

Royal Statistical Society response to DFE consultation on secondary schools accountability

1 May 2013

Introduction

This is a response to the consultation document issued on 7 Feb 2013. A separate document for primary schools is to be issued and will be responded to separately, although many of the issues raised here will apply to primary schools also. It should be noted that subsequent to publication of the consultation document, proposals for reform of the GCSE system were withdrawn. Nevertheless, the issues concerning accountability remain, and the change of policy is likely to have little general effect on these.

The first issue that arises concerns what is meant by 'accountability'. There is no clear definition within the document, save that the data to be published are intended to be used by parents, governors, head teachers and others in order to improve administration, make choices of schools and improve learning outcomes by identifying 'strengths and weaknesses'. The consultation document makes no attempt to explain exactly how the data can be used for such ends and in this response we shall address this issue. Nor is there any indication of the skills that data users will need in order properly to interpret and understand the published data.

As statisticians, we believe that the provision of extensive data does present opportunities for useful insights to be made. We are also aware, however, that a proper appreciation of the limitations of the data is central to their proper use and we will address ourselves particularly to this issue. The document deals with two kinds of data. The first, to which most attention is devoted, is the 'aggregate' data published by schools. The second is the sample data collected for 'national monitoring' purposes. We shall comment on both.

Accountability

While the document makes no clear distinction among the needs of different users, these do in fact differ considerably. Thus, for example, parents will typically, qua parents, wish to use information for school choice, whereas governors and head teachers will generally be concerned with how to use available resources more efficiently. Likewise, many companies offering services to schools will be seeking information that allows them to maximise their income.

We also note that institutions such as schools will be accountable not only for the performance of their students, but also for such matters as their emotional development and their understanding of social affairs more generally. This means that a concentration on readily measurable 'outcomes' such as exam passes should not play too dominant a role. This will in addition help to reduce teaching to the test. We therefore welcome the documents references to 'perverse incentives' that can arise from over-emphasis on one set of factors out of many other relevant ones.

We now look at needs of different users in more detail. Before doing so, we would like to draw attention to the issue of statistical uncertainty. We note that currently the publication of school performance tables on the DFE website does contain confidence intervals and also a description of how these may be used. Unfortunately, these are not headlined and very rarely find their way into media reporting of results. Little attempt, as far as we know, has been made to inform users of the tables and the public at large about the crucial importance

of providing measures of uncertainty. As we outline below, these typically show that many schools cannot be distinguished from the overall average or from their neighbours, and without such uncertainty measures proper interpretation of the tables is impossible. We accept that it is not necessarily an easy matter to convey to a non-technical audience what is involved, but we do believe that it has to be attempted, and indeed is possible using carefully constructed illustrations. The RSS is very willing to lend its expertise in this respect.

School choice

The document states that 'publishing more data helps parents make choices about the school for their children'. No evidence, however, is provided for this statement. Parents and others concerned with choice require essentially a prediction of whatever outcome is being measured, such as exam score, at the time when the students themselves will be sitting for such exams. In the case of secondary schools, this implies a prediction some 6 years ahead. Recent research, using the DFE National Pupil Database, has shown that the uncertainty surrounding such predictions is so large as to make it extremely difficult to separate schools in terms of likely performance of their pupils. Thus, for example, Leckie and Goldstein (2011) in a value added analysis conclude that 'school league tables have very little to offer as guides for school choice'. It follows that, in reporting performance results, this understanding ought to be communicated, and we urge that this is headlined in the future. In terms of published results as an indication of current performance, it is also the case that large uncertainties exist, which makes comparisons among institutions highly problematic.

School improvement

For schools themselves, including teachers and governors and also inspectors, a balanced reporting of strengths and weaknesses is important in deciding on appropriate measures to take. As the document suggests, publication of comparisons can distort the educational aims of schools, as they strive to maximise performance on a few targets at the expense of less easily evaluated factors such as those that may be studied during inspections. We suggest, therefore, that the inspection system is required to reduce the way it is currently strongly informed by headline performance data, especially given the uncertainties involved. We also suggest that inspection reports could take the lead in informing their audience about the issue of uncertainty. In particular, we suggest that there be no attempt to link, in any way, the results of performance tables to resources, directly or indirectly, going to schools. One consequence of this would be a more nuanced view of what constitutes 'failure' and 'success'.

In this respect, we note with approval that the document accepts that a 'value added' approach is an advance on the reporting of a simple average score or pass rate. We hope, too, that this will help to remove the current incentive – supported by anecdotal evidence – to concentrate extra attention on students who are seen as close to target thresholds. We are puzzled, however, by the recommendation that both are 'headlined', since they will often provide quite different pictures. We suggest, therefore, that the simple 'unadjusted' results are downgraded in importance to the status of supporting material, and that the headlines are given by the value added results, together with associated uncertainty. It is unclear, however, how proposals to remove levels for the National Curriculum, which the government says in that consultation document will mean 'the Department could only collect very limited information at Key Stage 3 in future' will be reconciled with the need to provide a baseline for measures of progress.

Floor targets

The document proposes a 'floor' target for schools to achieve. From what we have already said, it is clear that any such target should not be based upon unadjusted results, whether

these are based upon the proportion of 'A*-C' results or a point score. If such targets (or any other targets) are used, they should at the very least have a value added element. We note the implicit recognition of this in paragraph 2.4 where (value added) progress is stressed. In addition, publication should take account of the uncertainty associated with achieving any given level. In particular, in the interests of equity, schools that cannot reliably be distinguished should not be treated differently. We are, for these reasons, concerned about paragraph 2.6 and 5.4, which highlight floor standards in terms of unadjusted results, and we are unaware of any evidence that existing floor standards have been effective in raising standards. We believe that the proposed threshold measures and associated sanctions are unnecessary and are likely to lead to even more 'perverse incentives', distorting the nature of the educational experiences offered to students. We note (paragraph 7.1) that 'relative' standards are to be used initially, and these will be particularly suspect.

Differential progress

In principle, we welcome the intention to promote information about the progress of different types of students, for example disadvantaged ones. This implies that there are several value added estimates available, one for each type of student. Unfortunately, while this in theory provides more nuanced information, the uncertainties associated with results provided for subgroups of students have even larger uncertainties than those associated with the totality of students. This implies that, for accountability purposes, they are likely to be largely useless.

The current procedure for publishing these in the secondary school performance tables is inadequate: uncertainty intervals are not provided, and numbers involved are typically small, so that most comparisons will be scientifically meaningless. We strongly urge, therefore, that uncertainty intervals are made available so that the validity of any comparisons can be established easily by users.

Sample assessments

We welcome the proposal to carry out an independent sample assessment to gather information about the system as a whole. We assume that this would be linked with the existing national pupil database, and its usefulness would be enhanced by including all schools, especially independent ones. While the examples of PISA and TIMSS are given, the usefulness of these is limited by their cross-sectional nature. If a new system is introduced, it should, from the outset, collect longitudinal data, which through the existence of the NPD should be easy to implement. It should also not be restricted to performance measures, but collect social and other relevant data as well as data on the school context.

Evaluation

Finally, we urge the DFE to give serious attention to the thorough piloting and evaluation of the accountability framework. We are concerned that changes to the monitoring of the educational system have been implemented without thorough piloting of crucial aspects, and that evidence about negative consequences has been discounted.

Like the government, we are concerned that students' lives can be strongly affected by decisions such as those discussed, and we believe that such decisions must therefore be based upon the best evidence available, and every effort made to obtain such evidence. Insofar as the RSS has a body of expertise among its fellows, it is very happy to contribute to any such effort.

References

Leckie, G. and Goldstein, H. (2011). A note on the limitations of school league tables to inform school choice. *JRSSA*, **174**, Part 3, pp. 833–836.