

Response to the Department for Education’s “Secondary School Accountability Consultation”

***Sarah Cattan, Claire Crawford, Jonathan Cribb, Lorraine Dearden,
Ellen Greaves, Iftikar Hussain, Michelle Jin, Imran Rasul, Luke
Sibieta, Anna Vignoles***

Institute for Fiscal Studies

May 2013

Key points

- Our response to this consultation is based on the principle that accountability can help improve standards in schools, and promote the attainment and progression of all pupils (Muriel and Smith, 2011¹). The objective is to structure the system of accountability to maximise the attainment and progression of all pupils while minimising the opportunity for schools to “game” the system.
- Key headline measures should be easily accessible, comparable across schools, and understood by parents; the Data Portal should highlight the key headline measures from other information provided.
- Threshold and progress standards should be clearly defined and objectively measured, and not subject to arbitrary increases.
- Future school accountability measures should include “destinations” information, for example the proportion of pupils continuing to further and higher education or employment. While it is less straightforward to produce the proportion going into employment, we see this as an important future priority, as it would help incentivise schools to invest in their pupils in ways beyond the formal curriculum.
- Absolute standards over time should be monitored through additional tests. These should be introduced for a sample of pupils, on the condition that:
 - The additional tests cover material taught in core GCSE subjects (including English and maths);

¹ Muriel, A. and Smith, J. (2011), On Educational Performance Measures. Fiscal Studies, 32: 187–206.

- The additional tests can't be gamed (which may mean that the exam format is randomly changed across schools or across years);
- Schools and teachers are made fully aware that the scores of their pupils will not be publicly reported;
- An individual pupil will only take one additional test, to minimise the burden on pupils around the time of high stakes exams;
- Further consideration should be given to the possibility of other changes to the exam system, e.g. removing exam board choice (to avoid exam boards competing in a “race to the bottom”) and/or strengthening the role of a truly independent body whose aim is to monitor overall standards.
- Ofsted inspections should be produced entirely independently from published headline measures.

Supplementary points

- The reformed GCSEs should be tested against the old GCSEs to provide a comparison of standards across time, by a sample of pupils taking both reformed and old GCSEs as part of their development.
- A relative floor standard in the first year of reformed GCSEs should not be produced, as it is redundant (it gives the same information as league tables) and arbitrary.
- Results from internal assessments voluntarily chosen by schools to inform their pupil's progress should not be published on the Data Warehouse, as it may distort schools' incentives and the results are subject to manipulation.
- Supplementary assessments should be made available so that a larger proportion of pupils can access assessment (or demonstrate high ability) to enable a wide range of schools to be appropriately held account for their pupils' attainment.
- The value of KS3 teacher assessments and the cost implications to schools should be considered before the decision to stop collecting KS3 data.
- Clarification of the process behind classification of the “high value” subjects that can be included as part of these measures should be produced.

Introduction

The Secondary School Accountability Consultation asks how the Department for Education should publish and use data about schools' performance to "promote pupils' deep understanding across a broad curriculum and maximise progress and attainment for all pupils".

The consultation recognises that the previous system of accountability created some perverse incentives, such as an excessive focus on improving the attainment of pupils on the borderline between a grade D and C at GCSE to the detriment of other pupils, as this is the threshold which contributed to the headline accountability measure. An aim of the consultation is therefore to reduce the perverse incentives for schools to focus on a narrow group of pupils.

Our response to this consultation is based on the principle that accountability can help improve standards in schools, and promote the attainment and progression of all pupils (Muriel and Smith, 2011²; Burgess, Wilson and Worth, forthcoming³). The objective is to structure the system of accountability to maximise the attainment and progression of all pupils while minimising the opportunity for schools to "game" the system.

The objective for all schools should be to maximise the attainment and progression for all its pupils, irrespective of pupils' ability, characteristics and most suitable future career. If each pupils' academic potential and ideal post-compulsory schooling destination were known, then it would be relatively easy to hold schools to account: a school where a large proportion of pupils achieved their potential and had the best preparation for their future life would be ranked above a school where a low proportion of pupils achieved this (in a way easily seen by parents and other interested parties). Schools would be held to account by this published information through a number of mechanisms: the concern for number of pupils attending the school, the views of the local community, career concerns for the head-teacher, and the recruitment and retention of teachers. It is worth noting that when pupil numbers in the local area are high, the accountability mechanism is weakened, as even poorly performing schools must be allocated the maximum number of pupils.

² Muriel, A. and Smith, J. (2011), On Educational Performance Measures. *Fiscal Studies*, 32: 187–206.

³ Burgess, S., Wilson, D., and Worth, J. (forthcoming) A natural experiment in school accountability: the impact of school performance information on pupil progress and sorting. *The Journal of Public Economics*.

This ideal measure would reduce the presence of perverse incentives for schools, as all pupils would be encouraged to achieve their potential. (It may still be the case that some pupils are ignored if thought too unlikely to reach their potential.) Of course, this ideal measure does not exist, as each pupil's potential and most suitable direction in life is hard to measure and predict. In the absence of this measure, it is very difficult to promote the attainment and progression of all pupils without introducing perverse incentives, for example to change the range of subjects offered, but the aim to do so is worthwhile.

Our response to this consultation aims to maximise the positive impact of school accountability measures, discussing our points of agreement with the consultation, our concerns, and suggestions for improvement with regard to the proposed changes to the accountability system.

We discuss each consultation question in turn.

Question 1: do you agree with the proposals for the headline accountability measures?

The headline measures are:

1. A threshold measure in English and maths (the percentage of pupils achieving a "good standard").
2. An average points score measure based on best 8 qualifications (English, maths, any combination of 3 other Ebacc subjects, and 3 other high value subjects) for all pupils.
3. A progress measure (value added methodology) based on best 8 subjects (defined as above) for all pupils between KS2 and KS4.

General comments

We agree that performance tables should continue to show headline measures, defined as the most important information about schools: research in progress shows that clear, well-presented information about schools has an impact on parents' school choices, especially for parents from more disadvantaged circumstances (Hussain, 2013). It is therefore **vital that the most important information about schools is easily accessible and comparable across schools for parents**, and that the amount of information on the Data Portal does not obscure the key details.

It is not clear what the most important information about schools is, however. Accountability will increase standards through the demand side only if parents care about what is published, but little is known about what parents understand

and care most about (for example threshold or progress measures). The demand side is also limited by capacity constraints, as low performing schools in areas with limited capacity have no incentive to attract pupils by increasing standards (as their seats will be filled anyway). Below, we assume that parents value the headline measures above, although it must be noted that if parents value the threshold measure only (and schools know this) then the perverse incentives will remain. Counter to this, if parents value Ofsted ratings, which are informed by all threshold measures (or aspects of school quality correlated with the threshold measures), then the schools will have an incentive to improve standards.

Specific comments

- We agree that only **high value qualifications should be used in performance tables**, subject to the definition of “high value”⁴.
- We **support the inclusion of the average points score** as a headline measure as a good attempt to encourage a broad curriculum. The **average number of high-value subjects taken** should be published as an additional incentive for schools to offer a broad curriculum.
- The value-added progress measure should be presented with **correct statistical inference**, to allow parents to be certain about the magnitude and significance of differences in average test scores. We agree that the proposed value-added measure will encourage the progression of a wide range of pupils.
- The difficulty for schools with intakes with very high previous attainment to maintain high levels of attainment and progression should be addressed; reform of GCSEs and earlier assessments to allow **further differentiation for those with high ability**, for example, or **supplementary assessments** for those of high ability.
- The consultation proposes that schools should have to meet a set standard on both the threshold and progress measure to be above the floor standard. We suggest that **this set standard should be clearly defined and objectively measured**, and not subject to arbitrary increases.

⁴ A “high value” qualification is not defined in this consultation, although the current qualifications included in performance tables is published here: <http://media.education.gov.uk/assets/files/pdf/q/2015%20ks4%20wolf%20list.pdf>. It is important to note that “value” defined by labour market returns will depend on the proportion of pupils that take the qualification, and so will vary over time. The principles of defining a “high value” qualification based on the opinion of the Department for Education should be clarified.

- This section of the consultation notes that GCSEs should be reformed by Ofqual with a focus on the value to pupils rather than the absolute reliability of assessments, which we presume means the absolute standards required by reformed GCSEs will be higher. If this is the case, then the **absolute standards of the new and reformed GCSE exams should be compared**, using a sample of students that sit both exams. This would allow comparison of standards over time and inform the performance of schools between years. To reduce the burden on the selected pupils and schools, different test elements of the reformed and old GCSEs could be tested independently.

Question 2: Is there any further information we should provide about the performance of disadvantaged pupils?

The consultation states that information on the attainment (and progress) of those eligible for the pupil premium, other children, and the gap between them will be published.

This information may increase schools' incentives to raise the standards of disadvantaged pupils if parents value this. However, if prominence is given to the gaps in attainment within schools, schools may also have an incentive to reduce the attainment of pupils not classified as disadvantaged to create an artificially small difference. Furthermore, pupils not eligible for free school meals or the pupil premium can come from very different backgrounds, which may drive the level of any attainment gaps across schools.

We therefore recommend that **the attainment and progress made by disadvantaged pupils should be compared with other schools with similar intakes**. No prominence should be given to the attainment gap within individual schools. This would reduce the disincentive to increase the attainment of pupils not classified as disadvantaged.

Question 3: Should we look to use a relative measure as the floor standard in the first year of the new exams

After the introduction of reformed GCSEs, the consultation questions whether the floor standard should be set as a relative measure in the first year, rather than a pre-determined standard.

This **relative floor measure is redundant**, as the lowest performing schools according to this new measure will be identified in the league tables of headline measures.

More generally, **floor standards should be pre-determined by an absolute standard of attainment that the Department expects.**

The **reformed GCSEs should be compared with the current GCSEs** in order to determine the relationship between them, which could in turn inform floor standards in future.

Question 4: Are there any other measures we should consider publishing?

We agree that a broad range of measures should be available to parents on the Data Portal, although we have some concerns about information overload. To prevent this, it is vital that **key headline measures should be more prominent and clearly displayed**, consistent with the beneficial impact of clear information on school choice.

Our suggestions for additional measures that should be considered are as follows (in order of priority):

- **Progress measures for those with different levels of prior attainment.**
- **The proportion of pupils continuing into further and higher education or employment:** although less timely, this is valuable information on which we believe schools should be judged. This information would require routine linkage of education and employment administrative datasets. The former is more straightforward than the latter, but we see this as an important future priority, as it would help incentivise schools to invest in their pupils in ways beyond the formal curriculum. It is worth noting that at this stage we do not support the inclusion of performance indicators on the provision of extra-curricular activities, as they are easily subject to gaming by schools.
- **Progress measures for pupils with Special Educational Needs, eligibility for free school meals, and English as an additional language.**
- **The proportion of teachers with an “outstanding” grade, and the tenure of the head teacher:** parents may also be concerned with the staff composition of the school. Information on average teacher effectiveness (and the variation in teacher effectiveness) would also be extremely valuable to prospective parents (and school governors and head teachers) and we

strongly recommend that the necessary data linkage (between pupil and teacher data) is undertaken in future.⁵

- **The proportion of pupils enrolled on each GCSE (or equivalent) course:** this would provide timely information on subject choices available at the school.

More generally, **we question whether headline measures should be used to inform inspection judgements** (as discussed in this section of the consultation). An inspection that forms an independent judgement of standards provides more information about a school than an inspection judgement informed by headline measures; in an extreme case where inspections are only informed by headline measures, they provide no additional information to parents.

Question 5: Do you think we should collect and publish test data from internal assessments through the Data Warehouse?

We do not agree that the data portal should be used to gather information about non-statutory (internal) tests. Internal tests are currently used to inform pupils' development, not for accountability. Making the scores of internal tests publicly available would have a number of perverse incentives that may reduce pupils' development:

- Schools' use of internal tests could be distorted
- Schools could choose to selectively enter pupils
- Schools could choose the most flattering internal test
- Cheating may not be prohibited

Question 6: What other data could be published to create the right incentives for schools, including special schools, to ensure the best progress and attainment for all of their pupils?

Post-16 outcomes for all pupils (for example participation in further education and employment) should be published.

Supplementary assessments should be made available so that a larger proportion of pupils can access assessment, and this information should be made

⁵ Creating a quantitative measure of teacher effectiveness relies on a link between pupil attainment in a given subject, their prior attainment, and their class teacher.

publicly available (with comparison to the attainment of pupils with a similar background).

Question 7: Do you agree that the Department should stop the collection of Key Stage 3 teacher assessment results?

Before responding to this question, **more information is required about the nature and cost of this collection, as well as information on the reliability and value of the teacher assessments.** We agree that the burden on schools should be reduced, but that submitting KS3 teacher assessments (along with other mandatory information, for example the annual pupil census) is likely to be relatively low cost for schools.

We would also like to know **whether attainment according to KS3 teacher assessments is a useful indicator of pupils' progression.** If not, then there is a case for abolishing (or reforming) KS3 assessment entirely. If so, then there is value in collecting this information for research purposes. To determine whether KS3 teacher assessments give useful information about pupils' progression, we recommend that the **correlation between KS3 teacher assessment levels and other tests (including commercially available tests) is calculated to see how closely KS3 assessment tracks pupil ability.** This should be done before the collection is stopped. We have planned to conduct similar work as part of an evaluation of a school-level intervention in April 2014, and **would be happy to share our results with the Department.**

Question 8: How should we ensure that achievement beyond formal qualifications is recognised?

We agree that achievement beyond formal qualifications can be important for pupils' long term outcomes. For example, the development of non-cognitive skills has been shown to be an important determinant of employability (Caneiro, Crawford and Goodman, 2012⁶). Activities that encourage development of wider skills **should therefore be incentivised** in some way. There is **a large risk of gaming** by schools in whatever measure was designed, however. Instead, the **Department should determine the eventual outcomes it thinks are important** (and affected by wider skills and learning), for example employability

⁶ <http://www.ifs.org.uk/publications/5986>

and engagement in society, and find suitable and robust ways to measure these independently.

Question 9: How can national sample tests be introduced?

There are currently three functions of GCSE and equivalent qualifications: first, to document the achievement of pupils at a particular age; second, to make schools accountable for the achievement of their pupils; third, to track national standards of achievement over time.

We agree that while schools are held to account for pupils' performance on these exams, they have an incentive to game the system (for example by teaching to the test, manipulating the subjects chosen and choosing the most preferable exam board) as well as through increasing pupils' understanding. This creates a problem when tracking national standards, as it is unclear how much increases in standards are due to increases in learning or effective gaming.

This problem can be **partly addressed through changes to the exam system** (for example eliminating the choice between exam boards) and a completely independent body to monitor the standard of GCSE exams over time. However, it may be necessary to have separate assessments if schools can artificially increase pupils' scores without affecting underlying understanding (gaming). In this case, it is preferable to monitor standards over time through separate assessments. **We broadly support the introduction of separate tests, under the conditions that:**

- The additional tests cover material taught in **core GCSE subjects** (including English and maths);
- The additional tests **can't be gamed** (which may mean that the format for the exam is randomly changed across schools or across years);
- Schools and teachers are made **fully aware that the scores of their pupils will not be publicly reported;**
- **An individual pupil will only take one additional test**, to minimise the burden on pupils around the time of high stakes exams.

One could go further and use these external tests to moderate the different exam board grades. GCSE grades could be rescaled to account for the varying difficulty levels of the different exam boards in a given subject. This would reduce the incentives for exam boards to undertake a "race to the bottom" in terms of exam difficulty. It would also aid employers who could be sure of the comparability of GCSE exams from different boards.

A comment on the implications for school inspection

We agree that it is sensible for Ofsted to use new headline measures rather than old, if these are to contribute to overall Ofsted rankings of schools. However, we would **prefer Ofsted inspections to be produced independently of performance information**; inspections would then provide more information to interested parties, and could be used to validate (or otherwise) the headline measures produced separately. **The value of Ofsted inspections is reduced if they are too reliant on performance indicators.**