprising given the degree of crossover – the English and maths threshold measure retains precisely the same problematic patterns as the established measure. Again, by definition, the performance of large numbers of pupils does not contribute to the overall score. More strikingly, the immediate incentive structure it generates continues to drive attention towards only half of the cohort. That is, changing the grades of the other half (49\%) either up or down one in both core subjects has no impact whatsoever on the institutional score.

\textit{Impact on ‘the Tail’}

The contrasting implications of the proposed measures – particularly the two intended floor target indicators – is especially striking in relation to the key issue of the tail.

Notably, the incentives inherent to the new threshold measure continue to actively drive attention \textit{away from} rather than towards those entering secondary school most at risk. As Figure 10 illustrates, directly improving the grades of almost three quarters of these pupils in these core subjects – that is, the subjects providing the numeracy and literacy skills they so often lack – has no impact on the overall school score. By contrast, for those in prior attainment deciles 5 and 6, the altered performance of 80\% of pupils would ‘count’. Put another way, deploying resources here at the centre would be, from an institutional perspective, around three times as effective.

The stated aim for the measure is to ensure “extra focus on pupils who struggle in English and mathematics”.\textsuperscript{86} The analysis here suggests that there is actually a danger that it may do the exact opposite. The impetus driven by the measure does not align with the first core objective set for the system.

\textsuperscript{86} DfE, ‘Accountability Consultation’, 4.4.
By contrast, the Progress 8 measure both fully captures (and thus judges a school in relation to) the performance of pupils in the tail and ensures that a school is rewarded just as much for improving their outcomes as it is for any pupil on the spectrum. It is worth noting, however, that it does not provide any additional incentive to do so that could be seen to counteract the problematic thrust of the new threshold indicator.

**Impact on high attainers**

The same position exists in relation to high attainers. As Figure 10 illustrates, the proposed new threshold measure continues to provide a limited incentive to stretch the most able pupils in these key subjects. The altered performance of only one in five pupils in the top two deciles of prior attainment actually impacts on the institutional score. By contrast, under the Progress 8 measure, it is 5 in 5.

**Impact of threshold pressure**

In turn therefore, this new English and maths measure is also likely to continue to drive (and indeed, even intensify) undesirable strategic behaviour in these subjects. As the exam body AQA puts it:

“the proposed new threshold measure – a pass in English and maths – will continue to exert the same pressure on schools... the prominence of the threshold measure means that many schools are likely to retain a greater focus on those pupils at the C/D borderline.”

Thus, Ofqual have directly questioned the suitability of this measure - particularly as an aspect of the important floor target mechanism - precisely because of the stress it will continue to apply around the relevant boundary:

“If GCSEs in key subjects are put under too much pressure, they will not reliably measure the knowledge and skills they are designed to assess.”

Again, therefore, there is a potential disjoint between the likely operation of the measure and the intended rationale behind it.

By contrast, the proposed progress measure not only alleviates the pressure around any particular grade boundary, it is also likely in turn to discourage patterns of double and early entry (with a school

---

87 AQA response to DfE ‘Secondary School Accountability Consultation’, April 2013.
88 Ofqual, Open letter to Michael Gove MP, 1 May 2013.
being rewarded for the best grades it can secure for their pupils in a broader number of subjects).

To return therefore to the David Laws quote above, instead of replacing a ‘target which distorts’ with one that ‘gives incentives to stretch all pupils’, the proposals in fact create two separate measures that each do one of these things. That is, the Progress 8 measure aligns directly with the first core goal. The other, the English and Maths threshold, does not.
6 Accountability measures and closing the gap

“Where the spotlight focuses attention on the needs of our poorest children, improvement does follow. They emerge from the darkness of educational failure when we as a country resolve to do something about them.”

Sir Michael Wilshaw, 2013

This chapter analyses the coherence between the alternative measures (established and new) and the second core goal: closing the gap between pupils from disadvantaged backgrounds and the rest. To what extent will the potential key quantitative drivers in the system help to facilitate this?

Established measure – 5 A*-C EM

Once again, there is a clear (and damaging) disjoint between the established headline measure and this second objective. This is apparent from the distribution of the disadvantaged pupils in our test case local authority – around 15% of the cohort (equivalent to the national average) – across the prior attainment deciles (Figure 11).

Almost 40 per cent of the FSM pupils fall into ‘the tail’ (and almost half into the bottom three deciles) – the areas on the ability spectrum significantly underrepresented in the judgement the measure makes about a school. Similarly, as Graham Stuart puts it, the “terrible irony” of the current measure is that it “virtually ensures that attention will be focused not on the lowest performing, which sadly we know are typically the poorest, but will instead divert attention away from them.”

This is clear from a comparison of the pattern of the distribution of disadvantaged pupils with that of the incentives operating under the current measure (Figure 12).

89 Sir Michael Wilshaw speech, Church House, 20 June 2013.
90 Graham Stuart quoted in TES, ‘League tables receive a blow ... from the right’, 13 January 2013.
Figure 11 – Distribution of disadvantaged pupils (% falling into each prior attainment decile)

Figure 12 – Normalised FSM distribution against 5 A*-C EM incentive structure

The areas where a school can most effectively maximise its institutional score with the least effort (deciles 5, 6 and 7) – i.e. those with the greatest potential for return on resources allocated – are
those with significantly lower proportions of FSM pupils. Not only, therefore, does the existing headline marker of a ‘good school’ not provide an impetus towards closing the gap, it in fact generates an incentive structure that operates to widen it.

This issue is particularly problematic in schools with small but significant proportions of FSM pupils. As Figure 13 identifies, schools with 35% or more FSM children perform better for these children than schools with between 5% and 21% (i.e. schools that can achieve comfortable headline scores without a significant contribution from their disadvantaged pupils). As the recent Ofsted *Unseen Children* report has shown, it is precisely in these latter schools that, collectively, the largest numbers of FSM pupils are educated. As such:

“It is too easy to lose sight of disadvantaged pupils. In too many secondary schools, the poor performance of pupils from low income backgrounds is masked by the generally strong performance of other pupils.”91

**Figure 13 – KS4 attainment by school FSM band 2012**

Source: Department for Education data

---

In a recent interview, David Laws attacked the “disgrace” of “outrageously low” performance of FSM pupils in some of the leafiest parts of the country.\(^9^2\) He insisted that “schools need to understand that they cannot hide behind good headline figures if they fail a large cohort of pupils”.\(^9^3\) One of the major problems with the established measure is that this is precisely what they have been able to do.

### Proposed alternative headline measures

The coherence between the proposed measures and this key objective can be analysed using the same framework. Figure 14 compares the proportion of FSM pupils within the cohort to the respective contribution they actually make to the institutional score under each of the measures.

Figure 15 then breaks this down to the individual learner level, illustrating the average contributions made by each FSM pupil against those of each non-FSM pupil. It is only under the Progress 8 measure that a school is judged on the performance of all of its FSM pupils in direct relation to their actual numbers.

### Figure 14 – Reflection of a good school? Percentage contribution made by FSM pupils

\(^9^2\) David Laws quoted in Grice, A., ‘Schools in well-off areas ‘are failing’ poorer pupils - who get better exam results in deprived areas’, The Independent, 23 April 2013.

\(^9^3\) Ibid.
Interestingly, Figure 15 also points to the fact that the mechanics of the alternative headline measures actually matter more for disadvantaged pupils. The contribution of the average non-FSM pupil is relatively consistent under each of the measures. By contrast, the raw attainment measures each – to differing degrees – mask some element of the performance of the disadvantaged pupils in the cohort.

However, particularly striking again is the relationship between the distribution of disadvantaged pupils and the incentives driven by the proposed new floor target measures. Figure 16 illustrates this by superimposing this FSM distribution onto the incentives analysis presented in the previous chapter.
Figure 16 – Driving good school behaviour?
Normalised FSM distribution against accountability measure incentive structures

Not only does the proposed new threshold measure fail to provide any impetus to close the gap, it in fact continues to drive a pattern of incentives that may hold down the performance of disadvantaged pupils and thus perpetuate (or indeed widen) the gulf.

By contrast, the incentives generated by the Progress 8 measure again align directly with this second core objective by providing an equal (but not additional) reward to a school for advancing the prospects of disadvantaged pupils. Crucially, it will therefore shine a light on those “poor, unseen children” masked by both threshold measures. It will thus work in tandem with the thrust of other key policies such as the pupil premium and the move by Ofsted to only award ‘Outstanding’ status to schools that also perform outstandingly for their disadvantaged pupils. Indeed, as such, this issue reaches beyond the Department’s specific aim and goes to the very heart of the drive to improve social mobility – the “principal goal of the coalition’s social policy”.  

---

Figure 17 – Coherence with overarching objectives

Objective 1: Securing the best outcomes for all

1. judgement about school heavily reflects the performance of some pupils more than others
2. drives school behaviour heavily towards a focus on some not all

Objective 2: Closing the gap

1. judgement about school significantly underrepresents performance of disadvantaged pupils
2. drives focus away from significant numbers of disadvantaged pupils

1. judgement about school proportionately represents the performance of disadvantaged pupils
2. drives focus towards disadvantaged pupils just as much as all other pupils
7 Accountability measures and school performance

“[Headline performance measures] too often reward and sanction principals and teachers for outcomes over which they have little control. Not surprisingly, this situation results in frustration, conflict and a lot of strategic behaviour.”

The significantly different contributions made by different groups of pupils to the judgement the established measure makes about a school goes to the heart of the second key problem identified in Chapter 2: it captures far more about the nature of the intake a school ‘inherits’ than the contribution it makes to their development. That is, it is a marker that fails to make any distinction between school and pupil performance and is thus driven primarily by the latter. In the context of a system of school accountability – in which a reliable means of monitoring and comparing the performance of institutions across the board is required - this is problematic.

Building on the analysis above, this chapter therefore examines the extent to which the alternative measures can provide this marker of school performance.

Established measure – 5 A*-C EM

The extent of the link between prior attainment and school outcomes under the established measure is evident from the correlation graph in Figure 18 below.

This analysis seeks to compare two variables and gauge the strength of the relationship between them. In doing so it generates a correlation co-efficient as a measure of this strength – the closer the figure is to 1, the greater the interconnection between the variables.

Thus, in the graph below, each of the points represents a maintained mainstream school in England, plotting its score under the current headline measure (on the x axis) against the average KS2 points score of its pupils on entry. There is a clear and very strong relationship between the two (with a correlation co-efficient of 0.8).\(^{97}\)

**Figure 18 – Relationship between 5 A*-C EM and prior attainment**

Similarly, from Figure 19, we can see that there is also a clear inverse relationship between levels of disadvantage and school performance under the measure. That is, schools with lower proportions of FSM pupils (y axis) tend to exhibit higher headline figure scores (x axis).

\(^{97}\) The closer the correlation coefficient is to 1, the stronger the relationship. As a general rule, a score of 0.8 - 1 constitutes a very strong relationship, 0.6 – 0.8 a strong relationship, 0.4 - 0.6 a moderate relationship, 0.2 – 0.4 a weak relationship, 0 – 0.2 a very weak relationship.
Simply put, schools with lower performing intakes (and thus frequently those in deprived areas) are at a significant, inherent disadvantage under the established headline measure. It is for this reason that ASCL have rightly raised concerns over “accountability measures which may damage the capacity or confidence of schools serving particularly less privileged communities”. By contrast, “entire schools with middle-class intakes can coast, safe in the knowledge that enough of their pupils would get the five C grades”. 99

As identified at the head of the chapter, this gives rise to a scenario where schools are in fact being held to account against a measure that is only loosely associated with what they have the power to control.

Indeed, our research has identified examples of ‘Outstanding’ rated schools with scores in the top decile under the existing value added measure that actually fall beneath the average score on the existing headline measure (and thus in the bottom half of the league table it drives).

---

Proposed alternative headline measures

This analysis can be repeated in relation to each of the proposed alternative headline measures. Thus, Figures 20, 21 and 22 plot the strength of the relationship between each school’s average KS2 score on entry (y axis) and their institutional outcome under the indicators.

**Figure 20 – Relationship between A*-C EM and prior attainment**

Fit (correlation = .8122)

**Figure 21 – Relationship between APS 8 and prior attainment**

Fit (correlation = .8625)
Once again, there is a very strong relationship between a school’s performance under the proposed new threshold indicator and the level of prior attainment of its intake (with a correlation coefficient of 0.8). As the APS 8 measure is also a measure of raw attainment, the same is true (indeed, the relationship is moderately stronger at nearer to 0.9).

The picture is very different in relation to the Progress 8 measure, with only a weak relationship between school performance and intake level (0.39). The fact that a weak relationship does exist is a result of the fact that the underlying APS 8 measure is influenced by pupil entry patterns as well as raw performance. Schools with higher levels of prior attainment also tend to be the ones providing a curriculum offer to all of their pupils (including those lower down the attainment spectrum) that leads to greater levels of qualifying entries and thus higher APS 8 scores. The lower attaining pupils that do attend these schools also therefore receive a slight entries driven boost as a result. These issues are explored fully in Part 3.

---

100 There is thus an even weaker relationship between school prior attainment and a progress measure calculated across a straight ‘best 8’ that does not contain the curriculum related stipulations – i.e. one driven purely by performance.
A similar pattern emerges below from the analysis of the relationship between each measure and the level of disadvantage within a school (Figures 23, 24 and 25).

**Figure 23 – Relationship between A*-C EM and disadvantage**

![Graph showing the relationship between A*-C EM score and percentage of pupils eligible for Free School Meals. The correlation is -.5516.](image)

**Figure 24 – Relationship between APS 8 and disadvantage**

![Graph showing the relationship between APS 8 score and percentage of pupils eligible for Free School Meals. The correlation is .6475.](image)
Figure 25 – Relationship between Progress 8 and disadvantage

There is a clear (and relatively strong) inverse relationship between the FSM level within a school and the outcome for that school against both proposed raw attainment driven accountability measures (0.55 and 0.65 respectively). That is, schools with lower proportions of FSM pupils are likely to achieve higher scores.

Again, this relationship is much weaker in relation to the Progress 8 measure. The weak relationship that does exist is also the result of the curriculum aspect of the underlying APS 8 measure - schools with lower proportions of FSM pupils also tend to be those with a curriculum offer that leads to higher qualifying entries for all pupils and thus higher scores across the board. However, this impact attributable to entry patterns is not intrinsically the result of the prior attainment of the pupils and is something that will be within the power of the institution to influence.

Thus, simply put, schools with both low performing and high disadvantage intakes are again at a significant, inherent disadvantage under both proposed raw attainment measures. By contrast, those at the opposite end of the scale have the potential to coast on the strength of the cohort they inherit. In this sense, both measures continue to judge schools in relation to factors they do not fully control. By contrast, the Progress 8 measure removes this inherent (and potentially warping) unfairness. As such, it provides
a measure of school accountability that is a genuine indicator of school performance.

**Impact within a ‘high autonomy, high accountability’ framework**

This point is increasingly important within the context of a decentralising secondary school system. If schools are indeed to ‘float free’ and compete for pupils, this only further raises the stakes as to competition in relation to what? If, as they will, league tables are to continue to be a primary means of manifesting excellence (and thus appeal to parents) and these tables continue to be driven by a measure that does not in fact reflect or differentiate the quality of schools as institutions, this will become even more problematic. The sharper the competition, the sharper the incentive to follow the drivers created by whatever becomes the shorthand marker of a ‘good school’.

An area where this has the potential to manifest itself is in relation to school admissions. This issue has already been much debated in the context of the devolution of power towards individual schools and the process of academies becoming their own admission authorities. Thus, Machin and Vernoit demonstrate that the post-conversion KS4 performance of early academies improves significantly against the established headline measure.\(^{101}\) However, this also corresponds with a sharp increase in the ‘quality’ of their intake. Indeed, they find that “schools that gain the largest increase in autonomy experience the greatest increase in their pupil quality and the greatest increase in their [KS4] pupil performance”.\(^{102}\)

This has raised concerns over a possible association between schools having greater control over their admissions and unfair, exclusionary practices with a particular impact in terms of social segregation.\(^{103}\)

However, it is important not to overemphasise the point at this stage. The Admissions Code militates against any such exclusionary practices and strenuous efforts have been made in relation to academies to ensure equity in relation to intake. Indeed, as the first wave of academies were generally high deprivation schools, the willingness of parents from more affluent backgrounds to send

---

102 Ibid., p.37.
their children to places previously considered ‘sink schools’ is a good thing.

The general point, however, is to again highlight the importance of the direction in which the incentives line up. The scenario to be feared is that outlined by one of our interviewees of a school that:

“was one of the worst 8 schools in England and it sits in a poor white area, one of the most deprived wards in London. Now you give it a shiny new multi-million pound building but the cohort and their backgrounds remain the same - so for many years as an Academy it was still achieving poor GCSE results. Now a new headteacher took over and changed their pupil intake, focusing on attracting new pupils from out of the area, and results improved. But all you are doing there is improving a school by excluding two-thirds of the kids who would have normally gone to that school - exactly the pupils you are supposed to be helping.”

In itself, competition within the system is a good thing that has the potential to drive up standards. However, if the mechanisms by which success is defined (i.e. the basis on which competition takes place) are flawed, then it raises the prospect of even more powerful warping effects.

It is here again therefore that the Progress 8 measure – if given sufficient prominence - has the potential to be genuinely transformative. By removing the powerful effect of prior attainment and judging schools in relation to the value they add, it could shift the very nature of competition with regard to school choice. In effect, it makes the important move away from the question ‘What do the pupils at this school get?’ towards the key parental issue: ‘How much do the children at this school learn?’ This would again work with rather than against the thrust of key policies such as the pupil premium (which was specifically intended to make disadvantaged pupils more attractive to schools).

It also has a further potentially powerful knock on effect in light of the recognised importance of high quality teaching (which has a particular impact on those from disadvantaged backgrounds). Many good teachers - quite understandably – will often want to teach in schools with a ‘good’ reputation and thus a higher degree of status.

104 CentreForum interviewee, 2013. Our analysis of supports this assertion with the school experiencing an increase in its KS2 scores of around two and a half times the national average in relation to its KS4 cohorts in 2010 and 2012 (matched by a sizeable increase in KS4 performance against the current headline measure).
Shifting one of the key markers of a ‘good school’ away from being the sole preserve of institutions with the highest performing intakes would therefore again be a very positive step.
8 Conclusion and recommendations: Putting progress at the heart of the system

The proposed measures – a focus on ‘more’ but not ‘all’?

The proposed reforms therefore have the potential to be genuinely transformative. Within the package of new headline measures lies the prospect of redressing the perverse incentives generated by the established measure and thus removing the damaging incoherence between the defining goals of the system and one of the key drivers within it. **In and amongst its more high profile battles, if this can be achieved, it will be perhaps the single most positive education reform of the coalition government.**

However, the above analysis also makes it very clear that this outcome is not guaranteed. Indeed, the extent to which it is reached will depend directly on the implementation of the measures and the actual incentive structure that results from the interaction between them in practice.

There is reference in the consultation document to the three measures combining to strike a ‘balance’.\(^{105}\) However, in relation to the two core objectives – raising standards for all and closing the gap – the Progress 8 indicator is itself a balanced indicator. That is, the performance that it reflects and the behaviours it is likely to drive are directly aligned with these central goals. As such, by definition, it cannot provide a counterbalance to measures pulling off centre in either direction.

Thus, the more prominence given to the proposed threshold indicator the more its flawed incentive structure will drive attention towards *some* and not *all* (with a damaging impact on those

---

entering secondary school most at risk of failure). In turn, the more truth there will be to the conclusion reached by ASCL:

“It is highly likely that the issues and pressures that the consultation proposals seek to avoid are likely to remain.”\(^{106}\)

By contrast, **the more prominence achieved by the Progress 8 measure, the greater the coherence will be between the drivers in the system and the admirable goals that have been set for it.** In turn, the more light will shine on the performance of pupils in the tail and from disadvantaged backgrounds.

### Putting progress at the heart of the system

This issue of the interaction between the measures – assuming all three are to remain – must therefore become the crucial point of focus. Which measure will become the most pronounced driver of behaviour in the system? Which measure will be chosen as the principal basis for league tables with the profound implications for parental choice and school behaviour? Which measure will be ‘up in lights’ on the banner at the school gates? Which measure (if any) will, in short, become the shorthand marker of a ‘good school’ that permeates general debate, informs the public and influences media coverage, politicians and even policy makers themselves?

The implications arising from the varying incentive structures that could prevail depending on the balance struck are too important to be simply left to fall as they may. They require active thought and engagement.

Indeed, the default position is perhaps likely to be one that needs to be changed. A form of threshold indicator has been the established answer to each one of these questions since the inception of performance monitoring over 20 years ago. It is deeply ingrained. As such, its new incarnation – if retained – may simply take over. It is simple, accessible and familiar. Moreover, it can be generated and conveyed to media outlets swiftly.

The concept of the Progress 8 measure is, by contrast, far less familiar (particularly to those outwith the education sector). In terms of its calculation, it is also far more complicated and, partly as a result, there is presently a time delay in collecting the data and calculating the score.

Furthermore, as presently presented (as a score of above or below

the benchmark of 1,000) it is less accessible to parents. Indeed, individuals taking part in our stakeholder discussions with direct experience and expertise suggested that there is – as a result - often real difficulty in getting the public to engage with value added measures which, from the perspective of media outlets, actually creates a commercial incentive to stick with a simple raw attainment measure as the basis for league tables.

As such, there is a risk – despite the real potential for radical improvement – that the actual change to the incentives operating in schools could still be limited unless further steps are taken.

**Replacement of the English and maths threshold indicator**

The most desirable first step would be to replace the proposed new threshold indicator (either altogether or as a floor target lever) with an alternative mechanism that can provide the desirable additional focus on English and maths without the problems identified above. Indeed, as demonstrated, rather than providing the intended “extra focus on pupils who struggle in English and mathematics”, the incentives it drives may actually operate to shift resources away from those who struggle the most.

Similarly, due to the potential warping effects of the increased threshold pressure, Ofqual has urged the government to consider “alternative [floor target] bases which would place less pressure on the most important GCSEs”.  

One such alternative would be to factor the important additional focus on the core subjects directly into the proposed Progress 8 measure. **This could be achieved by double weighting the contribution of these subjects within the mechanics of the underlying APS 8 indicator.** By calculating value added in relation to this modified indicator (APS 8 EM), a measure could be created (and potentially used as a single floor target) that rewards schools for the progress they enable all their pupils to make in all subjects but with a particular emphasis on that made in English and maths (**Progress 8 EM**). Thus, by ensuring that this additional reward in the core subjects also applies in relation to every pupil – including those who struggle most – such a measure would in fact be better placed to achieve the stated aim. In turn, it would also provide an overall incentive structure directly in keeping with the desirable, balanced pattern outlined above.

The option would still exist, if it is felt that some degree of additional recognition of the ‘C pass’ really is necessary, to factor this into the measure in a more nuanced (and thus less corrupting) way through a (marginally) staggered grade point system under the APS 8 indicator. It is important to note, however, that this would potentially alter the incentive structure within the APS 8 measure and thus again raise the importance of the centrality of the progress indicator.

**Making Progress 8 the key measure**

However, even if the English and maths threshold is replaced, a second step is key. If it is retained, this step is even more important. Active engagement is required to make the new Progress 8 indicator accessible, understood and – as far as is possible – the key driver within the system. This will require a significant culture shift towards judging (and choosing) schools on this basis rather than the raw outcomes of their inherited pupils. **Ultimately, school league tables should be driven by the new progress indicator as a genuine measure of school performance.**

Indeed, the importance of this to the degree of coherence between the goals set for the system and the drivers within it require that it is something the Department must directly involve itself with (publicly or otherwise). In Australia, for example, the curriculum authority (ACARA) engage directly to promote constructive use of performance data in relation to high stakes accountability indicators, including working proactively to provide the media with success stories of schools with high value added scores, rather than a focus on raw attainment data. Indeed, there is a strong case that if it does genuinely wish to transform the incentives in the system, the Department may need to soften the (already slightly artificial) line that it is not involved or directly engaged with the production of league tables.

In particular, three key issues will need to be considered and addressed going forward:

1. **Robustness**

In light of potential future reforms to the operation of KS2 tests, it is important to ensure that this input measure from which progress is calculated (i.e. the baseline starting point) is sufficiently sensitive and robust to allow accurate judgements about the value added by secondary schools to be made.
2. **Timing**

Due to the need for more complex calculation and the fact that, at present, value added measures are not generally published until final confirmed exam results are available (i.e. post appeals and any adjustments), there is a danger that school figures under the proposed threshold measure may be available significantly earlier than the Progress 8 score. This needs to be addressed, particularly in the context of the publication of school league tables and the profound influence these have on school behaviour.

3. **Accessibility**

Perhaps most important, however, is the need to make the Progress 8 measure accessible to parents if it is to become a powerful driver of school choice (and thus school behaviour) in a decentralised system.

A key factor in this may relate to presentation of school outcomes under the measure and considerable thought should be given to how this can be most effective. This could potentially involve benchmarking around a more intuitive number (0?) or translating the Progress 8 outcome into a score out of 100. As is evident from Ofsted inspections, simply because the underlying judgement is complex does not mean that the presentation of the outcome needs to be. It may be therefore that some form of simple traffic-lighting or banding system of schools based on the progress measure might be the most effective way of making the measure immediately ‘graspable’.

The key point, however, is to find a way to facilitate the significant shift towards judging (and choosing) schools on the basis of the contribution they make to the development of all, rather than the raw outcomes achieved by some. Put simply, this involves a move away from the question ‘What do the pupils at this school get?’ towards the key issue: ‘**How much do the children at this school learn?**’.
Part 3: Headline accountability measures in practice – curriculum incentives
9 Introduction

“Accountability measures will directly influence the behaviour of schools, particularly related to the curriculum and qualification pathways they develop for their students.”

ASCL, 2013

Beneath the two overarching objectives analysed in Part 2 lies the coalition’s subsidiary curriculum related aim around which there is more contention. Specifically, it wants to encourage schools to offer a broad and balanced curriculum based on high value qualifications, with a particular emphasis on achieving an ‘academic core’ of subjects for all pupils.

As outlined in Chapter 4, the structured nature of the underlying APS 8 measure is intended to incentivise precisely this. This is part of an attempt to address the further problem with the established headline measure – the tension it has created between the interests of the institution and those of individual pupils with regard to qualification choice. As such, it builds upon the reform of the equivalence regime following the Wolf report.

From our stakeholder interviews, there was a full range of opinions as to the extent to which the headline accountability measures should be used to actively drive particular curriculum priorities. Indeed, to an extent, the desired balance is perhaps one of the values based questions that are inescapably the subject of (small p) political debate. However, for this debate to be meaningful, it is important to establish in practice the strength of the curriculum related incentives the new measure will drive. This section will attempt to do just this.

The following chapter will analyse the operation of the key APS 8 measure (against which progress will be gauged) at a national level.

109 DfE, ‘Accountability Consultation’
Chapters 11 and 12 then break this down to the level of the individual institution, with an in depth analysis of the implications of each of the proposed measures for the schools in a representative test-case local authority. This process will be supported by a qualitative analysis based on discussions with the relevant individuals within the local authority with detailed knowledge of the schools in question.

The analysis in all three chapters also draws on a second round of engagement with stakeholders in which these national and local results were presented for discussion. This involved contributions from many of the same individuals who were consulted in relation to the desirable overarching design principles outlined in Part 1 and also a specially convened panel of secondary head teachers.¹¹⁰

**Modelling and testing – important caveat**

Some key points on the modelling process are highlighted in Figure 26, with full details outlined in Annex 1.

However, an important caveat needs to be made clear at the outset. The testing is based on the National Pupil Database 2012 cohort data set (as the most recent available). That is, it is applying the proposed measures to the outcomes of schools that – clearly – had no knowledge of what the measures would reward. It is therefore unfair to base any normative judgements on what follows – the intended incentives the proposed measures will drive were not in place when schools made the relevant curriculum choices for this cohort.¹¹¹ The analysis presented therefore is exclusively a question of ‘fit’ – it highlights the extent to which the curriculum currently being quite legitimately offered by schools is conducive to scoring under the proposed measures. It therefore seeks only to identify the strength of the incentives that will be acting on schools when such measures become ‘live’.

In particular, it should be noted that the application of the post-Wolf 2014 approved qualifications list (upon which the new measures will be based) to 2012 data will therefore create a shortfall for some schools which they will already be in the process of addressing. Similarly, the 2010 introduction of the EBacc indicator is also likely to be already leading to some increase in uptake of these core subjects. As such, the incentives outlined below will mark an extension of a journey that many schools have already begun.

¹¹⁰ The latter generously convened with help from SSAT.
¹¹¹ This is one reason why details of the test-case local authority have been anonymised.
Figure 26 – Key APS 8 modelling points

- ‘Core’ and ‘additional’ science (i.e. the components of combined science) are treated as two separate qualifications filling two slots in the measure.
- English Literature GCSE is counted in the third basket (as suggested in the consultation document).
- The existing linear 6-point-interval grade scale is used ($C = 40 \pm 6$).
- Only high value vocational qualifications on the Department’s approved list for 2014 performance tables have been taken to qualify for the third basket.
- Any approved vocational qualification presently equivalent to multiple GCSEs has been reduced to a 1:1 equivalence (as will be the case from 2014).
10 National breakdown of APS 8 measure

Although not in fact itself operating as a floor target mechanism, the biggest new driver of behaviour in the system as presently proposed will come from the operation of the APS 8 measure (both as a free standing indicator and, more importantly, as the output indicator against which progress will be judged).

This chapter therefore seeks to break down the operation of this key measure on a national level, with particular regard to the issue of entry patterns. A reminder of the proposed structure of the APS 8 measure is presented in Figure 6 on page 38.

Overview

Applied to 2012 data, the average score under the new APS 8 measure across maintained mainstream schools in England would be 284 (or 61 per cent of the possible maximum). More interestingly, under the mechanics of the proposed ‘slot’ based structure, the contribution of the average pupil would be based on performance across 6.8 qualifying entries. Indeed, only just over half of pupils (52 per cent) would presently fill all 8 slots.

These entry patterns are broken down to the level of the individual baskets in Figure 28 below.

Figure 28 – Average qualifyfying entries by baskets

<table>
<thead>
<tr>
<th>Basket</th>
<th>Average qualifying entries per pupil (max entries)</th>
<th>% of pupils filling all slots</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (E&amp;M)</td>
<td>2 (2)</td>
<td>96%</td>
</tr>
<tr>
<td>2 (EBacc)</td>
<td>2.1 (3)</td>
<td>56%</td>
</tr>
<tr>
<td>3 (Other)</td>
<td>2.7 (3)</td>
<td>80%</td>
</tr>
</tbody>
</table>
As compulsory subjects, the first basket is effectively filled by all pupils. In turn, around 80 per cent of the 2012 cohort filled all three slots in the final basket (even with the exclusion of non-approved qualifications). However, only 56 per cent filled all three EBacc slots (with an overall average of 2.1 entries per pupil). This suggests a clear area with greater potential for development (and score enhancement) for many pupils / schools.

**Breakdown by school performance**

The impact of these entry patterns is more striking when broken down by levels of performance under the measure. Figure 29 divides the analysis into quartiles of school outcomes and shows a breakdown of the average performance of the pupils in these schools. This begins to provide a more specific picture of the curriculum offer that fits well with the proposed measure.

**Figure 29 – Breakdown by quartile of school performance**

<table>
<thead>
<tr>
<th>Quartile (by school APS 8 score)</th>
<th>Average qualifying entries</th>
<th>% of pupils filling all 8 slots</th>
<th>Basket 2 (EBacc) average entries</th>
<th>% of pupils filling basket 2 slots</th>
<th>Basket 3 average entries</th>
<th>% of pupils filling basket 3 slots</th>
<th>Average APS 8 total score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (low)</td>
<td>5.54</td>
<td>23%</td>
<td>1.39</td>
<td>28%</td>
<td>2.26</td>
<td>58%</td>
<td>207</td>
</tr>
<tr>
<td>2</td>
<td>6.60</td>
<td>43%</td>
<td>1.99</td>
<td>48%</td>
<td>2.65</td>
<td>78%</td>
<td>264</td>
</tr>
<tr>
<td>3</td>
<td>7.08</td>
<td>59%</td>
<td>2.34</td>
<td>63%</td>
<td>2.77</td>
<td>85%</td>
<td>298</td>
</tr>
<tr>
<td>4 (high)</td>
<td>7.59</td>
<td>80%</td>
<td>2.71</td>
<td>83%</td>
<td>2.90</td>
<td>94%</td>
<td>353</td>
</tr>
</tbody>
</table>

Simply put, there is a clear relationship between the level of average qualifying entries (both in total and within the individual baskets) and overall school APS 8 score. Thus, while only around 1 in 5 pupils in the bottom quartile of schools are filling all 8 slots, in the top quartile the figure is 4 in 5. While this is partly being driven by a difference in qualifying entries in the 3rd basket (with a gap of 0.6
entries), the difference is particularly striking in the EBacc basket, with the top quartile exhibiting almost double the level of average entries (a gap of 1.3 entries on average).

**A measure of both entry patterns and performance**

This pattern with regard to overall entries is broken down further into deciles of school performance in Figure 30. This identifies a range of a full 3 qualifying entries on average between the top and bottom deciles (4.8 compared to 7.8).

Imposed onto the graph is a line showing the average points per entry for schools in each decile. It highlights that while the average number of entries per decile shows a direct increase, so too does the average points score per entry. This in part reflects the fact that schools with higher levels of prior attainment also generally tend to be those offering a curriculum that fits well with the proposed measure.

**Figure 30 – Average qualifying entries and average points per entry by decile of school performance**

This therefore raises a crucial question: how much are the variations in school scores under the new measure being driven by differences in qualifying entry patterns and how much by differences in pupil performance within the individual entries?
Returning to the breakdown by quartiles in Figure 29, pupils in the top group outperform those at the bottom by 146 points (or 70 per cent). Statistical analysis indicates that 59% of this differential is driven by entry patterns alone and 41% by performance within these entries.\textsuperscript{112} For quartiles 2 and 3, it is 58% by entries and 42% by performance per entry. The variations between different deciles are, in turn, highly consistent: 58% of the differential between both deciles 1 and 10 and between deciles 5 and 6 respectively can again be explained solely by differences in entry patterns.

Notably, the average total number of qualifications taken by pupils in schools across the deciles is in fact remarkably steady (at generally just above 12). The key factor therefore is that, on 2012 data, pupils in schools within the lower deciles are, on average, failing to take a combination of qualifications that result in high levels of qualifying entries under the APS 8 structure. This is because:

- they took more subjects that would be excluded from the proposed measure as they are not on the 2014 approved list; and
- they took fewer EBacc subjects (see Figure 31).

The clear point, therefore, is that both this attainment measure and the progress indicator it facilitates will be ‘performance’ measures in a broad sense in that they will be sensitive to both the subjects pupils take and grades they achieve.

As heavily emphasised by stakeholder and head teacher groups on seeing this analysis, the measures are therefore likely to drive two principal curriculum-related behaviours in schools looking to demonstrate enhanced performance: a move away from qualifications not deemed to be suitably high value, and an increased uptake of core academic subjects.

As noted above, these are both, to an extent, extensions of a journey schools have already begun (particularly with regard to the post-Wolf move towards high-value vocational qualifications). Furthermore, schools will continue to base curriculum decisions on what they genuinely think will benefit their pupils, not simply on how they can maximise their institutional performance. However, with almost three quarters of the 75.8 point (or 148%) EBacc basket differential between schools in the bottom and top quartiles driven purely by entry patterns, the intended incentive for schools to encourage uptake of ‘an academic core of subjects’ is likely to be a reasonably strong one.

\textsuperscript{112} This and all following contributions breakdowns are calculated through a simple decomposition analysis using the product rule.
Breakdown by pupil socio-economic background

In light of the overarching objective of closing the gap, it is also illustrative to analyse the performance of pupils from differing socio-economic backgrounds. Figure 32 therefore breaks down the constituent aspects of this performance for the average FSM and non-FSM pupil in the cohort.

<table>
<thead>
<tr>
<th></th>
<th>Average qualifying entries</th>
<th>% fill all 8 slots</th>
<th>Basket 2 entries</th>
<th>Basket 3 entries</th>
<th>Basket 1 average score</th>
<th>Basket 2 average score</th>
<th>Basket 3 average score</th>
<th>APS 8 score</th>
</tr>
</thead>
<tbody>
<tr>
<td>FSM</td>
<td>5.95</td>
<td>29%</td>
<td>1.59</td>
<td>2.41</td>
<td>67.8</td>
<td>58.7</td>
<td>94.2</td>
<td>220.7</td>
</tr>
<tr>
<td>Non-FSM</td>
<td>6.97</td>
<td>56%</td>
<td>2.25</td>
<td>2.74</td>
<td>81.2</td>
<td>96</td>
<td>119.6</td>
<td>296.9</td>
</tr>
</tbody>
</table>
On 2012 data, FSM pupils would average a full entry less than their non-FSM counterparts and would be around half as likely to fill all 8 slots. Notably, they would be significantly less likely to be entered for EBacc subjects, with an average of 1.59 entries compared to 2.25 for non-FSM pupils.

In scoring terms, the interaction between the impact of qualifying entry patterns and performance within these entries is again apparent. In the compulsory basket 1 (i.e. when FSM and non-FSM pupils are taking the same subjects), there is a 20 per cent gap in scores driven exclusively by performance.

In basket 3 (i.e. ‘Other’), the overall gap is 25.4 points (27 per cent). This gap is driven almost equally by a difference in entries and by a difference in performance per entry (54:46). The points gap widens significantly however in relation to basket 2. Here, non-FSM children outperform their FSM counterparts by 64 per cent. Strikingly, statistical analysis shows that almost three quarters of this difference is driven solely by non-FSM children taking more EBacc subjects. This finding strongly supports the concern raised by the Sutton Trust that disadvantaged pupils are less likely to be entered for core subjects over and above any differential that would be expected based on ability levels.\footnote{113 The Sutton Trust, ‘Research - Summary: Attainment gaps Between the most Deprived and Advantaged Schools’, May 2009, p.4.}

The government – and Michael Gove in particular - have rightly emphasised the need to combat any paucity of ambition for those from disadvantaged backgrounds if the barriers to greater social mobility are to be broken down. To the extent that the Sutton Trust concern is valid, therefore, the new measure provides a clear incentive to address it (and thus ‘close the gap’ in this specific sense).

**Altering the balance between entry patterns and performance?**

The specific balance struck between the reward to institutions for securing high levels of pupil performance and for ensuring that this performance takes place within a particular curriculum is particularly sensitive to the precise modelling of the APS 8 indicator.

For example, there is a degree of uncertainty in the wording of the consultation as to whether ‘core’ and ‘additional’ science could actually be treated as filling only one slot, not two.\footnote{114 DfE, ‘Accountability Consultation’, 5.6.} If so, this...
would significantly increase the gap in entries already faced by many schools in the EBacc basket (and the resulting incentive to alter the curriculum offer to close it). However, if these qualifications are to continue to be of real value, allowing them to fill two slots is both important and equitable if many schools allocating substantial curriculum time to them are not to be severely disadvantaged.

Another subject that will impact on the balance struck is English Literature GCSE. As modelled, this is allocated to the third basket (as per the indication in the consultation).\textsuperscript{115} This is on the rationale that, although arguably an EBacc subject (in that it must be taken by any pupils studying English Language GCSE), a pupil does not in fact need to pass English Literature to secure this meta-qualification as defined. However, for many pupils, the qualification will therefore automatically come to fill one of the third basket slots, leaving less scope for ‘non-traditional’ subjects. Allocating English Literature to the EBacc basket, therefore, would correspondingly reduce the gap in this basket faced by many schools and therefore shift greater weight towards the impact of performance per entry.

A further means of altering this balance would be to change the grade point scale on which the APS 8 indicator operates. The position identified above is based on a linear grade point scale (with equal jumps of 6 points). Altering this towards a non-linear award system that provides increasing reward for grades further up the scale (i.e. a staggered points system) would increase the impact of performance per entry against that of patterns of qualifying entries. However, this would also have implications for the overarching incentives as to the particular pupils schools are encouraged to focus on (as discussed in Part 2).

\textsuperscript{115} Ibid, at 5.6.
11 Test-case local authority

APS 8 breakdown

Test-case local authority

This chapter and the one that follows look at the operation of each of the proposed measures in detail in relation to the schools in our representative test-case local authority area. This area was selected based on its relative size (over 5,000 secondary school pupils) and the broadly representative demographic of its school population (with an FSM level and current headline measure performance closely equivalent to the national average).

The authority has around 30 maintained mainstream schools in total ranging from selective grammars to comprehensives with very challenging intakes. The area is also relatively well advanced in the process of academisation, with over half of the schools discussed below either already converted or in the process of doing so.

The following sections will illustrate the impact of the proposed measures by presenting and comparing mini localised league tables in relation to each measure. This chapter focuses on the attainment indicators, the next on the proposed progress measure. The relative movements of different schools (and the underlying reasons for them) are discussed. This analysis draws directly on extensive input from individuals in the local authority best placed to provide in-depth, day-to-day knowledge of the schools themselves.

For this reason and also due to the caveat about the unfairness of making normative judgements about schools retrospectively, both the local authority and the schools within it have been anonymised. Thus, the name of each school has been replaced with a fictional alternative.116 For ease of presentation and understanding, the tables below have also been simplified to concentrate on the performance of only 16 schools.

---

116 Any resulting similarity between the fictional names chosen and that of any real school is therefore purely coincidental and unintended.
5 A*-C EM vs A*-C EM

The first basic comparison is between the present headline measure and the proposed English and maths threshold measure. However, due to the degree of direct crossover, there is limited movement in the schools’ individual scores (and no alteration in the localised ordering). Indeed, there is in fact very little change at all: very few pupils in these schools are achieving Cs in English and maths but then failing to do so in another three qualifying subjects. Thus, in 13 of the 16 schools below, the percentage of children reaching both thresholds is identical. This is in keeping with the national picture, where the overall percentage of pupils presently achieving the English and maths benchmark is only half a percent higher than for the five subject measure (60% against 59.5%). The measures are, in effect, very similar in their operation and outcomes. As such, the localised comparison table is presented in Annex 2 for reference.

5 A*-C EM vs APS 8

Figure 33 therefore presents a comparison between the current measure and the substantively different proposed attainment indicator - the APS 8 measure. The table thus presents the different outputs for the local schools under these alternative mechanisms of capturing desirable pupil performance.

As both are indicators of raw attainment, the schools that remain towards the top under the new measure are, unsurprisingly, those with higher levels of prior attainment. However, there are some notable movements in school outcomes. An analysis of these – drawing on the breakdown of qualifying entry patterns under the APS 8 in Figure 34 – follows on directly.
### Figure 33 – Current headline measure vs APS 8 measure

<table>
<thead>
<tr>
<th>#</th>
<th>5 A*-C EM</th>
<th>ND</th>
<th>5 A*-C EM</th>
<th>APS 8 measure</th>
<th>APS 8</th>
<th>ND*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>99%</td>
<td>97</td>
<td>Benfield Grammar School</td>
<td>Benfield Grammar School</td>
<td>426.0</td>
<td>99</td>
</tr>
<tr>
<td>2</td>
<td>80%</td>
<td>90</td>
<td>Deanbrook College</td>
<td>Deanbrook College</td>
<td>328.4</td>
<td>84</td>
</tr>
<tr>
<td>3</td>
<td>75%</td>
<td>86</td>
<td>Greenacre School</td>
<td>Portland School</td>
<td>314.5</td>
<td>75</td>
</tr>
<tr>
<td>4</td>
<td>74%</td>
<td>84</td>
<td>Huntington College</td>
<td>Queen Margaret High School</td>
<td>313.9</td>
<td>75</td>
</tr>
<tr>
<td>5</td>
<td>70%</td>
<td>78</td>
<td>Queen Margaret High School</td>
<td>Norton Academy</td>
<td>311.1</td>
<td>73</td>
</tr>
<tr>
<td>6</td>
<td>69%</td>
<td>76</td>
<td>Portland School</td>
<td>Eastwood Hill School</td>
<td>309.3</td>
<td>72</td>
</tr>
<tr>
<td>7</td>
<td>65%</td>
<td>67</td>
<td>Norton Academy</td>
<td>Greenacre School</td>
<td>303.5</td>
<td>67</td>
</tr>
<tr>
<td>8</td>
<td>65%</td>
<td>67</td>
<td>Oakwood Community College</td>
<td>Monkton Girls’ School</td>
<td>291.8</td>
<td>59</td>
</tr>
<tr>
<td>9</td>
<td>63%</td>
<td>61</td>
<td>Parkside Academy</td>
<td>Huntington College</td>
<td>287.3</td>
<td>55</td>
</tr>
<tr>
<td>10</td>
<td>60%</td>
<td>53</td>
<td>Eastwood Hill School</td>
<td>Milton School</td>
<td>281.2</td>
<td>50</td>
</tr>
<tr>
<td>11</td>
<td>56%</td>
<td>42</td>
<td>Shuttleworth School</td>
<td>Oakwood Community College</td>
<td>264.6</td>
<td>39</td>
</tr>
<tr>
<td>12</td>
<td>55%</td>
<td>39</td>
<td>RiversideTrust School</td>
<td>Longcroft Academy</td>
<td>254.2</td>
<td>31</td>
</tr>
<tr>
<td>13</td>
<td>54%</td>
<td>37</td>
<td>Monkton Girls’ School</td>
<td>Shuttleworth School</td>
<td>248.6</td>
<td>27</td>
</tr>
<tr>
<td>14</td>
<td>53%</td>
<td>34</td>
<td>Milton School</td>
<td>Parkside Academy</td>
<td>246.5</td>
<td>26</td>
</tr>
<tr>
<td>15</td>
<td>51%</td>
<td>29</td>
<td>Longcroft Academy</td>
<td>Elmhurst High School</td>
<td>204.1</td>
<td>8</td>
</tr>
<tr>
<td>16</td>
<td>31%</td>
<td>2</td>
<td>Elmhurst High School</td>
<td>RiversideTrust School</td>
<td>195.1</td>
<td>6</td>
</tr>
</tbody>
</table>

* Percentile position within the national distribution of maintained mainstream schools (1 as lowest, 100 as highest).
Figure 34 – Qualifying entries

<table>
<thead>
<tr>
<th>#</th>
<th>Rank by average qualifying entries</th>
<th>Qualifying entries</th>
<th>% all 8</th>
<th>B2 entries</th>
<th>all B2</th>
<th>B3 entries</th>
<th>all B3</th>
<th>B2 points</th>
<th>B3 points</th>
<th>Total APS8</th>
<th>Av score per entry</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Benfield Grammar School</td>
<td>8.0</td>
<td>99%</td>
<td>3.0</td>
<td>99%</td>
<td>3.0</td>
<td>100%</td>
<td>158.5</td>
<td>162.7</td>
<td>426.0</td>
<td>53.3</td>
</tr>
<tr>
<td>2</td>
<td>Norton Academy</td>
<td>7.5</td>
<td>77%</td>
<td>2.7</td>
<td>82%</td>
<td>2.8</td>
<td>88%</td>
<td>115.1</td>
<td>113.8</td>
<td>311.1</td>
<td>41.5</td>
</tr>
<tr>
<td>3</td>
<td>Portland School</td>
<td>7.4</td>
<td>73%</td>
<td>2.7</td>
<td>80%</td>
<td>2.7</td>
<td>81%</td>
<td>114.0</td>
<td>117.3</td>
<td>314.5</td>
<td>42.5</td>
</tr>
<tr>
<td>4</td>
<td>Milton School</td>
<td>7.4</td>
<td>66%</td>
<td>2.5</td>
<td>68%</td>
<td>2.9</td>
<td>93%</td>
<td>95.2</td>
<td>111.5</td>
<td>281.2</td>
<td>38.0</td>
</tr>
<tr>
<td>5</td>
<td>Queen Margaret High School</td>
<td>7.4</td>
<td>68%</td>
<td>2.4</td>
<td>68%</td>
<td>3.0</td>
<td>98%</td>
<td>105.8</td>
<td>125.8</td>
<td>313.9</td>
<td>42.4</td>
</tr>
<tr>
<td>6</td>
<td>Deanbrook College</td>
<td>7.3</td>
<td>71%</td>
<td>2.4</td>
<td>71%</td>
<td>3.0</td>
<td>98%</td>
<td>107.1</td>
<td>134.4</td>
<td>328.4</td>
<td>45.0</td>
</tr>
<tr>
<td>7</td>
<td>Greenacre School</td>
<td>7.1</td>
<td>56%</td>
<td>2.3</td>
<td>57%</td>
<td>2.9</td>
<td>90%</td>
<td>96.8</td>
<td>121.4</td>
<td>303.5</td>
<td>42.7</td>
</tr>
<tr>
<td>8</td>
<td>Eastwood Hill School</td>
<td>7.1</td>
<td>57%</td>
<td>2.5</td>
<td>66%</td>
<td>2.7</td>
<td>76%</td>
<td>112.7</td>
<td>115.3</td>
<td>309.3</td>
<td>43.6</td>
</tr>
<tr>
<td>9</td>
<td>Monkton Girls’ School</td>
<td>6.7</td>
<td>44%</td>
<td>1.8</td>
<td>44%</td>
<td>2.9</td>
<td>95%</td>
<td>81.1</td>
<td>133.3</td>
<td>291.8</td>
<td>43.6</td>
</tr>
<tr>
<td>10</td>
<td>Huntington College</td>
<td>6.7</td>
<td>55%</td>
<td>2.0</td>
<td>56%</td>
<td>2.7</td>
<td>82%</td>
<td>87.2</td>
<td>119.0</td>
<td>287.3</td>
<td>42.9</td>
</tr>
<tr>
<td>11</td>
<td>Longcroft Academy</td>
<td>6.4</td>
<td>29%</td>
<td>1.9</td>
<td>32%</td>
<td>2.5</td>
<td>66%</td>
<td>76.6</td>
<td>102.8</td>
<td>254.2</td>
<td>39.7</td>
</tr>
<tr>
<td>12</td>
<td>Parkside Academy</td>
<td>6.4</td>
<td>45%</td>
<td>1.9</td>
<td>53%</td>
<td>2.5</td>
<td>64%</td>
<td>75.0</td>
<td>95.9</td>
<td>246.5</td>
<td>38.5</td>
</tr>
<tr>
<td>13</td>
<td>Oakwood Community College</td>
<td>6.4</td>
<td>33%</td>
<td>1.5</td>
<td>33%</td>
<td>2.9</td>
<td>96%</td>
<td>58.7</td>
<td>125.2</td>
<td>264.6</td>
<td>41.3</td>
</tr>
<tr>
<td>14</td>
<td>Shuttleworth School</td>
<td>6.2</td>
<td>34%</td>
<td>1.6</td>
<td>36%</td>
<td>2.6</td>
<td>77%</td>
<td>63.5</td>
<td>111.6</td>
<td>248.6</td>
<td>40.1</td>
</tr>
<tr>
<td>15</td>
<td>Elmhurst High School</td>
<td>5.6</td>
<td>18%</td>
<td>1.1</td>
<td>19%</td>
<td>2.5</td>
<td>70%</td>
<td>41.6</td>
<td>98.7</td>
<td>204.1</td>
<td>36.4</td>
</tr>
<tr>
<td>16</td>
<td>Riverside Trust School</td>
<td>5.6</td>
<td>8%</td>
<td>1.6</td>
<td>20%</td>
<td>2.1</td>
<td>45%</td>
<td>53.2</td>
<td>69.9</td>
<td>195.1</td>
<td>34.8</td>
</tr>
</tbody>
</table>
Huntington College against Queen Margaret High School

A good example of the potential impact of the APS 8 measure is evident in the contrasting performance of Huntington College and Queen Margaret High School. Under the existing localised league table rankings, these schools occupy fourth and fifth position respectively. Both are academy convertors and have very similar (relatively middle class) intakes, with almost identical levels of prior attainment and FSM pupils.

Both schools score well under the established measure – at Huntington, 74% of pupils reach the threshold, at Queen Margaret it is 70%. However, under the proposed APS 8 measure, the relative ranking of Huntington College – an Outstanding school - suffers significantly, falling from 4th to 9th position locally and dropping down 30 percentile positions within the national distribution (to 55th). Queen Margaret, by contrast, holds up well under the new measure, jumping over Huntington into 4th position locally and staying broadly static nationally (75th percentile).

Figure 34 identifies the reason for this significant swing. In total, Huntington averages almost a full qualifying entry less per pupil under the APS 8 measure than at Queen Margaret (6.7 against 7.4). Indeed, in terms of average performance per entry, Huntington actually slightly outscores Queen Margaret (42.9 against 42.4). As such, the difference in the fortunes of the two schools is – in this instance - driven entirely by differences in entry patterns.

As the local authority analyst put it, Huntington has – as a specialist technology college – sought to meet the needs of its pupils by “going for alternative qualifications for pupils right across the board”. Thus, despite its relatively strong intake, only 55% of pupils at Huntington would fill all eight slots under the new measure. This is a result of a gap both in the 3rd basket (due to certain qualifications not being approved) and, particularly, in the EBacc basket (where, with an average of 2 qualifying entries, it loses almost half an entry per pupil against Queen Margaret).

Oakwood Community College against Norton Academy

A similar pattern is evident in a comparison of Oakwood Community College and Norton Academy. The two schools have matching scores under the current measure (at 65%, putting them in the
67th percentile nationally) and also have relatively similar intakes. However, they diverge sharply under the new measure, with a 34 percentile position difference emerging.

This is again primarily a matter of entry patterns. The average performance per entry in each school is almost identical (41.3 and 41.5). However, Norton Academy averages over a full qualifying entry per pupil more than Oakwood Community College. The overall points discrepancy is therefore again being driven almost entirely by curriculum choices and, in this instance, this relates exclusively to entries in the EBacc basket.

Both schools come close to filling the third basket slots (with Oakwood Community College actually slightly higher at 2.9 average entries compared to 2.8). However, Norton Academy has explicitly retained a “traditional curriculum”117, coming second only to the selective grammar Benfield in terms of EBacc basket entries (2.7). By contrast, Oakwood Community College has the second lowest average entry rate in the EBacc basket (at only 1.5).

These two comparisons therefore emphasise the curriculum related aspect of the new APS 8 measure. However, the Norton Academy example also begins to highlight the dual nature of the drivers under the measure. Although second highest in terms of overall entries, it remains in 5th place locally on the APS 8 table. This is almost 20 points and 11 national percentile positions behind Deanbrook College (a school with lower average qualifying entries) – a gap driven exclusively by performance per entry.

**Monkton Girls’ School against Parkside Academy**

The specific impact of performance is well illustrated in relation to Monkton Girls’. A school with a relatively challenging intake, it falls 24 percentile positions behind Parkside Academy (a school with a similar intake) under the current measure. Under the proposed APS 8 indicator, however, Monkton Girls’ comes out 34 percentile positions ahead, jumping up in the process from 13th to 8th position locally.

As Figure 34 highlights, this dramatic swing between the schools (58 percentile positions in total) is not driven heavily by entry patterns. Both schools have only around 45% of pupils filling all eight slots and are similarly positioned on average EBacc entries (at around a relatively low 1.8). As such, three quarters of the differential

117 Local authority analyst
Measuring what matters

between their outcomes under the new measure is driven *purely by performance*.

Notably, this is not a difference in performance across the board. In the first two baskets, the schools’ scores are relatively similar – 158 compared to 151. As such, 83% of the total points difference between the schools is isolated to scoring in the third basket. As a specialist arts college, Monkton Girls’ has large numbers of motivated pupils performing well in high value vocational and non-vocational arts qualifications (Art and Design, D&T, Film Studies, Performing Arts and Music Studies) – all of which contribute to the third basket score. Indeed, looking at the breakdown, the schools’ average score in this basket is markedly superior to almost any other school in the area (excluding the selective grammar). This in turn ensures the school’s overall average score per entry is the third highest in the local authority.

Thus, given this strength in depth, Monkton Girls’ benefits significantly from the move towards an attainment measure that rewards performance across a broader curriculum. This is in direct contrast to Parkside, which was identified as perhaps the one school in the area most adept at “playing the game” with regard to performance tables under the current measure. Its score of 95.9 points in the third basket is the second worst in the whole area.

As such, over 60 per cent of the sizeable points difference between Monkton Girls’ and Parkside is being driven exclusively by superior scoring by the former in non-EBacc subjects. This would suggest that the proposed measure is able to reward (and continue to encourage) excellence in non-core subjects (an important concern explicitly recognised in the consultation).

---

118 Local authority analyst
12 Test-case local authority - progress measure breakdown

APS 8 vs Progress 8

It is important to understand the likely operation of this APS 8 measure not only (or indeed even primarily) as an indicator in and of itself but as the foundation on which the proposed progress measure will be based. As highlighted above, such a measure is designed to make comparisons of school performance fair (and meaningful) by isolating the value added by the institution.

In effect, the measure provides a snapshot of the distance travelled by pupils from a baseline position (i.e. KS2) against a chosen marker of KS4 performance (i.e. APS 8). If the new Progress 8 indicator is to be central, therefore, then so too will be the mechanics (and resulting incentives) of this facilitating attainment measure.

Having established the operation of this underlying measure in relation to the schools in the test-case local authority, it is therefore necessary to take the next step and assess the operation and impact of the proposed resulting Progress 8 measure. This is illustrated in Figure 35.

Modelling of Progress 8

The Progress 8 measure has been modelled using the same value added methodology presently used to calculate existing progress measures.\(^\text{119}\)

As proposed in the consultation, it uses English and maths scores at KS2 as the baseline and APS 8 scores as the KS4 output. The resulting measure therefore benchmarks average school performance at a score of 1000 – that is, a school scoring precisely

1000 will be securing scores for its pupils on the APS 8 measure that directly equate to the national average performance for pupils of that combined demographic of prior attainment. The relative performance of individual institutions is then compared against this benchmark. ‘Good’ schools adding more value than this average will have scores proportionally in excess of 1000. Schools adding less value will have scores that are proportionately below this figure.\(^\text{120}\)

Notably, the overall average school score in the local authority is very close to the national benchmark (at 1002).

---

\(^{120}\) See Annex 1 for details of the progress measure calculation.
**Figure 35 – APS 8 vs Progress 8**

<table>
<thead>
<tr>
<th>#</th>
<th>APS 8</th>
<th>ND</th>
<th>APS 8 measure</th>
<th>Progress 8 (P8) measure</th>
<th>P8</th>
<th>ND</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>426.0</td>
<td>99</td>
<td>Benfield Grammar School</td>
<td>Deanbrook College</td>
<td>1028.7</td>
<td>86</td>
</tr>
<tr>
<td>2</td>
<td>328.4</td>
<td>84</td>
<td>Deanbrook College</td>
<td>Monkton Girls’ School</td>
<td>1027.9</td>
<td>85</td>
</tr>
<tr>
<td>3</td>
<td>314.5</td>
<td>75</td>
<td>Portland School</td>
<td>Milton School</td>
<td>1027.6</td>
<td>85</td>
</tr>
<tr>
<td>4</td>
<td>313.9</td>
<td>75</td>
<td>Queen Margaret High School</td>
<td>Benfield Grammar School</td>
<td>1025.7</td>
<td>83</td>
</tr>
<tr>
<td>5</td>
<td>311.1</td>
<td>73</td>
<td>Norton Academy</td>
<td>Queen Margaret High School</td>
<td>1021.3</td>
<td>77</td>
</tr>
<tr>
<td>6</td>
<td>309.3</td>
<td>72</td>
<td>Eastwood Hill School</td>
<td>Longcroft Academy</td>
<td>1013.2</td>
<td>66</td>
</tr>
<tr>
<td>7</td>
<td>303.5</td>
<td>67</td>
<td>Greenacre School</td>
<td>Norton Academy</td>
<td>1010.2</td>
<td>62</td>
</tr>
<tr>
<td>8</td>
<td>291.8</td>
<td>59</td>
<td>Monkton Girls’ School</td>
<td>Elmhurst High School</td>
<td>1004.2</td>
<td>53</td>
</tr>
<tr>
<td>9</td>
<td>287.3</td>
<td>55</td>
<td>Huntington College</td>
<td>Shuttleworth School</td>
<td>1003.7</td>
<td>52</td>
</tr>
<tr>
<td>10</td>
<td>281.2</td>
<td>50</td>
<td>Milton School</td>
<td>Eastwood Hill School</td>
<td>1003.4</td>
<td>52</td>
</tr>
<tr>
<td>11</td>
<td>264.6</td>
<td>39</td>
<td>Oakwood Community College</td>
<td>Portland School</td>
<td>999.7</td>
<td>46</td>
</tr>
<tr>
<td>12</td>
<td>254.2</td>
<td>31</td>
<td>Longcroft Academy</td>
<td>Greenacre School</td>
<td>998.0</td>
<td>43</td>
</tr>
<tr>
<td>13</td>
<td>248.6</td>
<td>27</td>
<td>Shuttleworth School</td>
<td>Huntington College</td>
<td>991.4</td>
<td>35</td>
</tr>
<tr>
<td>14</td>
<td>246.5</td>
<td>26</td>
<td>Parkside Academy</td>
<td>Parkside Academy</td>
<td>975.8</td>
<td>22</td>
</tr>
<tr>
<td>15</td>
<td>204.1</td>
<td>8</td>
<td>Elmhurst High School</td>
<td>Oakwood Community College</td>
<td>966.6</td>
<td>14</td>
</tr>
<tr>
<td>16</td>
<td>195.1</td>
<td>6</td>
<td>Riverside Trust School</td>
<td>Riverside Trust School</td>
<td>960.6</td>
<td>11</td>
</tr>
</tbody>
</table>
Portland

The performance of this school in isolation gives a clear insight into both the rationale behind the Progress 8 measure and its operation. Portland is the school with the second highest level of prior attainment in the local authority (after the selective grammar). Indeed, half of its intake are designated as ‘high attainers’ on entry and only 5% as ‘low attainers’.

However, of these low attainers, not one achieved the A*-C threshold in English and maths (compared to almost 10% for these pupils across the local authority). In turn, only 50% of middle attainers reached this mark compared to a local authority average of over 60% (and respective performance in Greenacre School and Deanbrook College of 78% and 82%). As such, Portland was described as a “school coasting on its intake” with “chronic underachievement across the board”.

Thus, while this high level of relative pupil prior attainment (combined with retention of a firmly traditional curriculum) ensures a seemingly strong score on the APS 8 measure (75th percentile nationally), the actual school performance on the value added indicator falls below the national average benchmark (to the 46th percentile). By contrast, Queen Margaret – which secures the same APS 8 score with more than twice the proportion of low attainers and almost 50% fewer high attainers – sees its position hold up under the Progress 8 measure.

However, notably, the Progress 8 score for Portland does not fall as significantly as it would do if the value added calculation was based on an output attainment indicator without any curriculum related specifications – against a straight ‘best 8’ measure, the school’s score falls to well below average at 973. Part of the notable underperformance of Portland was attributed by analysts within the local authority to a somewhat ‘stuffy’ or overly traditional approach and a resulting unwillingness to tailor a curriculum to suit and enthuse individual pupils (in direct contrast, for example, to Monkton Girls’). However, this same approach ensures that – when calculated using an underlying attainment indicator that rewards both particular entry patterns and raw performance – the value added outcome is protected more than might otherwise be the case.

121 Local authority analyst
122 Annex 3 provides a full comparison of the Progress 8 scores for each school against their existing value added performance under the present progress indicator.
Huntington College, Eastwood Hill, Greenacre and Norton Academy against Queen Margaret

A clear example of how the Progress 8 indicator operates as a measure of school performance is evident from a comparison of the varying fortunes of a series of schools with relatively similar, relatively middle class intakes. These five schools - Queen Margaret, Norton Academy, Eastwood Hill, Greenacre and Huntington College – occupy roughly the upper middle positions within the localised APS 8 rankings.

Of the five, Queen Margaret achieves the highest APS 8 score (314) with, in relative terms, slightly the weakest intake (as outlined in Figure 36). As such, it also scores well (77\textsuperscript{th} percentile) under the Progress 8 measure: while it has a relatively advantaged intake, it is performing well for this intake in relation to the relevant national average. Each of the other schools, however, by achieving respectively lower scores with similar (or slightly more advanced) intakes perform correspondingly worse under the value added measure. They are, to varying degrees, penalised for being “leafy lane coasters”.

Figure 36 – School intake demographics

<table>
<thead>
<tr>
<th>School</th>
<th>% of low attainers on entry</th>
<th>% of high attainers on entry</th>
<th>Average KS2 score on entry</th>
<th>% FSM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eastwood Hill</td>
<td>6%</td>
<td>41%</td>
<td>29.1</td>
<td>6.7%</td>
</tr>
<tr>
<td>Greenacre</td>
<td>11%</td>
<td>38%</td>
<td>28.8</td>
<td>11.9%</td>
</tr>
<tr>
<td>Norton Academy</td>
<td>9%</td>
<td>44%</td>
<td>28.7</td>
<td>12.7%</td>
</tr>
<tr>
<td>Huntington</td>
<td>11%</td>
<td>37%</td>
<td>28.5</td>
<td>11.3%</td>
</tr>
<tr>
<td>Queen Margaret</td>
<td>11%</td>
<td>35%</td>
<td>28.3</td>
<td>10.4%</td>
</tr>
</tbody>
</table>

For Norton Academy, the drop is not dramatic (around 10 percentiles nationally). However, as we have seen, this school is a significant ‘winner’ under the proposed APS 8 measure due to its existing (strong EBacc) entry patterns. It is thus deemed to add more value in relation to this structured measure than it would under a value

123 Local authority analyst
added indicator calculated against a straight ‘best 8’ (where it is nearer to the 1000 average, as evident from Annex 3).

Huntington College, by contrast, suffers significantly under the APS 8 measure as detailed above, with almost a full qualifying entry less per pupil than Norton Academy. Its resulting relatively low score with a strong intake ensures that the corresponding Progress 8 outcome is noticeably poor.\textsuperscript{124} It would thus occupy the 35\textsuperscript{th} percentile of a national league table based on the proposed value added measure, compared to its original position in the 84\textsuperscript{th} percentile under the current threshold measure.

This comparison again raises an important point. Just as differences in scores between schools under the proposed APS 8 measure are driven by a combination of entry patterns and performance per entry, so too therefore are differences in the Progress 8 scores. That is, Huntington College would be in a position to enhance its value added score under the proposed measure by both improving the grades of its pupils and altering its curriculum offer. At nearer to full coverage of the 8 slots, however, Norton Academy would be more heavily reliant on only the former.

This illustrates the clear implications as to curriculum design (and resulting resource allocation) in schools that will initially underperform against the measure. That is, schools making early marked improvements in their value added scores are likely to provide evidence of shifting patterns of qualifying entries as well as enhanced raw outcomes.

\textbf{Oakwood Community College}

One further school that could have been added to the table of similar intakes above was Oakwood Community College. Indeed, at an average KS2 score on entry of 28.6, it has a marginally stronger intake than both Queen Margaret and Huntington College. However, as identified above, its performance under the APS 8 measure is strikingly poor (due largely to very low performance in the EBacc basket). As such, it falls from the 67\textsuperscript{th} percentile under the established measure to the 39\textsuperscript{th} percentile under APS 8. Given such a strong relative intake, therefore, this in turn gives rise to a Progress 8 score falling into the 14\textsuperscript{th} percentile nationally.

This school therefore presents an interesting example of how the

\textsuperscript{124} Under the value added measure based on a straight ‘best 8’, Huntington therefore scores better (at 1003, slightly above the average benchmark).
proposed dual floor target may have an impact. At 65% on the A*-C EM threshold measure, this school would be in no danger in relation to this first aspect. However, by dropping towards the bottom decile of schools under the Progress 8 measure, this would be likely to be a school that would warrant concern and potential intervention. Notably, therefore, it has in fact been recently placed in special measures following an inspection by Ofsted.

**Monkton Girls’**

As we saw above, Monkton Girls’ benefitted significantly from the move towards a broader attainment measure based primarily on the strength of its performance per entry. It thus moved up the localised tables under this new measure, occupying a position between Greenacre School and Huntington College.

As evident from Figure 36 above, these schools both have relatively affluent intakes with high levels of prior attainment. Monkton Girls’, by contrast, has a markedly more challenging intake, with double the proportion of FSM pupils (at 23.5%) and around 2 full points lower prior attainment at KS2 (26.6). As such, this relative over-performance in relation to its intake is rewarded. Indeed, the school rises from the 37th percentile nationally based on the current (narrower) threshold measure to the 85th percentile under the Progress 8 indicator.

**Shuttleworth School and Longcroft**

These schools are both in high deprivation areas with over 30% low attainers and FSM pupils in each. However, they were identified qualitatively as two of the best schools in the local authority – Shuttleworth is rated ‘Good’ by Ofsted and Longcroft is an established ‘Outstanding’ school. Thus, by securing APS 8 scores within 20-25 points of somewhere like Huntington College, these schools are performing “exceptionally well for the pupils they have”.

Indeed, both have now been enlisted to provide partnership and support for struggling schools in the area.

As such, Shuttleworth moves from the 27th percentile under APS 8 to the 52nd percentile under Progress 8. The move for Longcroft is from the 31st percentile to the 66th. In effect, both schools cross the Rubicon from below to above average.

---

125 This will of course depend at what level the initial benchmarks are set.
126 Local authority analyst
The real strength of the Progress 8 measure is therefore evident from the scope for genuine comparison it allows between these schools and one like Portland. With seven times the level of low attainers in the cohort, a comparison of straight outcomes for the pupils at Longcroft with those at Portland tells us very little about the performance of Longcroft as a functioning institution.
13 Conclusion

In keeping with the subsidiary objective, the curriculum related analysis of the proposed measures indicates that the model of the ‘good school’ they point towards is one that not only secures the best possible outcomes for all pupils, it also secures them across a broad and at least partially structured range of subjects.

Both the APS 8 measure and the progress measure it facilitates will therefore be ‘performance’ measures in a broad sense in that they will be sensitive to both the subjects pupils take and grades they achieve. Schools with higher average numbers of qualifying entries per pupil will generally perform better. As such, schools making early marked improvements are likely to provide evidence of shifting curriculum decisions as well as enhanced raw pupil outcomes.

In particular, the intended incentive for schools to encourage uptake of ‘an academic core of subjects’ is likely to be a reasonably strong one. This will be, to an extent, an extension of a journey many schools have already begun. However, testing also reveals the clear potential for schools to achieve good scores under the measures based primarily on strong pupil performance in high value vocational and creative subjects.

The ‘correct’ balance with regard to the curriculum that schools should be encouraged to offer their pupils is perhaps inevitably a values based judgement subject to political debate. However, it is only by presenting the operation of the proposed measures - and the strength of the curriculum incentives they will drive – in practice that a full and meaningful debate can be had.

Having done so, the government should consult on the appropriate balance in the context of the decision about the points-per-grade scale to be used under the new APS 8 / Progress 8 measures.
Annex 1 – Data and modelling

Data

For the purposes of this analysis, we use data from the National Pupil Database. The National Pupil Database (NPD) is a dataset containing key stage attainment records for the population of all pupils in England. This is combined with census information for pupils in state schools, which contains information on pupil characteristics such as ethnicity and gender. The NPD is an administrative dataset administered by the Department for Education.

Our dataset, and the results in this paper focus exclusively on pupils in state schools (which comprise maintained mainstream schools, academy converters, and free school converters) who were studying for Key Stage 4 (GCSE or equivalent) level in 2012. Our population of interest comprises 556,382 pupils and 3,065 schools.

The key measure used in this analysis is the ‘APS 8’ measure of performance as outlined by the Department for Education in their Secondary School Accountability Consultation.127

Modelling

‘APS 8’ measure of performance

The APS 8 measure has been modelled according to the proposals set out in the consultation document.

The following specific modelling decisions have been followed:

- ‘Core’ and ‘additional’ science (i.e. the components of combined science) are treated as two separate qualifications filling two slots in the measure.
- English Literature GCSE is counted in the third basket (as suggested in the consultation document).
- Only high value vocational qualifications on the Department’s approved list for 2014 performance tables have been taken to qualify for the third basket.
- Approved vocational qualifications presently equivalent to multiple GCSEs have been reduced to a 1:1 equivalence (as will be the case from 2014).

Points system

We use the standard GCSE points scheme as follows:

- A* = 58
- A  = 52
- B  = 46
- C  = 40
- D  = 34
- E  = 28
- F  = 22
- G  = 16

For GCSE equivalents, standard DfE scoring systems are used.

Calculation of the ‘value added’ measure of performance

The value added measure is calculated using the standard methodology, as outlined in Department for Education (2012), ‘A guide to value added key stage 2 to 4 in 2011 school & college performance tables & raise online’, Department for Education, London


Discounting

Standard discounting rules apply in the calculation of all measures above. Discounts (where a pupil has two or more entries for the same qualification, and thus only one qualification should be counted) apply in the following cases:

- Claimed Internally – Pupil is shown twice at same school and this pupil is incorrect
- Internal Claim – Pupil is shown twice at same school, but this pupil is the correct one
- Transferred In - Pupil has been transferred in from another school
- Transferred Out - Pupil has been claimed by another school
- Pupil has been counted twice (not currently transferred out or sixth form centre)
- Pupil not at end of Key Stage (forced out)
- Results amended
- Grade Amended
- New entry
- Transferred In - Results have been claimed from another student at same school.
- Internal Claim – Pupil is shown twice at same school, claim this result from another student.
- Claimed Internally – Transfer of result to another student.
- Withdrawn
## Annex 2 – Current headline measure vs proposed English and maths threshold measure

<table>
<thead>
<tr>
<th>#</th>
<th>5 A*-C EM</th>
<th>5 A*-C EM</th>
<th>A*-C EM</th>
<th>A*-C EM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>99%</td>
<td>Benfield Grammar School</td>
<td>Benfield Grammar School</td>
<td>99%</td>
</tr>
<tr>
<td>2</td>
<td>80%</td>
<td>Deanbrook College</td>
<td>Deanbrook College</td>
<td>81%</td>
</tr>
<tr>
<td>3</td>
<td>75%</td>
<td>Greenacre School</td>
<td>Greenacre School</td>
<td>79%</td>
</tr>
<tr>
<td>4</td>
<td>74%</td>
<td>Huntington College</td>
<td>Huntington College</td>
<td>74%</td>
</tr>
<tr>
<td>5</td>
<td>70%</td>
<td>Queen Margaret High School</td>
<td>Queen Margaret High School</td>
<td>70%</td>
</tr>
<tr>
<td>6</td>
<td>69%</td>
<td>Portland School</td>
<td>Portland School</td>
<td>70%</td>
</tr>
<tr>
<td>7</td>
<td>65%</td>
<td>Norton Academy</td>
<td>Norton Academy</td>
<td>65%</td>
</tr>
<tr>
<td>8</td>
<td>65%</td>
<td>Oakwood Community College</td>
<td>Oakwood Community College</td>
<td>65%</td>
</tr>
<tr>
<td>9</td>
<td>63%</td>
<td>Parkside Academy</td>
<td>Parkside Academy</td>
<td>63%</td>
</tr>
<tr>
<td>10</td>
<td>60%</td>
<td>Eastwood Hill School</td>
<td>Eastwood Hill School</td>
<td>60%</td>
</tr>
<tr>
<td>11</td>
<td>56%</td>
<td>Shuttleworth School</td>
<td>Shuttleworth School</td>
<td>56%</td>
</tr>
<tr>
<td>12</td>
<td>55%</td>
<td>Riverside Trust School</td>
<td>Riverside Trust School</td>
<td>55%</td>
</tr>
<tr>
<td>13</td>
<td>54%</td>
<td>Monkton Girls’ School</td>
<td>Monkton Girls’ School</td>
<td>54%</td>
</tr>
<tr>
<td>14</td>
<td>53%</td>
<td>Milton School</td>
<td>Milton School</td>
<td>53%</td>
</tr>
<tr>
<td>15</td>
<td>51%</td>
<td>Longcroft Academy</td>
<td>Longcroft Academy</td>
<td>51%</td>
</tr>
<tr>
<td>16</td>
<td>31%</td>
<td>Elmhurst High School</td>
<td>Elmhurst High School</td>
<td>31%</td>
</tr>
<tr>
<td>ND</td>
<td>VA</td>
<td>Progress 8</td>
<td>Current value added measure</td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>----</td>
<td>-----------</td>
<td>-----------------------------</td>
<td></td>
</tr>
<tr>
<td>86</td>
<td>85</td>
<td>1028.7</td>
<td>Monkton Girls' School</td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>83</td>
<td>1027.6</td>
<td>Milton School</td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>77</td>
<td>1025.7</td>
<td>Benfield Grammar School</td>
<td></td>
</tr>
<tr>
<td>83</td>
<td>76</td>
<td>1021.3</td>
<td>Queen Margaret High School</td>
<td></td>
</tr>
<tr>
<td>82</td>
<td>66</td>
<td>1013.2</td>
<td>Longcroft Academy</td>
<td></td>
</tr>
<tr>
<td>82</td>
<td>62</td>
<td>1010.2</td>
<td>Norton Academy</td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>53</td>
<td>1004.2</td>
<td>Elmhurst High School</td>
<td></td>
</tr>
<tr>
<td>78</td>
<td>52</td>
<td>1003.7</td>
<td>Shuttleworth School</td>
<td></td>
</tr>
<tr>
<td>77</td>
<td>46</td>
<td>1003.4</td>
<td>Portland School</td>
<td></td>
</tr>
<tr>
<td>76</td>
<td>43</td>
<td>998.7</td>
<td>Greenacre School</td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>35</td>
<td>998.1</td>
<td>Oakwood Community College</td>
<td></td>
</tr>
<tr>
<td>73</td>
<td>22</td>
<td>985.8</td>
<td>Riverside Trust School</td>
<td></td>
</tr>
<tr>
<td>72</td>
<td>14</td>
<td>985.6</td>
<td>Oakwood Community College</td>
<td></td>
</tr>
<tr>
<td>71</td>
<td>11</td>
<td>973.0</td>
<td>Portland School</td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>11</td>
<td>960.6</td>
<td>Elmhurst High School</td>
<td></td>
</tr>
<tr>
<td>69</td>
<td>8</td>
<td>918.1</td>
<td>Oakwood Community College</td>
<td></td>
</tr>
<tr>
<td>68</td>
<td>8</td>
<td>958.1</td>
<td>Riverside Trust School</td>
<td></td>
</tr>
<tr>
<td>67</td>
<td>8</td>
<td>956.6</td>
<td>Oakwood Community College</td>
<td></td>
</tr>
<tr>
<td>66</td>
<td>7</td>
<td>950.6</td>
<td>Portland School</td>
<td></td>
</tr>
</tbody>
</table>