CMPO Working Paper Series No. 03/058

Which Ranking? The Use of Alternative Performance Indicators in the English Secondary Education Market

Deborah Wilson

CMPO, University of Bristol
Department of Economics and International Development, University of Bath

January 2003

Abstract
Performance tables for UK secondary schools have been published annually since 1992. In 2003, for the first time, these tables additionally include a measure of the educational 'value added' by a school to its pupils. This paper provides the first large scale analysis of the likely impact of the new value added performance indicator on the rankings of schools in the resulting league tables. Our analysis employs a national dataset of matched exam results, recently released by the Department for Education and Skills (DfES), which includes the results of the cohort of pupils who sat Key Stage 3 (KS3) exams at age 14 in 1997, and GCSE (or equivalent) exams at age 16 in 1999; this yields data on over half a million pupils. Using this dataset we have replicated five performance indicators which have been or will be published in the UK. In particular, we focus on the key pre-2002 PI, the percentage of pupils gaining at least five GCSEs or equivalent at grade C or above (%5A*-C), and the new value added indicator (VAcap). At a national level, we investigate the relationships between both the indicators themselves and the rankings which result. We then focus on one LEA, Bristol, and show to what extent school positions in the league tables are sensitive to the PI employed. We find a low degree of correlation between %5A*-C and VAcap and the resulting rankings, both at national and local level. This is reflected in the degree to which Bristol schools' ranking positions change when different PIs are employed. We conclude that value added does provide a more accurate measure of school performance and hence should help parental choice. We provide evidence, however, which suggests that a single PI, representing a school average value added score, may not be sufficiently informative.

JEL Classification: H4, I21, L33

Acknowledgements
We thank the Leverhulme Trust for funding this research. Many thanks to Brendon McConnell for excellent research assistance and to Adele Atkinson , Simon Burgess and Carol Propper for comments on previous drafts. Thanks to the Department of Education and Skills for supplying the data. The views expressed in this paper are solely those of the author, and may not reflect those of the DfES.

Address for Correspondence
Department of Economics
University of Bristol
8 Woodland Road
Bristol
BS8 1TN
Tel: +44 (0)117 928 7795
D.Wilson@bristol.ac.uk

CMPO is funded by the Leverhulme Trust.

1 Introduction

Performance tables for UK secondary schools have been published annually since 1992. In 2003, for the first time, these tables additionally include a measure of the educational 'value added' by a school to its pupils. This paper provides the first large scale analysis of the likely impact of the new value added performance indicator on the rankings of schools in the resulting league tables, and investigates its effect on both schools' behaviour and parents' ability to choose the right school for their child.

The UK value added performance indicator (PI) essentially measures the progress of a school's pupils, relative to the national average, between the ages of 14 and 16. Pre-2002, the indicators provided alternative measures of pupils' attainment in the national exams sat by 16 year olds (GCSEs and GNVQs). Such indicators, based on raw test results, have been criticised on the grounds that they may be measuring differences in schools' intakes as well as differences in their performance. The value added indicator incorporates a proxy for intake (performance at age 14), and hence should be better able to isolate the actual performance of a school with regard to the educational progress made by its pupils between the ages of 14 and 16. As we show, the rankings of schools are sensitive to which type of indicator is used to measure performance.

But why do rankings matter? Annual school league tables were introduced in the UK as part of the introduction of the quasi market in education. We shall consider two implications of possible changes in school ranking positions for the working of this market. First, league tables provide data on relative school performance which in turn should enable informed parental choice. If it is the case that the publication of the new value added PI alters those rankings, how informative have previous league tables been? Second, how may the schools react to such a change in ranking position: by improving performance or improving league table position? We discuss the incentives created by the alternative types of indicator and show how their publication may have undesired consequences in terms of behavioural response.

Our analysis employs a national dataset of matched exam results, recently released by the Department for Education and Skills (DfES), which includes the results of the cohort of pupils who sat Key Stage 3 (KS3) exams at age 14 in 1997, and GCSE (or equivalent) exams at age 16 in 1999. We restrict our analysis to English state maintained secondary schools, for which we have data on over half a million pupils. Using this dataset we have replicated five performance indicators, which have been or will be published in the UK. In particular, we focus on the key pre-2002 PI, the percentage of pupils gaining at least five GCSEs or equivalent at grade C or above (%5A*-C), and the new value added indicator (VAcap). At a national level, we investigate the relationships between both the indicators themselves and the rankings which result. We then focus on one LEA, Bristol, and show to what extent school positions in the league tables are sensitive to the PI employed.

We find a low degree of correlation between %5A*-C and VAcap and the resulting rankings, both at national and local level. This is reflected in the degree to which Bristol schools' ranking positions change when different PIs are employed. We conclude that value added does provide a more accurate measure of school performance and hence should help parental choice. We provide evidence, however, which suggests that a single PI, representing a school average value added score, may not be sufficiently informative.

The rest of the paper is structured as follows. Section 2 describes the use of performance indicators in the UK education sector. In section 3, we focus on specific types of PI and the incentives created by their publication. In section 4 we present evidence using our replication of the UK PIs, both at national level and for Bristol Local Education Authority (LEA). Section 5 concludes.

2 Performance Indicators in the UK Education Sector

The UK education sector is large and complex. In 2001 there were over 400,000 Full Time Equivalent teachers and 8.4 million pupils in 25,760 schools (of which 7% are independent) (Burgess et al 2001). Each pupil receives at least 15,000 hours of compulsory ‘treatment’ (Fitz-Gibbon 2000). Until recently education has been treated as a procedural organisation with concomitant emphasis on the role of professionalism. Now it is subject to relatively high levels of public monitoring. There are two main systems for measuring performance in education in the UK, OFSTED reports[1] and the publication of summary performance indicators.

PIs appeared as part of the introduction of quasi-market forces in education following the Education Reform Act of 1988 (Le Grand 1991; Glennerster 1991). The basic principles on which this original reform was based are still present in the current education system: local management of schools, with devolved budgets calculated on a per-capita basis, overlapping catchment areas and open enrolment. In order to maintain resource levels a school must attract sufficient pupil numbers; the overlapping catchment areas create the potential for competition for pupils which should drive up the quality of provision. Parental choice of school is informed by the annual publication of performance indicators, now commonly referred to as the league tables.

Currently two key indicators are published: absences (authorised and unauthorised) and pass rates at GCSE. There is evidence of manipulation of the former by schools through the reclassification of truancy to be ‘excused absences’ (Fitz-Gibbon 1996). With regard to the latter, data is now published in four categories: the percentage of a school’s pupils gaining at least 5 grade A* - C passes, the percentage gaining at least 5 A* - G, the percentage who pass no GCSEs and the average point score.[2] The first three of these are examples of target indicators, and we shall discuss the incentives created by such

[1] See Burgess et al (2002) for more on OFSTED reports.
[2] For further information, go to http://www.dfes.gov.uk/performancetables/ .

performance measures in education in more detail below. From 2002 a value added measure for each school will additionally be published. The key target indicator, i.e. the one to which most attention is paid, is the percentage of pupils gaining 5 or more GCSE passes at grades A* - C (perhaps because this is the minimum requirement for continuation of study). Out of the current indicators, therefore, it is this one which we concentrate on in this paper, comparing it to alternative measures of value added, both in terms of the incentives created and the different rankings which are produced.

There are usually various, non mutually exclusive, aims of any performance monitoring system. In the UK education sector, the aim of the publication of summary PIs is generally considered to be twofold: to improve the performance of individual schools, thereby raising standards in education, and to provide information on individual school performance in order that parental choice is better informed. Of course the two are linked through the working of the quasi market.[3] Any one PI will be at best an imperfect summary measure of the complex multiple goals facing a school; indeed, some of these goals or tasks may be inherently unmeasurable (Dixit 1999). So any single PI will only measure a specific subset of tasks. This has consequences regarding possible undesired behavioural responses of self-interested schools. As we discuss below, the specific responses depend on the actual PI employed or, more specifically, published.

We now consider the distinct incentives created by first target indicators and then alternative measures of value added. What we concentrate on is the perverse or undesired behavioural responses to the publication of such performance indicators. We are not suggesting that desired responses have not also been induced. Indeed, it is certainly the case that performance in the UK education sector – as measured by raw exam scores – has improved since the introduction of the quasi-market and the publication of the first league tables. It is, however, difficult to isolate which element(s) of such a huge programme of reform has had an impact on outcomes (Burgess et al

[3] There is a parallel here with the use of report cards in some US states (Ladd and Walsh 2002). Published performance data is increasingly being used to achieve an improvement in US school performance without increasing overall levels of spending (Figlio 2001). Indeed, in January 2002 President Bush signed the "No Child Left Behind" Education Bill into law. This mandates the use of accountability systems based on student test performance (Cullen and Reback 2002).

2002). Bradley et al (2000) provide evidence of the impact of the use of PIs in the UK education market; Hoxby (2001) of their impact in the US.

3 Alternative Performance Indicators and Incentives Created

(a) Target Indicators

Target indicators have so far been the most commonly employed type of PI in education. They represent the "% > X" criterion in reporting (Fitz-Gibbon and Thymms 2002). So in the UK, for example, the key indicators include the percentage of a school's pupils gaining at least 5 A* - C grades at GCSE (or equivalent, see below); in the US, various target indicators are published in different states: see Figlio and Page (2001) for details on the Florida scheme. As Fitz-Gibbon and Thymms point out, the "% > X" criterion introduces an arbitrary dichotomy into continuous data. Moreover, this dichotomy – the X – is measured in terms of raw output such as test scores. It is this feature that creates the particular forms of gaming associated with target indicators. Of course, it should be remembered that it is the publication of such indicators that creates the incentive to game: the use of such performance measures as an internal management tool should be treated quite separately (Thomas 1998; Saunders 1999).

Three main responses to such target indicators have been identified. The evidence of such behaviour is primarily anecdotal. First, whenever an arbitrary dichotomy is introduced, it will focus agents' attention on the borderline (Fitz-Gibbon 1996). So, given that a key performance target in the UK is the percentage of children achieving 5 or more GCSE passes graded A* - C, we may expect schools to shift their activities or target their resources to pupils who are expected to just miss the target in the absence of (extra) intervention. This may be to the detriment of pupils at either end of the ability distribution in terms of the resources allocated to them and may or may not be welfare improving. It certainly seems to be a distortion of the activities of teachers from what they would do in the absence of such a target and provides one example of 'measure fixation' as discussed by Smith (1995) (Fitz-Gibbon 1997). Examples of such behaviour include the use of volunteer helpers with weaker students; strategic mentoring by teachers; after-school coaching and holiday revision courses (West and Pennell 2000). Such practices may be employed in general if attendance on the course is compulsory. If, however, this is not the case, the incentive would be to remove weak students from the course (Fitz-Gibbon 1996). Again, there is anecdotal evidence from the UK that schools are removing weak students from GCSE courses and putting them into GNVQ equivalents (Times Educational Supplement 2002).[4]

Second, there may be the incentive to reclassify weak students in order that they are not eligible for sitting the tests that are the subject of the target indicator. Figlio and Getzler (2002) and Cullen and Reback (2002) both provide evidence of such practices in the US. Since the signing of the No Child Left Behind Act states must design accountability systems based on the fraction of students attaining a certain level in reading and mathematics (Figlio and Getzler 2002). As well as creating incentives to 'teach to the test', schools have the incentive to reclassify weak students into the special education categories that are exempt from such mandatory testing. In the UK, there is some indirect evidence that the publication of league tables created the incentive to exclude certain types of pupils from the school (not just from sitting certain exams as in the US). Gillborn (1996) quoted in West and Pennell (2000) reports a tripling of permanent exclusions in the three year period from 1993/94 (the first league tables were published in 1992). Gerwitz et al (1995) introduce the term 'constructive exclusion' to describe ways in which schools may put pressure on certain children in order that they leave 'voluntarily' (West and Pennell (2000)). See also Wilson (2001) for a longer discussion of the issue of constructive exclusion and its parallels with involuntary disenrolment from health care service providers.

Third, schools now have the incentive to also exclude weak students ex ante, i.e. to engage in selection or cream skimming at the point of admission (Wilson 2001). This is

---

[4] The DfES has calculated a rate of equivalence between GCSE and GNVQ for the purposes of the PI. For example, one full intermediate GNVQ is worth 4 GCSE passes at grade A*-C. See http://www.dfes.gov.uk/performancetables/vap_01/docD.shtml for further details.

one consequence of the use of a raw output measure of performance, from which it follows that, for the same quality of education received, the better the input, i.e. the higher the ability of children admitted to the school, the better the output and hence the higher the school's relative position in the league tables. The deregulation of school admissions procedures as part of the 1988 reform package gave locally managed schools greater discretion over which pupils they admitted. Gerwitz et al (1995), Whitty et al (1998) and West and Pennell (2000) all discuss ways in which a school can design the procedure in order that only certain types of pupils (and parents) are attracted to the school. These include the use of complicated admissions forms and the use of pre-admission interviews. In 2002 the highest achieving comprehensive school in England (it scored 100% at 5 A* - C or over) was reprimanded by the Local Government Ombudsman for its pre-admissions interview policy (Cassidy 2002). Coopers' Company and Coborn School is a faith school and is therefore allowed to interview prospective pupils only to ascertain their religious commitment. It was additionally asking questions about their hobbies and where they lived, which was seen as potentially forming part of a non-faith based selection procedure.

Finally, note that the second and third responses to target indicators described above are both ways of altering who takes the relevant tests (see Meyer (1997) for additional 'creaming' strategies). This matters since who takes the test is correlated with measured performance (raw output), i.e. a target PI doesn't explicitly account for heterogeneity in any population of students, so the school has the incentive to tailor the population to improve its indicator. It follows, therefore, that one way to reduce the incentive for such behaviour is for the indicator itself to better account for such heterogeneity.[5] This is one argument supporting the introduction of value added measures of school performance. Before we consider the incentives created by such measures, however, it is worth discussing what we actually mean by the term itself. As we shall show, the way value added is defined and/or measured has implications for the incentives thereby created.

[5] There is a parallel here with the use of risk adjusted capitation payments in health care systems (Wilson 2001).

(b) Value Added

An alternative performance indicator is a school level measure of 'value added'. The aim of such a PI is to isolate the impact of the school on the progress made by its pupils. We usually consider value added in terms of the progress made, or change in performance of, one cohort of pupils across successive time periods. In the UK, for example, the value added PI is a measure of the relative change in performance of pupils between Key Stage 3 and GCSE (exams sat by all pupils at ages 14 and 16 respectively). We therefore concentrate on this formulation. This notion of the value added within the same cohort through time builds on the economic definition of the term. For economists, 'value added' is used to describe the difference in value between the materials 'bought in' and the finished product, i.e. it measures the value added by the process of production. The term was adopted by educationalists interested in the value added by the education production process (see Saunders (1999) for a discussion of the history of its use). To investigate the alternative ways educational value added may be defined, and its implications, consider the following education production function (taken from Ladd and Walsh (2002, page 3)):

$$A_{it} = \lambda A_{it-1} + \alpha_t \mathbf{S}_t + \beta_t \mathbf{F}_{it} + \varepsilon_{it} \qquad (1)$$

where A_{it}, A_{it-1} represents the achievement of student i in year t and year t-1 respectively; $\mathbf{S}_t$ is a vector of school characteristics; $\mathbf{F}_{it}$ a vector of measurable family background characteristics and ε_{it} the random error term.

$\mathbf{S}_t$ represents the impact of the school on the change in pupil performance between the two time periods, once we account for prior attainment and family background characteristics. In his formulation, Meyer (1997) talks in terms of η_s, the total school performance parameter. Two points should be noted at this stage. First, both $\mathbf{S}_t$ and η_s are absolute measures. Meyer (1997), however, employs what he calls the conditional mean format to transform educational value added into a relative concept. Indeed, in contrast with the economic use of the term, educational value added is generally considered relative to some benchmark. We discuss this in the context of the PI published in the UK below. Second, there are issues surrounding both the inclusion of $\mathbf{F}_t$ and what characteristics such a vector should contain. We do not consider these issues here because the UK calculation of value added does not incorporate any such explanatory variables. For more on this, see Thomas and Mortimer (1996); Meyer (1997); Goldstein (2001); Fitz-Gibbon and Thymms (2002;); Ladd and Walsh (2002).

We can isolate the impact of the school in greater detail, at least in theory. Ladd and Walsh (2002) point out that the vector $\mathbf{S}_t$ includes factors that are both within and outside school control. Reconsider the education production function (again, from Ladd and Walsh (2002, page 3)), which now splits $\mathbf{S}_t$ into exogenous and endogenous factors:

$$A_{it} = \lambda A_{it-1} + \alpha_{Rt}\mathbf{R}_t + E_t + \beta_t\mathbf{F}_{it} + \varepsilon_{it} \qquad (2)$$

where $\mathbf{R}_t$ is a vector of factors outside the control of the school and E_t a measure of the effectiveness of the school's staff and administration. $\mathbf{R}_t$ may include the exogenously determined resource levels and school composition, for example.

Again, several points should be noted (Ladd and Walsh 2002). First, the true value added by a school's environment should only comprise the change in performance that can be attributed to factors endogenous to the school, represented here by E_t. In Meyer's (1997) terms, this is the measure of intrinsic school performance. Second, one implication of this formulation of the education production function is that it is not possible to measure the intrinsic performance (or effectiveness) of a school without controlling for resource levels. This way of calculating value added is therefore sympathetic to the original economic definition of the term. As we shall see, however, in practice, resource levels are generally not included in educational value added calculations. Indeed, it may be difficult to specify the exogenously determined level of resource. As Ladd and Walsh state, there is a further difficulty in separating the endogenous from the exogenous elements of a school's environment. Specifically, to what degree are we able to consider school composition as exogenous, given that peer group is one factor in neighbourhood and/or school choice decisions?

Perhaps whether or not we actually want to separate the exogenous from the endogenous elements of school environment depends on what the resulting performance indicator is trying to capture. More specifically, the alternative PIs may be applicable to the objectives of the different actors in the education market. The parent wants to choose the school which maximises his child's educational achievement, broadly defined. The total school performance parameter is hence relevant: the parent is interested in the whole package: teacher effects, resource levels and peer group. The government, however, is more concerned with improving performance for a given level of resource, hence the measure of intrinsic school performance is relevant, as that better isolates school (teacher plus administration) effectiveness.

So the term value added can be used to describe a range of measures of school performance, each of which capture distinct elements of the education production process. How does the above analysis relate to how value added PIs are actually calculated in practice? The next section describes how value added is calculated in the context of the secondary education market in the UK; for descriptions of alternative US state measures, see Ladd and Walsh (2002) and Figlio and Page (2001).

Value Added: UK Calculation

A school level measure of value added is published for the first time in the 2002 secondary school performance tables. This PI provides information on the progress made by one cohort of pupils between the ages of 14 and 16. All pupils sit Key Stage 3 (KS3) exams at the age of 14 in English, maths and science. At the age of 16, the end of compulsory schooling, they take their GCSE and/or GNVQ exams. In order to calculate value added, all pupils are categorised into one of 18 bands depending on the mean score they attained at KS3 across the three subjects. This provides the input measure for the value added calculation. The output score is total GCSE (or equivalent GNVQ) points.

The median output score is computed for each KS3 band: this is the expected outcome for all pupils within that input band. Each pupil's total GCSE score is then compared with the median point score for all pupils in the same category. If it is equal to the median, they have gained zero value added; a positive value added implies a higher than expected score, given KS3 performance, while negative value added suggests a lower than expected performance at GCSE. A school level measure is then calculated by taking the mean of individual pupils' value added scores, so a positive value added PI indicates that, on average, the school's pupils have performed better than expected at GCSE, given their KS3 results.[6] The published PI is actually centred around 100 and not 0, presumably to avoid confusion about the meaning of a 'negative' value added score. In fact, what is published in the performance tables is a variant on this basic method: the output score – and hence the value added calculation – is capped at the eight best GCSE results or equivalent.

We discuss the particular implications of both versions of the value added PI below, but first it is useful to consider this method of calculating value added with reference to the preceding discussion. The UK educational value added PI employs value added as a relative concept. Family and/or background characteristics are not controlled for, nor is there any attempt to isolate exogenous from endogenous school effects and hence there is no consideration of efficiency in the education production process. Of course, much of the departure from the theoretical ideal outlined above is simply due to practical data limitations. What this PI does provide is a measure of total school performance; it may therefore be particularly suited to the issue of parental choice of school. Moreover, given that now prior attainment is explicitly accounted for, the new PI provides a more accurate measure of the impact of the school environment than the target indicator which uses just raw test scores.

Even if we put to one side the arguments for and against the inclusion of explanatory variables other than prior attainment, issues arise regarding the publication of such a

[6] Go to http://www.dfes.gov.uk/statistics/DB/PER/p0313/index.html for further information on the value added calculation.

summary measure of total school performance. In the next section we therefore consider the various alternatives, focusing in particular on the incentives created by them.

Value Added: Alternative Measures and Incentives Created

We have so far discussed two variants employed in practice (with reference to the UK): the basic value added calculation and the one which caps the output measure at the best eight exam scores. Let us now consider the incentives created by these alternatives. Recall that it is the publication of such indicators that may create incentives to game the system and that the responses to their publication depend on the actual PI employed.

By definition, a general feature of any value added measure is the inclusion of a measure of an input score. Ceteris paribus, the lower the input score, the higher the value added. It follows, therefore, that one incentive created by the publication of such indicators is to depress the input score used. So in the UK, following the publication of the 14-16 value added PI, schools will have the incentive to depress their Key Stage 3 scores as this provides one way to boost their position in the value added performance table. Crucially, currently, the input score is internal to the school and hence is within its control. As Fitz-Gibbon (1997) points out, this provides one argument for using an alternative, exogenous, input score: in the UK this would mean using the results from the Key Stage 2 tests, taken by pupils at age 11 prior to starting secondary school.

The choice of output score also has an impact on the calculation of value added. Let us consider the alternatives, each of which may be used to formulate a school level PI of mean value added across its pupil population. We first consider measures which are aggregated across at least several subjects, before looking at the issue of subject level value added.

Total Score

This was the output used in the DfES pilots for value added, with the method described above. As Fitz-Gibbon (1997) states, such a total measure reflects the teaching effort across all subjects. There may, however, be a confusion between quality and quantity: the incentive is to put students through more exams in order to boost the indicator.

Average Score Across All Subjects

Using average score as the output measure avoids the quantity / quality confusion from using total score: the two are only equivalent when all students take the same number and combination of subjects (Fitz-Gibbon 1997). Now the school has the incentive not to put students in for exams if a good result cannot be guaranteed.

Average Score Across a Subset of Subjects

Given the wide range of subjects taken at GCSE and GNVQ, a more accurate measure of relative school performance may be one which uses only those subjects taken by all pupils – the core curriculum. This will provide only a partial measure of school or teacher effort. Moreover, any measure which only incorporates the results from a subset of subjects will create an incentive to concentrate on maximising the results of those which count. This distortion of effort can be thought of as a form of 'tunnel vision' (Smith 1995), concentrating only on the subjects relevant for the indicator (Fitz-Gibbon 1997).

Total Score, Capped at Eight Best Results

As stated above, this is the value added measure to be published in the UK secondary school league tables from 2002. This formulation avoids the problem of distortion of effort towards a particular subset of subjects as the eight best can be any from the full range of subjects offered. There is no longer an incentive to enter students for as many subjects as possible: indeed, there is no incentive to enter students for more than eight: now the incentive is to maximise the scores of each pupil in each of the exams he is entered for, up to the cap. This measure therefore gets round the quantity versus quality issue, without being subject specific. There may be an incentive to distort effort away from the top end of the ability distribution, however, due to the imposition of the cap. In effect, there is a parallel here with the target indicator: capping the value added measure at the eight best scores introduces an arbitrary dichotomy into continuous data and thus focuses schools' attention on the borderline. The incentive now is to focus effort on those students who may not get eight (good) GCSEs in the absence of extra intervention. This may be to the detriment of those students particularly at the top of the ability distribution: if a student is expected to get eight good passes, there is now no incentive to improve her position beyond that point.

Subject Level Differences and Differential Effectiveness

All the indicators we have so far discussed employ an output score based on some aggregation across subjects in order to calculate the school mean value added. As Fitz-Gibbon and Thymms (2002) point out, the use of such aggregate measures may actually confuse or hide two issues. First, there may be differences in the value added across different subjects, and second there may be differences in the value added by the same school to pupils at different points in the ability distribution: in other words, a school mean value added PI will not encapsulate any degree of differential effectiveness in school performance (Goldstein 2001; Thomas1998).

First consider the incentives created by the publication of an aggregate measure, as opposed to subject level value added indicators. The publication of such a measure, which hides between-subject differences, creates the incentive to keep students out of difficult subjects and/or difficult syllabuses or exam boards (Fitz-Gibbon 1997). For this to be a problem, there needs to be systematic differences in performance (output score) across different subjects. Fitz-Gibbon and Thymms (2002) provides evidence that such differences do exist. Such an incentive has knock-on effects in terms of what students go on to take at A-level. For example, if it is the case that it is easier for a student to attain a certain score in English GCSE than in maths, the publication of an aggregate PI (value added or target) may mean fewer students are entered for maths. Hence fewer will go on to study maths at A-level. This then creates a wider problem if the economy needs more mathematicians. In the following section we provide evidence that systematic differences do exist between pupils' performance in English, maths and science at GCSE.

Second, how informative is school mean value added? Recall that a school's value added PI is calculated by taking the mean of its pupils' individual value added scores. It hence gives the average value added by the school (relative to the national expected outcomes). Alternatively, we can consider this PI as telling us the value added by the school to the educational attainment of the average pupil. Such an aggregate measure is consequently not able to provide information on any degree of differential effectiveness in school performance (Goldstein 2001; Thomas 1998), i.e. it may hide differences in the value added by a school to different groups of pupils. In what follows, we focus on possible differences across the ability distribution: see Thomas (1998) for more on differential effectiveness across alternative groupings (with regard to gender, ethnicity, income, for example).

Consider that a school has a range of pupils of differing abilities. It may be the case that the school is particularly effective – creates high value added – for the low ability pupils, while not performing so well for those at the high end of the ability distribution, or vice versa. The publication of a PI based on school level mean value added hides such differences. Why does this matter? Recall that one aim of the publication of such performance indicators is to help the parents make an informed choice regarding the best school for their child. If it is the case that schools exhibit a large degree of differential effectiveness with respect to ability, such PIs are only informative to the parents of the "average" child. If parents have some information regarding their own child's ability, they can only make an informed choice if they know how effective a particular school is for children of that ability. If, however, schools are consistently effective across the ability range (or parents have no knowledge of their child's ability), then the mean value added PI provides sufficient information to all parents. Below we present some preliminary evidence on the extent of differential effectiveness by ability in a subset of English secondary schools.

In summary, the publication of performance indicators may lead to unintended and/or undesired behavioural responses. The incentives created, and hence the responses induced, depend on which PI is employed. One aim of the publication of such indictors is to provide information to parents in order for them to exercise choice in the education market. The extent to which schools' positions in the resulting performance tables are dependent on the particular PI is hence central to the incentives created by its publication. There is evidence that alternative indicators do lead to different rankings (Thomas and Mortimer 1996; Thomas 1998; Figlio and Page 2001). The next section adds to this body of evidence.

4 Evidence from the English Secondary Education Market

The dataset we employ for this analysis is one of the national matched exam datasets recently released by the Department for Education and Skills (DfES). These datasets contain the matched exam results of pupils for Key Stage 3 and GCSE/GNVQ, national exams usually sat at the ages of 14 and 16 respectively. Here we present results for the 1997 – 1999 cohort. We restrict our analysis to state maintained schools in England (see Atkinson and Wilson (2003a) for more details on the dataset). Summary statistics are presented in Table 1. We have data on over half a million pupils in 3,129 schools, 89% of which are comprehensive. Table 2 shows the average results in each of the exams by gender. At Key Stage 3 (KS3), girls outperform boys in English, while boys narrowly outperform girls in science and maths. By GCSE, girls come out on top by 5 points on average, equivalent to an additional GCSE grade C (see Atkinson and Wilson (2003b) for more on the widening gender gap in English secondary schools).

Using this dataset we have re-created five of the performance indicators used in the English secondary schools performance tables:

(1) Percentage of pupils gaining at least 5 GCSEs at grade C and above (%5A*-C)
(2) Average point score at GCSE (APS)

(3) Value added between KS3 and GCSE (VA)

(4) Value added between KS3 and GCSE, capped at the best 8 exam results (VAcap)

(5) Average point score at GCSE, capped at the best 8 exam results (APScap)

(1) is the pre-2002 key target indicator; (2) is an additional output-based indicator; (3) is the value added measure employed in the DfES pilot schemes, calculated using the method described above; (4) is the value added PI which is to be published from 2002, using a capped output score. Because of this, the published average point score PI is now also to be capped, i.e. (1), (4) and (5) will all be included in the performance tables from 2002.

We examine three questions:

(i) What are the relationships between the alternative PIs? In particular, between %5A*-C and VAcap?

(ii) How sensitive are schools' positions in the performance tables to the use of these alternatives?

(iii) How informative is a school mean value added PI? Does such aggregation hide differences either at subject level or across the ability distribution?

Comparison of Alternative Performance Indicators

Consider Table 3, which shows the correlations between the five performance indicators using the national dataset. The key point to note is the low degree of correlation between the two different types of indicator, i.e. between the output based and the value added measures of performance: 0.3641 between VAcap and %5A*-C, for example. We investigate this further below. As expected, the three output based PIs are highly correlated, as are the two value added indicators. Note that there is not perfect correlation between the two value added PIs, suggesting that the introduction of the cap does have some impact. In Table 4 we split the pupil population into deciles based on KS3 performance and look at the correlation between VA and VAcap at different points in the distribution. The degree of correlation decreases as we move up the deciles, i.e. the two measures diverge, suggesting that the imposition of the cap may indeed provide the scope for specific incentives, not found with the basic VA measure, to distort effort away from the top end of the ability distribution.

So there is a lack of correlation between the output based and the value added measures of performance. In order to investigate this point further we focus on the relationship between the key pre-2002 performance indicator, %5A*-C, and the new indicator, VAcap. Consider Figure 1, in which each dot represents a secondary school and which illustrates the lack of relationship between the two indicators. In Figure 2 the same data is shown. Now, however, the schools are represented by the numbers 1 – 5. We calculated the mean KS3 score for each school and split the resulting distribution into quintiles, where 1 is the lowest and 5 the highest. The numbers 1 – 5 therefore proxy the relative average ability of each school's pupils at age 14. And now a clear pattern emerges: the schools which achieve the highest percentage of pupils gaining at least 5 GCSEs at grade C or above are consistently represented by the number 5, while the highest achievers in terms of VAcap show a spread of numbers across the KS3 quintiles. There is a high correlation between performance at 14 and at 16; output based PIs such as %5A*-C cannot distinguish between the two. Of course this provides one argument for the introduction of VAcap as a more accurate measure of school performance.

Comparison of Alternative Ranking Positions

Let us now consider the relationships between the rankings of schools which result from the use of the alternative performance indicators. We do this in two ways. First we created a national ranking, 1 – 3129, for each of our replicated PIs. The correlations between these rankings are presented in Table 5a. Given that performance tables, and hence school rankings, are generally presented for individual LEAs, however, we also created rankings tables for each LEA and then calculated the weighted average (based on pupil numbers), again for each of the five PIs. Table 5b shows the correlations between those rankings. In fact the two methods provide similar results. Not surprisingly, there is a similar picture between the resultant rankings as there is between the PIs which give rise to them. In particular, there is a consistent lack of correlation between the rankings based on the two different types of indicator (output based and value added): 0.4067 between rankVAcap and rank%5A*-C, for example. This suggests that the positions of schools in the league tables will change depending on which PI is employed, i.e. that relative school performance is sensitive to whether the basis for measurement is output or value added. We show this to be the case at LEA level below.

First, it is interesting to investigate further the relationship between the two key ranking systems at national level, as we did above for the underlying indicators. Consider Figure 3. Here we plot rankVAcap against rank%5A*-C, with the highest ranking schools being the furthest from the origin. This figure further illustrates the lack of relationship between the two bases for ranking performance.[7] In Figure 4 we replace the dots with numbers 1 – 5, representing school KS3 mean quintiles, as in Figure 2. And again a clear pattern emerges: the best performing schools according to rank%5A*-C are those represented by number 5, the worst by number 1, i.e. those schools which achieve a high position in the current league tables are generally those drawn from the highest quintile of our proxy for ability (mean KS3). If we consider the best performing (highest ranked) schools in terms of rankVAcap, however, there is a good spread of 1 – 5. The implication is that ranking schools on the basis of raw output may ‘flatter’ those with a high KS3 mean and do the reverse for those with a low KS3 mean; such rankings may in fact reflect differences in ‘input’ or ability rather than differences in actual school performance. Rankings based on value added account for such heterogeneity across school populations and as such create a more level playing field: Figure 4 shows that it is possible for schools to achieve a high ranking on the basis of their value added regardless of their position in the KS3 distribution.

So what is the likely impact of the introduction of the new value added performance indicator on individual schools’ relative positions in the performance tables? To what

[7] Note that for Figures 3 and 4 we have used the national ranking of schools, 1-3129, the results for which are given in Table 5a.

extent are rankings sensitive to the type of PI employed? In Table 6a we consider these questions with reference to state maintained schools within Bristol Local Education Authority. Using our data we have replicated the 1999 DfES performance table for these 22 schools for the key target indicator %5A*-C and have also calculated the rankings that would have resulted from the use of both VA and VAcap in that year.[8] Table 6a also provides information on the admissions policy of each school (comp = comprehensive; sel = selective (grammar)) as well as religious denomination. All state maintained schools in Bristol are co-educational.

As we would expect, given the analysis at national level, there are some differences between the rankings created by the use of the two value added PIs. But these are small compared to the differences between either of the value added rankings and that based on the target indicator. Let us concentrate on the comparison between rank%5A*-C and rankVAcap. Some of the movements up and down the performance table are substantial: St George Community School moves from 16th to 4th when we take account of its pupils' KS3 performance; Hengrove from 19th to 10th. Conversely, Henbury and St Thomas More move down from 8th and 9th to 21st and 22nd respectively. Table 6b shows just how little correlation there is between the two ranking systems for Bristol. It is obviously not possible to draw general conclusions or find specific patterns of such movements on the basis of one year's data from one LEA. What this exercise highlights, however, is just how sensitive schools' relative performance is to the basis for measurement of that performance.

In summary, performance indicators, and the rankings which result from their use, are heavily dependent on whether or not a measure of input is included, i.e. on whether output or value added is the basis for measurement. There is a systematic relationship between performance as measured by the target indicator and average results at Key Stage 3, which provides one argument for the use of a value added indicator as a more

---

[8] Our calculations produce identical rankings except that in the DfES table: (i) Brislington is 8th, Henbury 7th; (ii) Brislington and St Thomas More are tied at 8th; Whitefield, Monks Park and Lawrence Weston tied at 12th; Hengrove and Withywood tied at 19th. It should be noted that we were unable to replicate the published rankings for average point score.

accurate measure of school performance. League tables based on VAcap should, therefore, be more informative to parents choosing which school to send their child, but is one, school mean value added indicator sufficiently informative to enable effective choice?

How Informative is Aggregate Value Added?

Recall that the use of such an aggregate measure of value added may hide two issues: differences in value added across subjects, and differences in the value added by the same school to pupils at different points in the ability distribution (Fitz-Gibbon and Thymms 2002).

First, consider possible differences in subject level value added. We have calculated a school level value added indicator for English, maths and science. The correlations between these and the value added across all subjects are presented in Table 7.[9] The generally low results suggest that an aggregate measure may not be able to capture differences at subject level. What simple correlations cannot show, however, is whether there is a systematic difference in performance across subjects, which creates the scope for the subject level incentives discussed above. Figure 5 provides some evidence that such differences do exist. GCSE point scores in English, maths and science are plotted against Key Stage 3 mean score.[10] Again we are using performance at Key Stage 3 as our proxy for ability. The vertical lines represent the 25th, 50th and 75th percentiles of this distribution. There is a consistent pattern: given KS3 performance, it is easier to gain a higher score in English than in science or maths. This is true to a point beyond the 75th percentile. And science appears systematically 'easier' than maths for pupils below the median. If the PI is not subject specific, therefore, the incentive exists to put pupils – of a

---

[9] Two alternatives are presented for each subject. Given that pupils can sit more than one GCSE in each (English Language and English Literature, for example), we calculated value added on the basis of the maximum score gained in any English GCSE and the mean score across all English GCSEs if more than one was taken. Similarly for maths and science. In all cases, the two alternatives are highly correlated.
[10] Note that the maximum score for each subject was used (see footnote 9). The KS3 mean scores are split into the 18 bands used as the input for the DfES value added calculation.

wide range of abilities – into English rather than maths or science in order to secure a good pass and boost the indicator and resultant ranking position.

Finally, is there any evidence of differential effectiveness across the ability distribution, i.e. do schools provide different amounts of value added for different types of children? In Figure 6 we present evidence of such differential effectiveness. For each state maintained secondary school in Bristol, we have calculated VAcap across the Key Stage 3 distribution (split into quintiles in the figure). There is one line for each school and the identifying number is its ranking according to rank%5A*-C. The lines are not horizontal: there is some difference in the value added by each school at different points in the KS3 distribution (again, our proxy for ability). School 1 (St Mary Redcliffe and Temple) is fairly consistent, as is school 11 (Bedminster Down), so the aggregate measure of average value added may be sufficiently informative. But now consider school 12 (Whitefield Fishponds Community), whose value added decreases as we move up the KS3 distribution; or school 22 (Merrywood), whose value added becomes positive for the highest quintile. The publication of one measure of the average value added by a school may in fact hide differences which are relevant to parents when making the choice of school for their particular child.

5 Conclusion

This paper provides the first large scale analysis of the likely impact of the publication of the new value added performance indicator in the English secondary school performance tables. We find a low degree of correlation between output based PIs and value added PIs, as well as between the resultant rankings. The systematic relationship between rank%5A*-C and average KS3 score suggests that league tables to date have at least partly reflected the ability of a school's intake (at age 14) as well as providing some measure of the impact of school environment on pupil progress. This adds weight to the argument supporting the use of value added as a more accurate basis for the measurement of individual school performance, given that it explicitly takes account of the heterogeneity of the pupil population. Our replication of the Bristol league table illustrates just how sensitive schools' ranking positions may be to the type of PI employed, and suggests that we can expect significant movements up and down the rankings once performance is measured in terms of value added.

Changes in ranking position will of course create different competitive pressures for schools. The aim is that they respond by improving performance, but the complexity of the education production process leaves scope for schools to try and game the system in order to simply improve their league table position. While value added PIs reduce the incentives for selection, present when performance is measured solely on the basis of output, the imposition of the cap in the UK value added calculation may create the incentive for schools to distort effort away from those pupils at the top end of the distribution. A key point to note is that from 2002 the value added PI is to be published alongside the existing output based indicators. Published information on multiple outcomes prevents the school from focusing on only one, and may therefore reduce the incentive (or ability) to game the subsequently more complex performance monitoring system (Ladd 1999; Fitz-Gibbon 1997). Of course there is a trade off between complexity and transparency: the more alternative indicators published, the more difficult it becomes for parents to evaluate the information presented to them.

Given that the UK value added calculation provides a measure of total school performance it may be particularly suited to the aim of enabling informed parental choice. One school level measure of average value added across the curriculum may not be sufficiently informative, however, given our findings on both subject level differences and differential effectiveness. Not all parents have 'average' children: in order for parents to be effective drivers for improvement in the education market, it is essential that they are able to determine which may be the best school for their individual child. The introduction of a value added PI is certainly an improvement on the previous reliance on output based indicators: it may, however, be necessary to additionally consider the impact of differential value added across different student types.

References

Atkinson, A and D Wilson (2003a), *Measuring Pupil Attainment in English Secondary Schools: A Preliminary Analysis*, CMPO, University of Bristol, Working Paper, forthcoming

Atkinson, A and D Wilson (2003b), *The Widening Gender Gap in English Schools*, CMPO Bulletin number 8, CMPO, University of Bristol

Bradley, S, R Crouchley, J Millington and J Taylor (2000), Testing for Quasi-Market Forces in Secondary Education, *Oxford Bulletin of Economics and Statistics*, 62(3): 357-390

Burgess, S, C Propper and D Wilson (2002), *Does Performance Monitoring Work? A Review of the Evidence from the UK Public Sector, Excluding Health Care*, CMPO, University of Bristol, Working Paper 02/049

Cassidy, S (2002), Top Comprehensive School Ticked Off for Interviewing Prospective Pupils, *The Independent*, 07/03/02, page 1.

Cullen, JB and R Reback (2002), *Tinkering Towards Accolades: School Gaming Under a Performance Accountability System* (March), mimeo, University of Michigan

Dixit, A (1999), *Incentives and Organizations in the Public Sector: An Interpretive Review*, mimeo, Princeton University

Figlio, DN (2001), What Might School Accountability Do?, http://www.nber.org/reporter/fall01/figlio.html, 07/02/02

Figlio, DN and Page, ME (2001), Can School Choice and School Accountability Successfully Coexist?, http://www.nber.org/books/schools/figlio-page8-2-01.pdf, 14/02/02

Figlio, DN and LS Getzler (2002), Accountability, Ability and Disability: Gaming the System?, http://bear.cba.ufl.edu/figlio/fig_getz.pdf (April), 14/10/02

Fitz-Gibbon, CT (1996), *Monitoring Education: Indicators, Quality and Effectiveness*, London, Cassell

Fitz-Gibbon, CT (1997), *The Value Added National Project: Final Report: Feasibility Studies for a National System of Value Added Indicators*, London, School Curriculum and Assessment Authority

Fitz-Gibbon, CT (2000), Quality, Science and Soros's Reflexivity Concept: a Value-Added Approach, in: Balázs, É, F van Wieringen and LE Watson (eds), *Quality and Educational Management: A European Issue, Budapest*, Wolters Kluwer Group: 193-204

Fitz-Gibbon, CT and P Thymms (2002), Technical and Ethical Issues in Indicator Systems: Doing Things Right and Doing Wrong Things, *Education Policy Analysis Archives*, 10(6): http://epaa.asu.edu/epaa/v10n6, 12/02/02

Gillborn, D (1996), *Exclusions from School, Viewpoint*, Number 5, Institute of Education, University of London

Glennerster, H (1991), Quasi-Markets for Education, *Economic Journal*, 101: 1268-1276

Goldstein, H (2001), Using Pupil Performance Data for Judging Schools and Teachers: Scope and Limitations, British Educational Research Journal, 27(4): 433-442

Hoxby, CM (2001), Testing is about Openness and Openness Works, http://post.economics.harvard.edu/faculty/hoxby/papers/NAEP_results_jun01.pdf http://www-hoover.stanford.edu.pubaffairs/we/current/hoxby_0701.html

Ladd, HF (1999), The Dallas School Accountability and Incentive Program: an Evaluation of its Impacts on Student Outcomes, *Economics of Education Review*, 18(1): 1-16

Ladd, HF and RP Walsh (2002), Implementing Value-Added Measures of School Effectiveness: Getting the Incentives Right, *Economics of Education Review*, 21(1): 1-17

Le Grand, J (1991), Quasi-Markets and Social Policy, *Economic Journal*, 101: 1256-1267

Meyer, RH (1997), Value-Added Indicators of School Performance: a Primer, *Economics of Education Review*, 16(3): 283-301

Saunders, L (1999), A Brief History of Educational 'Value Added': How Did We Get To Where We Are?, *School Effectiveness and School Improvement*, 10(2): 233-256

Smith, P (1995), On the Unintended Consequences of Publishing Performance Data in the Public Sector, *International Journal of Public Administration*, 18(2&3): 277-310

Thomas, S (1998), Value-Added Measures of School Effectiveness in the United Kingdom, *Prospects*, XXVIII(1): 91-108

Thomas, S and P Mortimore (1996), Comparison of Value-Added Models for Secondary School Effectiveness, *Research Papers in Education*, 11(1): 5-33

Times Educational Supplement (2002), *League Table Bonus Attracts Schools to Vocational Option*, 23 August

West, A and H Pennell (2000), Publishing School Examination Results in England: Incentives and Consequences, *Educational Studies*, 26(4): 423-436

Whitty, G, S Power and D Halpin (1998), *Devolution and Choice in Education: the School, the State and the Market*, Milton Keynes, Open University Press

Wilson, D (2001), *Information, Incentives and Insurer Behaviour: An Analysis of Selection in the Health Insurance Market*, Unpublished PhD Thesis, University of Bristol

Table 1: Summary Statistics

Type of School (% in brackets)	Community	Foundation	Voluntary Aided	Voluntary Controlled	CTC*	Total
Comprehensive	1886	362	460	80		2788 (89.10)
Modern	101	44	12	4		161 (5.15)
Grammar	45	75	32	13		165 (5.27)
Total	2032 (64.94)	481 (15.37)	504 (16.11)	97 (3.10)	15 (0.48)	3129 (inc CTCs)
Number of Pupils	343047 (66.26)	80125 (15.48)	73987 (14.29)	18172 (3.51)	2364 (0.46)	517695

*CTC: City Technology College

Gender Split: Girls: 254931 (49.24%); Boys: 262764 (50.76%)

Table 2: Key Stage 3 (1997) and GCSE (1999) Results by Gender

	Girls: Mean (s.d.)	Boys: Mean (s.d.)	All Pupils: Mean (s.d.)
KS3 English	4.99 (1.38)	4.34 (1.66)	4.66 (1.56)
KS3 Maths	4.99 (1.35)	5.06 (1.35)	5.02 (1.35)
KS3 Science	4.90 (1.16)	4.96 (1.20)	4.93 (1.18)
Total GCSE score	42.55 (17.60)	37.54 (17.74)	40.02 (17.85)

Table 3: Correlations of Performance Indicators

	VA	VAcap	%5A*-C	APS	APScap
VA	1.0000				
VAcap	0.8801	1.0000			
%5A*-C	0.3418	0.3641	1.0000		
APS	0.4884	0.4288	0.9583	1.0000	
APScap	0.3953	0.4258	0.9768	0.9812	1.0000

Table 4: Correlations of VA vs. VAcap Across Ability Deciles*

All Pupils	0.8761
1	0.9311
2	0.9154
3	0.9047
4	0.8967
5	0.8799
6	0.8674
7	0.8581
8	0.8234
9	0.7988
10	0.7568

* as defined by KS3 score (1 = lowest; 10 = highest)

Table 5a: Correlations of Rankings (Based on National Ranking 1-3129)

	RankVA	RankVAcap	Rank%5A*C	RankAPS	RankAPScap
RankVA	1.0000				
RankVAcap	0.8813	1.0000			
Rank%5A*C	0.3631	0.4067	1.0000		
RankAPS	0.5005	0.4709	0.9594	1.0000	
RankAPScap	0.4174	0.4632	0.9793	0.9816	1.0000

Table 5b: Correlations of Rankings (Based on Weighted Average of LEA Rankings)

	RankVA	RankVAcap	Rank%5A*C	RankAPS	RankAPScap
RankVA	1.0000				
RankVAcap	0.8478	1.0000			
Rank%5A*C	0.3941	0.4523	1.0000		
RankAPS	0.5463	0.5089	0.9229	1.0000	
RankAPScap	0.4511	0.5097	0.9564	0.9581	1.0000

Table 6a: School Rankings: Comparison of Alternative Performance Indicators
Bristol LEA (1999)

School	Rank%5A*C	RankVA	RankVAcap	Adm	Denom
St Mary Redcliffe and Temple	1	1	1	comp	C of E
Cotham	2	4	3	sel	none
St Bede's	3	9	7	comp	RC
St Bernadette	4	6	9	comp	RC
Fairfield	5	5	2	sel	none
Ashton Park	6	10	8	comp	none
Brislington	7	12	13	comp	none
Henbury	8	21	21	comp	none
St Thomas More	9	22	22	comp	RC
Hartcliffe	10	3	5	comp	none
Bedminster Down	11	18	18	comp	none
Whitefield Fishponds Community	12	7	6	comp	none
Monks Park	13	20	19	comp	none
Lawrence Weston	14	13	12	comp	none
Lockleaze	15	16	14	comp	none
St George Community	16	2	4	comp	none
Portway Community	17	11	16	comp	none
Speedwell	18	14	11	comp	none
Hengrove	19	8	10	comp	none
Withywood	20	17	15	comp	none
Pen Park	21	19	20	comp	none
Merrywood	22	15	17	comp	none

Table 6b: Correlations of Rankings: Bristol LEA (1999)

	RankVA	RankVAcap	Rank%5A*C
RankVA	1.0000		
RankVAcap	0.9492	1.0000	
Rank%5A*C	0.4195	0.4884	1.0000

Table 7: Subject Level Value Added (National Data)

English:

	MeanVA	VAmaxE	VAmeanE
MeanVA	1.000		
VAmaxE	0.5768	1.000	
VameanE	0.5733	0.9658	1.000

Maths:

	MeanVA	VAmaxM	VAmeanM
MeanVA	1.000		
VAmaxM	0.4475	1.000	
VAmeanM	0.4311	0.9902	1.000

Science:

	MeanVA	VAmaxS	VAmeanS
MeanVA	1.000		
VAmaxS	0.4312	1.000	
VAmeanS	0.4204	0.9890	1.000

Figure 1: Correlating Performance Indicators

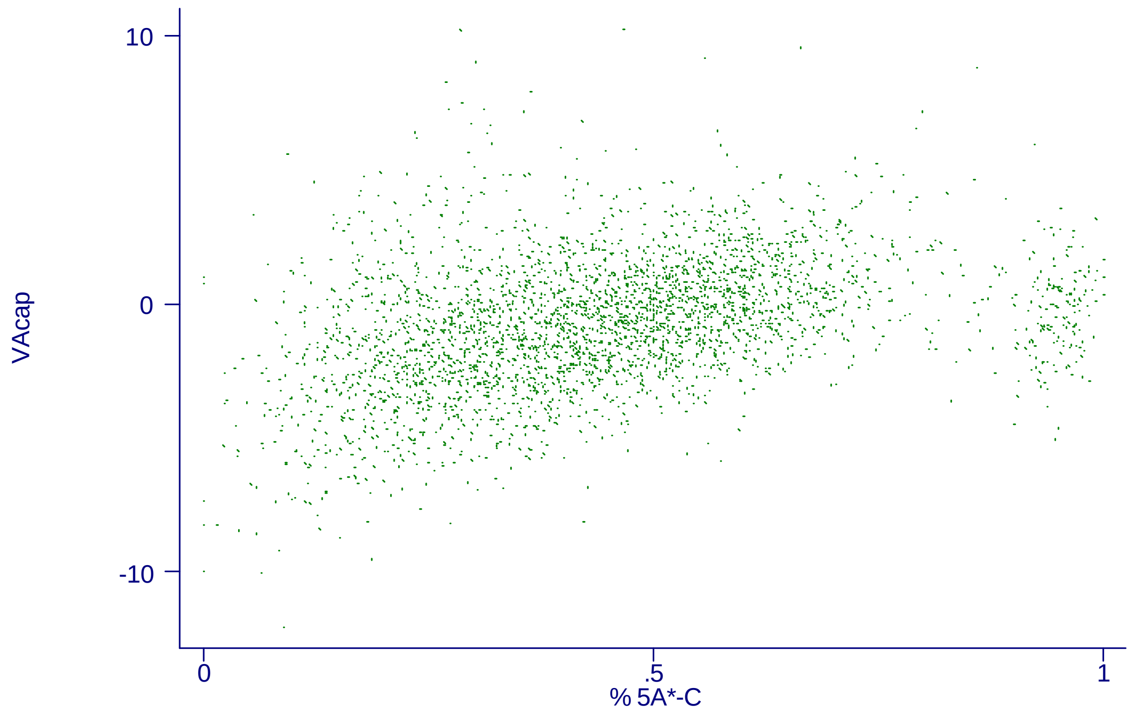

Figure 2: Correlating Performance Indicators

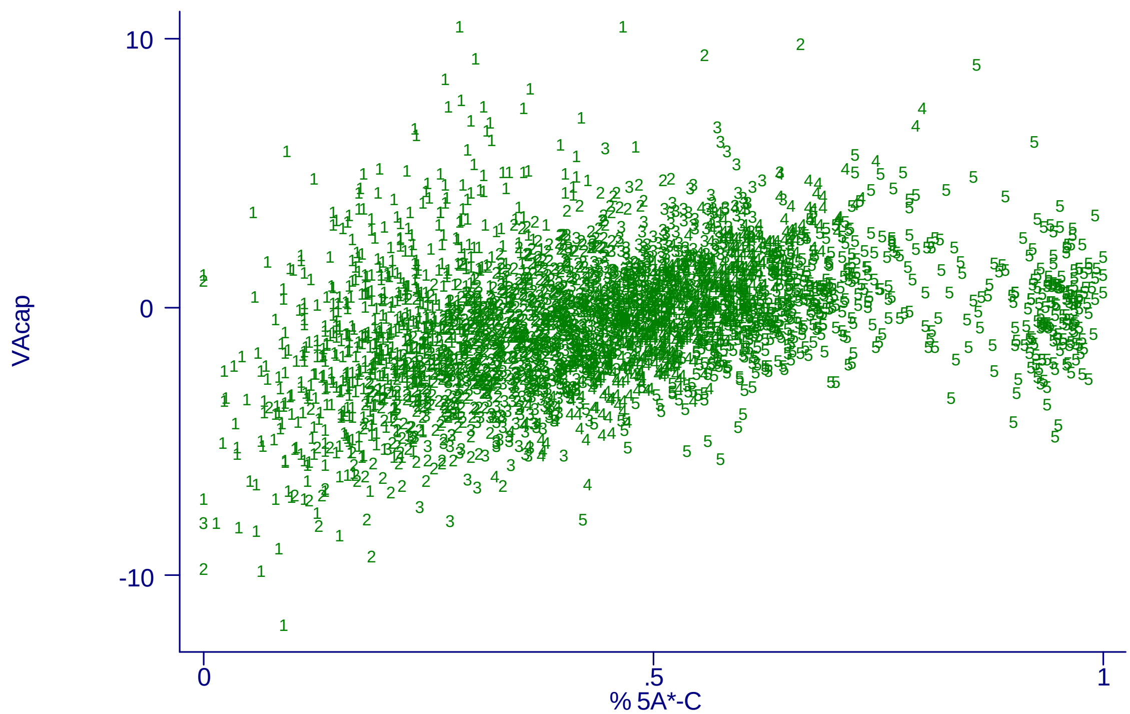

Figure 3: Correlating Rankings

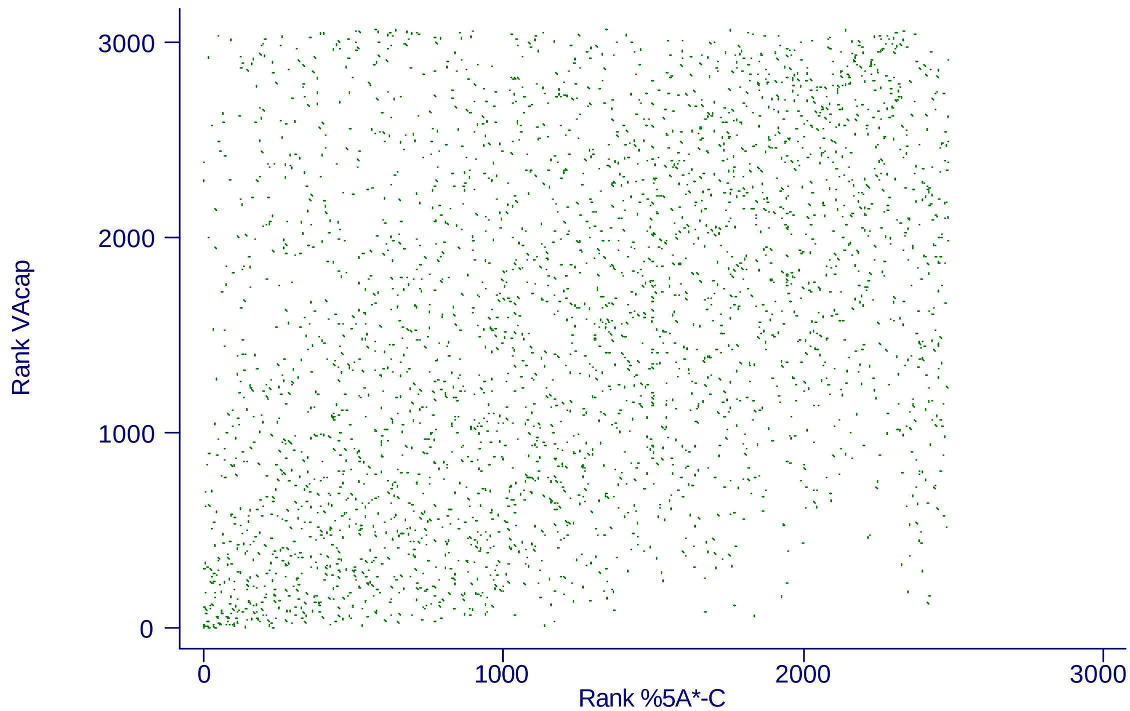

Figure 4: Correlating Rankings

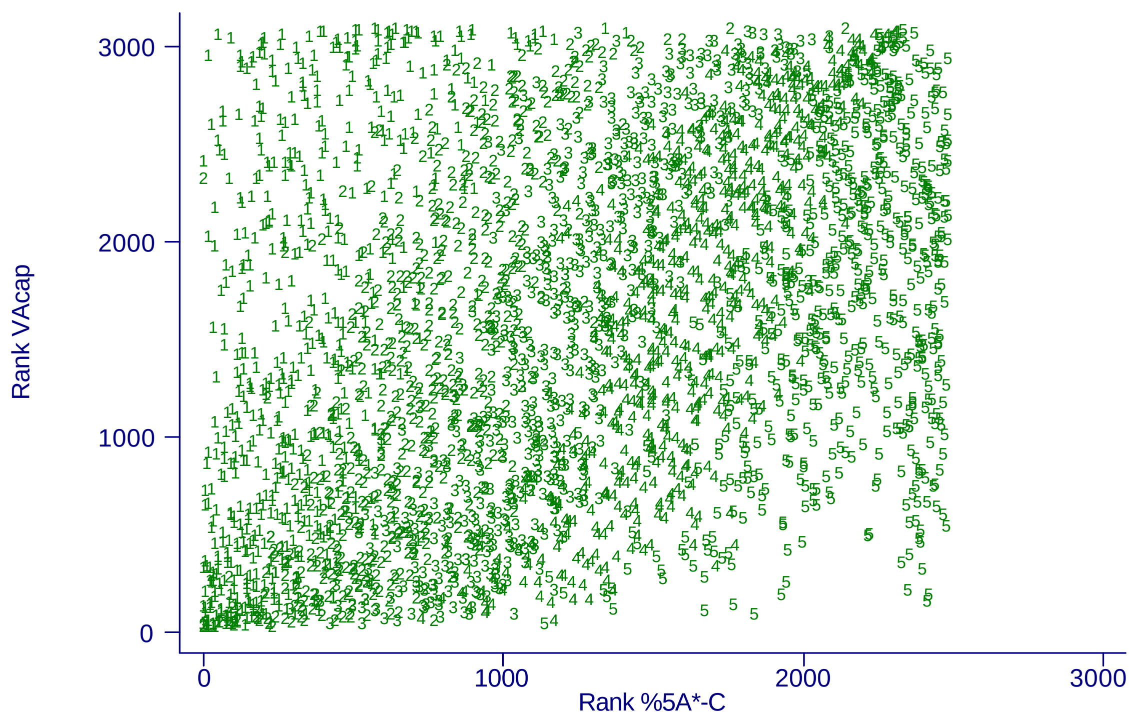

Figure 5: Pupil Performance by Subject

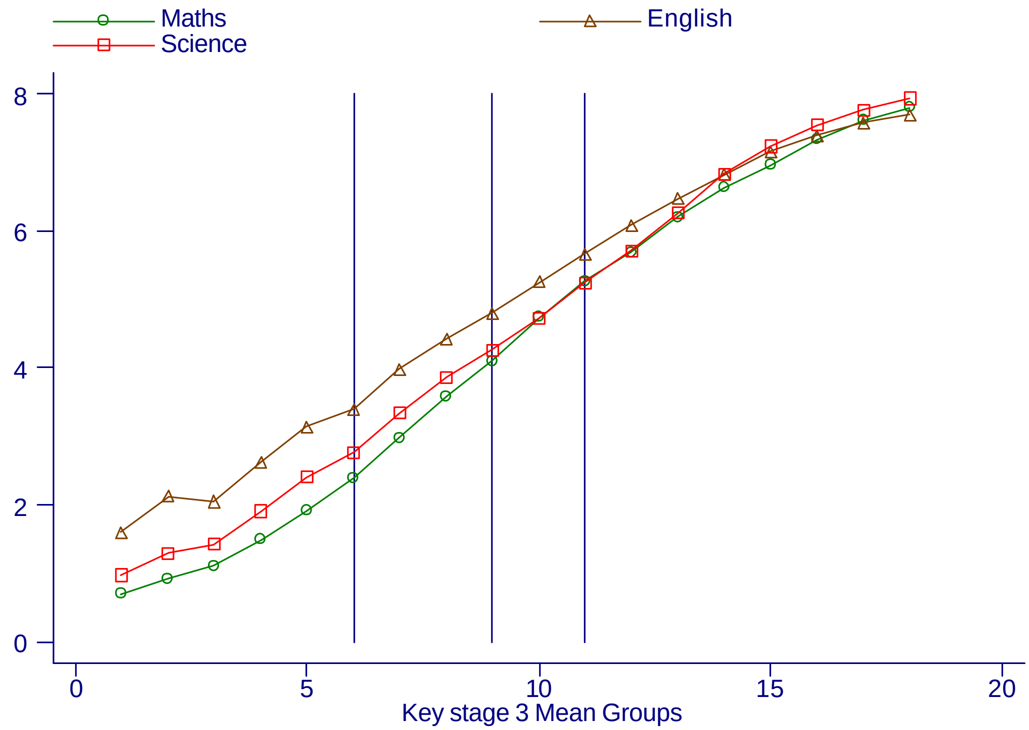

Figure 6: School Performance by Key Stage 3 Mean