

Value added in English Schools

1. Introduction

Value added indicators are now a central part of school accountability in England and value added information is routinely used in school improvement, both at national and local level. This paper describes the value added models that are being used in the academic year 2007-08 by schools¹, parents, school inspectors and other professionals in local and national government. It updates information provided for the OECD Project on the Development of Value-Added Models in Education Systems (Ray, 2006). The paper does not discuss in detail the rationale for using value added models or the wider academic debates concerning approaches to school improvement and school effectiveness and the ways these have affected education policy (see, for example, the papers in Teddlie & Reynolds, 2000)

Before discussing value added models, a few background facts are useful in order to put the English school system into an international context:

- There are approximately 8.1 million pupils in 25,000 state-maintained and independent schools (DfES 2007a) - 7% of pupils are in the independent sector. Some pupils with special educational needs are educated in maintained schools; others are educated separately in special schools².
- There are about 17,400 primary schools, which generally cover ages 4-11, and about 3,300 secondary schools, normally covering ages 11-16 (1,750 have 'sixth forms' covering post-16 as well).
- The average size of a secondary school is 980 pupils (approximately 140

¹ The focus here is on pupils of compulsory school age (age 4 to 16) and the paper does not discuss value added for older students at colleges.

² Separate value added models have been estimated for pupils in special schools

pupils per year group on average); primary schools have about 240 pupils on average, 40 per year.

- Maintained schools are funded through local government: there are 150 Local Authorities covering England. Local Authorities vary considerably in size and characteristics. The smallest is the Scilly Isles with just one school and the largest is Kent with 103 secondary schools and 466 primary schools.
- The school year runs from September to July (dates vary), divided into three terms. Assessment of attainment is made at the end of the school year, so data for the calculation of value added becomes available in the following autumn.

The development of value added models

In 1988 the National Curriculum was introduced in England, setting out the subjects and programmes of study which maintained schools are obliged to cover from ages 5-16. For Key Stages 1, 2 and 3, covering the ages 5-7, 7-11, and 11 to 14, a national system of testing and teacher assessment was established. Attainment is assessed against criterion-referenced national curriculum 'levels' at the end of each Key Stage. This means that there are currently no annual test results of each cohort: instead the national data provides the attainment of pupils at intervals of between two and four years. The testing system is run by the independent Qualifications and Curriculum Authority (QCA) and National Assessment Agency (NAA). Levels in the tests are equated between years using anchor and pre-tests. They are designed so that a Level 4 at Key Stage 2 is equivalent to a Level 4 at Key Stage 3, but the value added models do not rely on vertical equating or measure 'growth' on a common scale.

Prior to the Key Stage tests, the only examinations sat by almost all pupils in maintained schools were the GCSEs and similar qualifications at age 16 (now called

Key Stage 4). Although it was possible to benchmark or adjust these results using data about the school³, there was no information on pupil level characteristics, aside from gender, and no earlier prior attainment data with which to calculate value added scores. Nevertheless, it was possible to provide value added analysis if pupils in a group of schools took specific tests, such as in the programme developed by the CEM centre at Durham University (Tymms and Coe, 2003) and in studies like Mortimer et al (1988) and Goldstein et al (1993). A summary of these early studies is available in SCAA (1994).

The early 1990s saw a new emphasis on performance data to hold schools accountable, with the introduction of national Performance Tables showing the results obtained in tests and exams in each school. The development of the Key Stage tests offered the possibility of calculating value added scores for each school based on progress between each Key Stage, once national data was available for the relevant cohorts of pupils. On the basis of two reviews (SCAA, 1994, and Fitz-Gibbon, 1997), a simple value added method was piloted in the 1998 Performance Tables, making use of the fact that it was now possible to link Key Stage 4 outcomes to Key Stage 3 prior attainment (DfEE 1998b). This method compared each pupil's '*expected*' outcome, based on the national median GCSE result for each level of Key Stage 3 prior attainment, with their *actual* outcome. Value added scores for schools were then taken as the average of these differences for all their pupils. This method for deriving schools scores was consistent with the approach taken in presenting national charts and information on value added for the purposes of school improvement and target setting in the new Autumn Package (DfEE, 1998a).

Although there were positive responses to this 1998 pilot, the decision was taken not to publish VA scores nationally until the full secondary age range could be covered.

³ See for example Sammons et al (1994) which discusses ways of benchmarking results in the absence of national pupil level data on prior attainment.

In the first year this was possible: 2001, a further pilot of both KS 2-3 and KS 3-4 measures was undertaken, and the following year value added scores for these Key Stages were published for all schools (with a few exceptions, e.g. some private independent schools). Value added scores for *primary* schools were calculated using the same method to measure progress between KS1 and KS2; they were piloted in 2002 and published in 2003. By 2004, value added scores for secondary schools were calculated over the full 5-year span from KS2-4

Whilst the publication of value added measures in the Performance Tables was generally seen as a positive advance on the publication of 'raw results' only, there were some concerns about aspects of the methodology and presentation (discussed in Ray, 2006). At the same time, the developments in linking data, via unique pupil identifiers (introduced in 1999) and the first Pupil Level Annual Schools Census (PLASC) in 2002, which collected a range of characteristics at pupil level, offered scope to reconsider the possibility of including contextual data alongside prior attainment in value added models.

Once PLASC data matched to information on pupil progress became available towards the end of 2002, statisticians in the Department began analysing it to understand the relationships between the variables and what they said about national performance. Views of a selection of academics in the field were sought on the future direction of the value added work and, although there was no consensus of opinion, there was strong support from some for the development of more complex models that used the new data. Outside the Department, statisticians had begun building value added models that took into account contextual factors and external advice from the National Audit Office (2003) was to develop performance information for schools that would take into account not just prior attainment, but also 'other external influences on performance', based on the new PLASC data.

In October 2004 a prototype 'contextualised' value added (CVA) model, so called because it now added 'contextual factors' such as gender and ethnicity into the model in addition to prior attainment, was discussed with secondary schools. The following year a system of CVA scores was piloted for use in Performance Tables and in 2006 a similar model was rolled out to all schools in the KS4 tables. Piloting followed a year later for other Key Stages, ensuring that measures for all Key Stages were rolled out by 2007. At the same time, the contextualised value added model was introduced into the new software package RAISEonline⁴, which was developed for school improvement and inspection.

The contextual value added models

As noted already, the current models use both prior attainment and 'contextual' factors, and have become known as CVA models to distinguish them from the earlier VA models that used prior attainment only. Note that it would also be possible to produce 'contextualised attainment' models where prior attainment information is missing, as is currently the case in England for the assessments made of seven-year olds at the end of Key Stage 1. The approach taken in Dallas (Webster and Mendro, 1997) is to run a first stage in which attainment is adjusted for contextual factors and then construct value added estimates from the resulting residuals. In England, a simpler one stage model has been used.

Most of the models cover progress across the individual Key Stages: 2, 3 and 4 (periods of four, three and two years respectively). There are also models for progress between ages 11 to 16, i.e. Key Stages 3 and 4 combined: the five year period covered by most secondary schools. Models have been designed for individual subject outcomes and for average outcomes across the range of subjects tested. Value added models based on growth over time normally measure

⁴ This can be seen at <https://www.raiseonline.org/>

progression in a specific subject, whereas the CVA models use prior attainment for all the core subjects for which we have data, so as to improve the fit of the model. It would be relatively straightforward to calculate models that use English and mathematics only as inputs and outputs, but data is limited for other subjects.

Specification of the multilevel model

The example described here is the five-year model for progress made across all subjects for which we have test data by pupils who completed Key Stage 4 in 2007. The simplest value added regression model (1) for this, with no contextual factors, would include as the $y_{ij(t_2)}$ Key Stage 4 results for the i th pupil in the j th school, and as the explanatory variable $y_{ij(t_1)}$, the Key Stage 2 prior attainment average point scores they achieved five years earlier. The a and b are regression coefficients and the e_{ij} and u_j are the random quantities, which are independent, normally distributed deviations with means equal to zero and common variances. The model coefficients could be estimated using Ordinary Least Squares and the mean of the pupil residuals for each school used as a measure of value added.

$$(1) \quad y_{ij(t_2)} = a + by_{ij(t_1)} + u_j + e_{ij}$$

In the contextualised value added model for maintained mainstream schools we actually use four variables to model prior attainment (as described below), a total of 10 pupil level contextual variables and two school level variables. This is shown in (2), where, for example, x_{5ij} denotes the first of the pupil contextual factors and y_{qj} the first of the two school level factors. (Note that the explanatory variables are measured in 2007 and so could also be given the Key Stage 4 subscript, although most would be equally applicable to the earlier time point, e.g. gender, ethnic

classification).

$$(2) y_{ij(t_2)} = \beta_0 + \beta_1 y_{1ij(t_1)} + \dots + \beta_4 y_{4ij(t_1)} + \beta_5 x_{5ij} + \dots + \beta_p x_{pij} + \beta_q y_{qj(t_1)} + \beta_s y_{sj(t_1)} + \varepsilon_{ij}$$

Rather than use an Ordinary Least Squares approach, the CVA indicators are derived from a multilevel model (Goldstein, 2003), where the residual variance is partitioned between Level 1, e_{ij} , for each pupil, and Level 2, u_j , for each school (3).

This multilevel model was run in MLwiN, a software package that estimates the fixed effects - prior attainment and contextual factors - and calculates Level 1 and Level 2 residuals⁵. Level 1 residuals show variation in pupils' outcomes in relation to their schools. The Level 2 residuals show schools' outcomes in relation to the national expected results, given the factors measured by the fixed effects. These Level 2 residuals are the value added scores.

$$(3) \varepsilon_{ij} = u_j + e_{ij}$$

The choice of variables

The dependent variable being modelled is the capped total point score achieved at KS4, based on each pupil's grades in their best eight qualifications. This gives a continuous measure of attainment ranging from the most able pupils scoring eight or more A* grades, right through to pupils achieving smaller numbers of G grades, and includes lower level qualifications (such as in basic literacy and numeracy) taken mainly by pupils with special needs.

The point score outcome measure was chosen rather than simple thresholds (such as whether a pupil has achieved five or more A*-C grades) so that the resulting model reflects under- or over-performance of pupils at all parts of the range rather than focusing only on pupils at the C/D grade borderline. It also gives useful

⁵ For more on MLwiN see <http://www.mlwin.com/features/index.html>.

diagnostics at pupil level, showing what each pupil has achieved in comparison to “similar” pupils nationally.

Capping the point score at the best eight qualifications aims to capture both the quantity and quality of subject grades, without providing incentives for schools to enter pupils for excessive numbers of qualifications.

The most important explanatory variable is each pupil's prior attainment and this has been modelled as accurately as possible. Average point score (APS), based on the marks achieved in English, Maths and Science, is used, alongside two extra terms which measure the *difference* between the English, maths and science results. This is mathematically equivalent to including each of the subjects separately but has advantages in coping with missing data as well as simplifying presentationally. A quadratic APS term is also included, reflecting the fact that the relationship between KS4 outcomes and KS2 prior attainment is non-linear. We also make a post-hoc adjustment for floor and ceiling effects to deal with a very small group of pupils at the extremes of the range where the model predictions are out of line with the achievements recordable by the testing scale. Making the ceiling adjustments is important as, although only 2% of pupils are in this range, their treatment has a disproportionate effect on the small number of selective “grammar” schools which exist in certain areas. Further information is available on the DCSF Achievement and Attainment Tables website: <http://www.dcsf.gov.uk/performance/tables/>

Comment: Should we add a footnote to mention the ‘floor’ effect as well and perhaps quantify these: the ceiling effect affects approximately 2% of pupils in KS2-4 and the floor affects less than 1% of pupils? If we were to say this, should we acknowledge the fact that the former group of pupils are disproportionately distributed in grammar schools?

The decisions over which contextual variables to include in the model were based on a mixture of statistical, educational and practical considerations. Given the need to provide value added information for every school, it was necessary to restrict the choice to information for which there is national data. Since the aim was to generate value added residuals, the explanatory fixed effect variables needed to cover factors that are outside the school's control. The choice of additional pupil level contextual

variables took into account what was known from previous internal and external research about factors that explain variation in test results.

Table 1 provides the full list of explanatory variables and their coefficients. Most of the pupil-level variables are simple flags derived from school census data and

Table 1 shows how many pupils were coded '1' in the 2007 model. The majority of these categories apply to thousands of pupils nationally although, of course, many individual *schools* have no pupils from some of these groups.

Table 1 The 2007 Key Stage 2-4 CVA regression model

Dependant variable = capped KS4 point score
 -2logL = 6304400
 Number of pupils = 563,908

Explanatory factor	Variable	Estimate	Std. Error
	Intercept	162.10	13.10
Prior attainment	KS2 student APS	-5.94	0.24
	KS2 APS (using fine grades) – squared	0.38	0.01
	KS2 English PS deviation	1.40	0.07
	KS2 Maths PS deviation	-0.11	0.07
Deprivation	Does student have FSM?	-22.90	0.33
Deprivation of pupil's local area	Deprivation indicator – IDAC1 score	-59.51	0.67
In care	Has the student ever been in care at this school?	-27.10	1.17
Special Educational Needs (SEN)	Does student have SEN Statement/ Action Plus?	-65.76	0.38
	Does student have SEN - school action?	-34.37	0.31
Mobility	Student joined other than Jul/Aug/Sep?	-23.43	0.42
	Student joined within last 2 yrs?	-73.55	0.63
Gender	Is student female?	14.52	0.19
Age	Age within year	-12.94	0.30
English as Additional Language (EAL)	Is English not the student's first language?	-8.32	8.25
	EAL interacted with prior attainment	4.93	0.68
	EAL interacted with prior attainment squared	-0.14	0.01
Ethnic group	White Irish	-3.61	1.56
	White Irish Traveller	-64.92	7.66

	White Gypsy/Roma	-54.15	5.25
	White Other	10.52	0.79
	Mixed White/Black Caribbean	-3.08	1.03
	Mixed White/Black African	8.03	2.19
	Mixed White/Asian	10.81	1.38
	Any other Mixed ethnic group	6.03	1.08
	Indian	24.08	0.83
	Pakistani	17.53	0.92
	Bangladeshi	22.91	1.47
	Any other Asian ethnic group	24.76	1.35
	Black Caribbean	13.31	0.90
	Black African	29.19	1.04
	Any other Black ethnic group	11.16	1.62
	Chinese	32.78	1.77
	Any other ethnic group	20.68	1.39
	Unclassified ethnic group	-7.42	0.74
	White Irish and FSM	2.56	4.04
	White Irish traveller and FSM	11.36	11.98
	White Gypsy/Roma and FSM	25.62	7.79
	White other and FSM	26.53	1.79
	Mixed White/Black Caribbean and FSM	8.46	2.08
	Mixed White/Black African and FSM	3.87	4.46
	Mixed White/Asian and FSM	8.27	3.54
	Any other Mixed ethnic group and FSM	11.24	2.48
Ethnic group * FSM	Indian and FSM	16.04	1.94
	Pakistani and FSM	17.30	1.24
	Bangladeshi and FSM	19.48	1.89
	Any other Asian ethnic group and FSM	24.71	2.87
	Black Caribbean and FSM	18.20	1.79
	Black African and FSM	20.46	1.65
	Any other Black ethnic group and FSM	8.69	3.19
	Chinese and FSM	34.11	5.12
	Any other ethnic group and FSM	30.83	2.32
	Unclassified ethnic group and FSM	7.14	1.95
Level of school prior attainment	School KS3 APS (using fine grades) for CVA	1.96	0.36
Spread of school prior attainment	School std dev of KS3 APS for CVA	-4.82	0.92
	Random components:	Estimate	Std. Error
	Between school variance	319.42	8.77
	Within school variance	4135.60	7.81
	Variance partition coefficient	0.08	

Table 2 The prevalence of different pupil characteristics

Binary variables in the KS2-4 2007 CVA model		
	Number	Percentage
Does student have FSM?	68760	12.2%

Has the student ever been in care at this school?	3091	0.5%
Does student have SEN - school action?	60773	10.8%
Does student have SEN - Action Plus?	42411	7.5%
pupil joined school after Sept Yr 10	10750	1.9%
pupil joined school not in July /AUG/ Sept Yr 7- 9	25841	4.6%
Is student female?	279701	49.6%
Is English not the student's first language?	43321	7.7%
Is the student White Irish?	2079	0.4%
Is the student a White Irish traveller?	120	0.0%
Is the student White Gypsy/Roma?	278	0.0%
Is the student White other?	9557	1.7%
Is the student Mixed White/Black Caribbean?	5381	1.0%
Is the student Mixed White/Black African?	1154	0.2%
Is the student Mixed White/Asian?	2603	0.5%
Is the student any other Mixed ethnic group?	4536	0.8%
Is the student Indian?	11984	2.1%
Is the student Pakistani?	13035	2.3%
Is the student Bangladeshi?	5096	0.9%
Is the student any other Asian ethnic group?	3231	0.6%
Is the student Black Caribbean?	7541	1.3%
Is the student Black African?	7171	1.3%
Is the student any other Black ethnic group?	2204	0.4%
Is the student Chinese?	1619	0.3%
Is the student any other ethnic group?	3611	0.6%
Is the student in an unclassified ethnic group?	10622	1.9%
Is the student White Irish & FSM?	302	0.1%
Is the student a White Irish traveller & FSM?	49	0.0%
Is the student White Gypsy/Roma & FSM?	126	0.0%
Is the student White other & FSM?	1702	0.3%
Is the student Mixed White/Black Caribbean & FSM?	1309	0.2%
Is the student Mixed White/Black African & FSM?	277	0.0%
Is the student Mixed White/Asian & FSM?	395	0.1%
Is the student any other Mixed ethnic group & FSM?	848	0.2%
Is the student Indian & FSM?	1301	0.2%
Is the student Pakistani & FSM?	4507	0.8%
Is the student Bangladeshi & FSM?	2697	0.5%
Is the student any other Asian ethnic group & FSM?	646	0.1%
Is the student Black Caribbean & FSM?	1775	0.3%
Is the student Black African & FSM?	2509	0.4%
Is the student any other Black ethnic group & FSM?	552	0.1%
Is the student Chinese & FSM?	180	0.0%
Is the student any other ethnic group & FSM?	1203	0.2%
Is the student in an unclassified ethnic group?	1300	0.2%

No data on social class, family income or parental education levels is collected nationally for all pupils. However, some information relating to deprivation is available. Children whose parents receive the social welfare benefit Income Support,

and some related benefits, are entitled to claim free school meals (FSM). To become entitled to FSM the parents have to indicate a wish for their child to have a school meal and give a proof of benefit receipt. Even as a simple proxy measure for 'deprivation', FSM has disadvantages (Ray, 2006). For example, the fact that parents have to register an interest in their pupils' having school meals may discourage some from applying. Those not applying but eligible may be an unrepresentative group, e.g. choosing not to register for cultural reasons or on the basis of dietary preferences that may correlate with attitudes to education. The benefit rules may also exclude some families who could be considered deprived, e.g. some low paid workers.

In addition to FSM, the models use another measure of deprivation that takes advantage of the fact that the School Census collects pupils' home postcodes and that these can be linked to data on their local area. Various possible local indicators have been investigated but at present the models use a measure called IDACI: the Income Deprivation Affecting Children Index. This is calculated at small local area level: England is split into 33,500 areas with approximately 600 households in each and the percentage of the local area's children who are under 16 and are living in families in receipt of certain benefits, or on low incomes, is then calculated. The index is calculated by another government department using data and a methodology which can be checked and is open to scrutiny. This makes it more suitable for an official value added system than the various commercial indices such as ACORN or MOSAIC, which are widely used to classify small areas by their socio-economic characteristics.

Special Educational Needs (SEN) covers a wide range of needs that are often inter-related as well as specific needs that usually relate to particular types of impairment. Children with SEN will have needs and requirements which may fall into at least of

one of four areas: communication and interaction; cognition and learning; behaviour, emotional and social development; and sensory and / or physical needs. The model distinguishes between two levels of SEN: School Action (where the class teacher or SEN Co-ordinator provides additional help) and School Action Plus, where the intervention at School Action has not resulted in improvement and external advice is sought.

Pupils who switch between schools tend to make less progress and the CVA model takes this into account using data from the School Census on the date of entry into the school for each pupil. The current method for Key Stage 2-4 is to flag pupils who joined at non-standard times - any month other than July, August or September – and also pupils who joined within the last two years.⁶ Recent research has considered whether late joining pupils in ‘high-mobility’ schools experience the same disadvantage as other late joining pupils (Wilson et al., 2008). They found that late joiners in ‘average’ schools are more disadvantaged than late joiners in ‘high-mobility’ schools. This area warrants further investigation.

Pupils’ ages vary when they take the tests used in the value added model, although they are normally within a year of each other. On average, younger pupils tend to have lower outcomes but make more progress. Age within year is, therefore, included as one of the variables. Another variable associated with lower outcomes but faster progress due in part to ‘catching up’ is the first language of pupils – English or ‘other than English’. Since 2006, interaction terms between having English as an additional language (EAL) and prior attainment have been included in the model; this is to account for the fact that EAL pupils with low prior attainment make, on average, better progress than non-EAL pupils, but this gap tapers away for pupils with higher prior attainment. Difficulty with language limits attainment in the early phases of

⁶ This will create an anomaly in the case of newly-opened schools if they admit large numbers of pupils at non-standard times

education yet, pupils who subsequently master English regain much of the lost ground. For consistency, an interaction term between EAL and prior attainment squared is also included within the model in order that a curve of the same form be fitted for both EAL and non-EAL pupils.

The school census collects data for 18 main ethnic groups, with a 19th code available for 'unclassified' since provision of this data is voluntary. The unclassified group represented 1.9% of the pupils included in the 2007 Key Stage 2-4 contextualised value added model (having fallen from 4.9% in the earliest years of the PLASC collection). This unclassified group is not nationally representative – pupils in it tend to have relatively low attainment. Even with 18 ethnic groups, the codes obviously have to cover pupils with very different characteristics. For example, the Black African category covers pupils who may or may not speak English, or who may come from recent or long established immigrant communities. One non-significant ethnic category was included because it was felt that for practical and presentational reasons it would be better to include *all* the categories rather than combine two of them with other groups on the basis of the one year's data (including these variables makes very little difference to the overall model.) There are interaction terms in the multi-level model between each of the ethnic groups and FSM eligibility. These terms reflect the extent to which differential effects of poverty vary across different ethnic communities in their impact on progression.

The final pupil level contextual variable flags those who are in the care of their Local Authority. These children may be living with foster parents or prospective adopters, placed in children's homes or some other form of residential care, or placed at home with their parents. It is a relatively small proportion of pupils, but an important group who are relatively likely to be vulnerable and educationally disadvantaged. There have been some concerns that the numbers of these pupils are under-counted in

annual school census, partly because schools may not know whether pupils are in the care of their authority (which may not be the same authority in which they go to school). In 2008 the plan is to try to merge data from the Local Authority directly onto PLASC, rather than rely on schools to supply this information.

There are advantages and disadvantages in including school level contextual variables. For example, they can control for additional issues that affect all pupils in the school, but they may also be partly endogenous, measuring factors that have themselves been influenced by the school's effectiveness. They put schools on a more equal footing so that the value added estimates can provide a fairer indicator of school effectiveness, but in doing so make an adjustment that may not be useful to parents or others who are less interested in making comparisons on this basis. The CVA models are used for a variety of purposes and the decision was taken to include two school level factors in the models for secondary schools, describing the overall prior attainment of the intake, but not additional factors like the school's ethnic mix, level of deprivation and so forth. These variables were the overall level and spread (as measured by the standard deviation) of APS prior attainment.

The model for Primary Schools

The model used for Primary Schools is broadly similar to the KS2-4 model outlined above. The main difference is that the school level terms are omitted: this is because the pattern of intakes is less variable than for secondary schools, and also because with the small numbers of pupils in each primary cohort, the use of the same prior attainment variable at two levels can lead to model instability.

There are also other minor differences such as in the mobility factor definitions.

The effect of shrinkage on the value added estimates

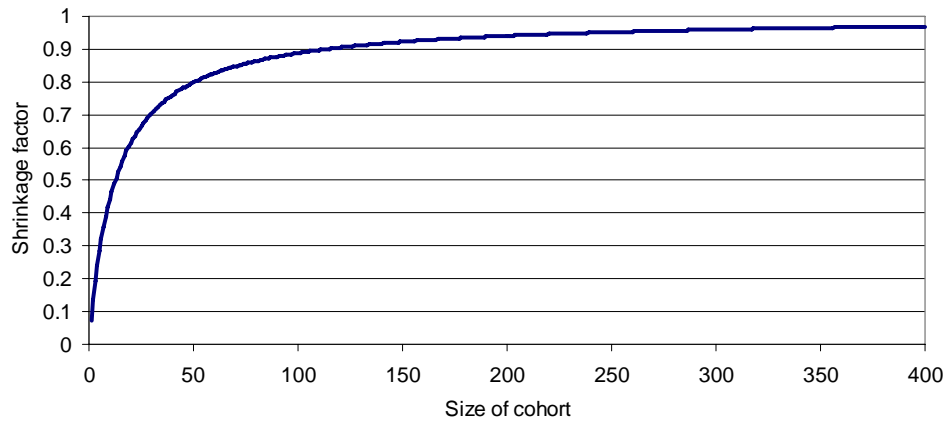
Multilevel modelling (MLM), in taking account of the structuring of the educational data, offers a more complex set of modelling options than Ordinary Least Squares (OLS). The added complexity is sometimes seen as a disadvantage, particularly when the models that result are fairly similar, both in terms of estimated fixed effects and value added residuals. In England both OLS and MLM versions of the models have been calculated and compared (Ray, 2006). The main practical difference is that the MLM value added estimates incorporate 'shrinkage'.

MLM uses an iterative process to produce maximum likelihood estimators for both the regression coefficients and the variance of the residuals at school and pupil level. To do this, it can be shown (Goldstein, 2003) that the estimated school level residuals \hat{u}_{0j} are equal to the raw residuals that would be obtained from the model's estimated fixed effects, adjusted by a shrinkage factor, c_j . The constant c_j is called the shrinkage factor because it is bounded by 0 and 1, so that the \hat{u}_{0j} residuals will be smaller (either negative or positive) than the raw residuals. As the size of the school increases, c_j becomes closer to 1. In other words, for large schools, the MLM and OLS value added residuals will be similar. For small schools on the other hand, the shrinkage factor will have an impact. The impact is larger when the within-school variation is large relative to the between-school variation.

The degree of shrinkage depends on the size of the school: smaller schools are 'shrunk' towards the national mean. Figure 1, based on the shrinkage factor for the KS2-4 CVA model, shows how the scores of schools with fewer pupils in their cohort are brought closer to the national average (zero). A school's 'raw score' is particularly affected by shrinkage for cohorts of less than about 50. However, although the main impact is seen for small schools, some shrinkage occurs in even

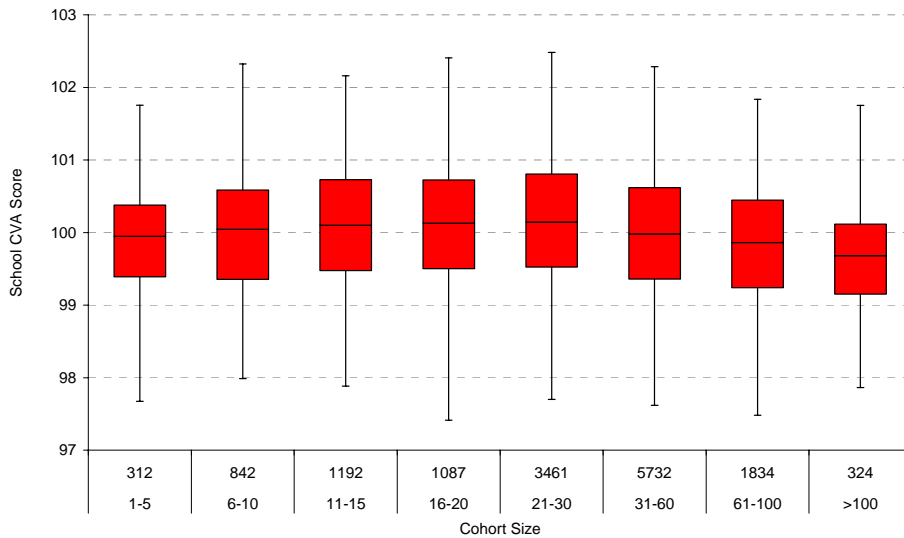
quite large schools. Earlier work for the OECD (Ray, 2006) showed the extent to which shrinkage affects the value added estimates. A shrinkage factor of, for example, 0.9 means that the value-added score is reduced to 90% of its 'raw' size: moving slightly closer to the national average.

Figure 1 The effect of shrinkage on Key Stage 2-4 CVA



Only 14% of secondary schools had a shrinkage factor below 0.9 and, among primary schools, 86% had a CVA score that was at least three quarters the size of the raw residual. Without shrinkage, the smallest primary schools have the best and worst value added scores, which in many cases is likely to be a misleading indication of their effectiveness. After shrinkage, as shown in Figure 2, the actual MLM-derived estimates have a similar spread across the range of schools.

Figure 2 Key Stage 1-2 school CVA scores by cohort size



In OLS, a small school's value added is based on the pupils it has, with a confidence interval implying a wide margin of error. The process of shrinkage in multilevel modelling provides a different approach, where the point estimate is calculated from both the school's pupils *and* the national school distribution (with narrower confidence intervals around the resulting estimates). This could be seen as problematic if it prevents a genuinely successful or ineffective small school from registering a significantly high or low value added score. However, in such circumstances, there is no way of telling from one year's data whether these small schools' raw residuals *are* good estimates of effectiveness, or whether their results have been subject to random error of some kind. The advantage of MLM is therefore that it explicitly allows for this uncertainty in calculating the value added score.

There is no easy solution to the problem of interpreting value added for small schools. Restricting value added to schools above a certain size means that school improvement and accountability systems leave some schools out. The Department's current policy is to calculate scores for all schools but only publish a school's value-added if they have more than 10 pupils in their cohort and greater than 50 percent

“coverage” (i.e. the proportion of the cohort for which value added can be calculated). Averaging over more than one year may provide more robust figures but prevents annual comparisons. Of course, interpreting significant changes in value added over short time periods is difficult even for larger schools, which is why this kind of information is used as a basis for discussion in school improvement and inspections, rather than directly driving any rewards or sanctions. Various studies have shown the way that value added scores fluctuate or change over time: see for example DfES (2004), Taylor, Peng and Gray (2007), Wilson and Piebalga (2008). The degree of volatility depends partly on the way value added scores are used (obviously it is possible to move around considerably in terms of ranks in the middle of a distribution). The CVA estimates are given confidence intervals and year-on-year changes in RAISEonline are tested for significance. Bearing this in mind, stability in value added may be considered reasonable whilst, over a longer time period, allowing sufficient change to indicate real changes in school effectiveness. However, further research on the extent to which trends in value added correlate with other indicators of school effectiveness would be welcome.

Judging the complexity of the model

Moving away from the use of unadjusted test results, simple growth scores or value added measures to a regression based approach risks confusing or alienating users. It needs to be born in mind that whilst OLS is widely familiar to those with statistical or econometric training, it is not familiar to most teachers or parents. The multilevel model set out here is one of the simplest possible and the Department has provided guidance on how it works both online and through presentations to schools and Local Authorities. This has explained and justified the effects of each variable in the model and explained how the residual is calculated and ‘shrunk’. A ‘ready reckoner’ is available that allows users to see how the model treats pupils with different

characteristics⁷.

The choice of variables, discussed above, focused on those factors that add something significant in statistical or educational terms to the model and improve the value added estimates for certain schools. The first version of the model excluded the two sets of interaction terms: it was only after further consultation with schools that these were included. As schools become used to a model it may be possible to add technical sophistication, although any changes will create discontinuities in the indicators. Further possibilities, for example, are the use of more than one year's prior attainment, considered in Goldstein and Sammons (1997), or a more complex approach to mobile pupils, where their progress is assigned to different schools, described in Goldstein, Burgess and McConnell (2007).

The model is restricted to two levels, rather than taking into account a third level above the school for Local Authorities, which fund and provide services to the maintained schools in their area. Another way in which this model has been kept simple is that there are no explanatory variables for the random part of the model. It assumes that a school is uniformly more or less effective for all its pupils, and that this can be encapsulated in a number, the value added score. A more complex approach is to assume schools vary in their effectiveness, e.g. between levels of prior attainment or for different ethnic groups. This could produce a range of measures for each school, or alternatively a set of charts showing the schools' value added compared to national expected levels. Although there are significant differences in, for example, the slope of attainment outcomes relative to prior attainment, this has not been taken into account for value added in Performance Tables, where the aim was to provide a single indicator. Value added information for school improvement in RAISEonline is differentiated within schools according to

⁷ See <http://www.standards.dfes.gov.uk/performance/1316367/CVAinPAT2005/>

whether pupils have low, medium or high prior attainment.

Data completeness and model estimation

To maximise the usefulness of the CVA model, it attempts to calculate scores for as many pupils as possible. Where individual data items are missing this leads us to use default values in preference to omitting pupils. The minimum data requirement is to have at least one item of prior attainment, and a valid outcome. For the prior attainment variables, the APS is taken as the average of whichever subject point scores are available, with any deviations which cannot be calculated defaulted to zero. For binary characteristics the defaults assumed are the most common (e.g. non-FSM, non SEN) and, for continuous measures, the national means are used. The main exception is the IDACI deprivation variable, where the mean value for the school is used.

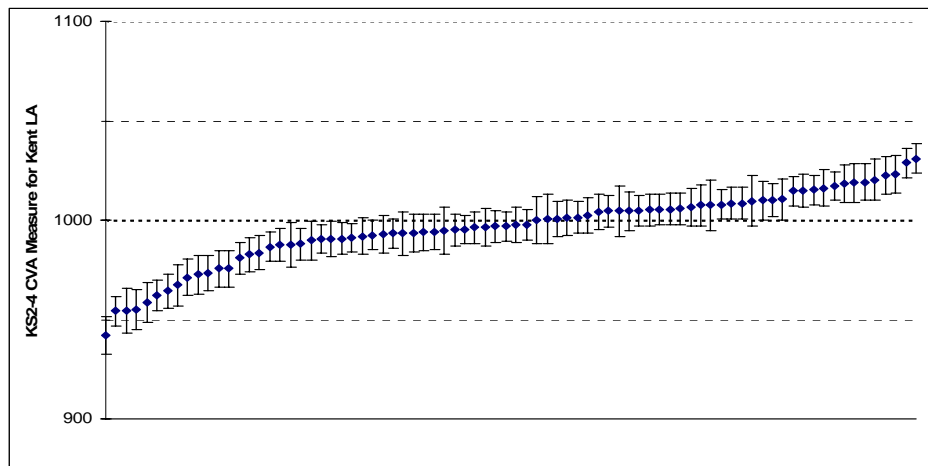
To avoid these defaults distorting the model estimation process, pupils whose entire PLASC characteristics are missing are not included in the MlwiN model estimation.

Use of value added in school and college achievement and attainment tables

The objectives of the annual performance tables are to provide consistent accessible national data on the performance of schools, to inform parents and the public more generally, and ensure that schools are accountable for their results. The tables are resource intensive to produce accurately every year and are deliberately kept to a limited range of key indicators. For CVA the tables show the single overall measure, together with upper and lower confidence intervals, and the coverage indicator (the percentage of the cohort included in the CVA measure). They do not, for example, provide results or value added for every subject taken at Key Stage 4. Users are directed to Ofsted inspection reports for a fuller picture of a given school.

One graphical representation that is frequently used to display schools' effectiveness is referred to as a 'caterpillar plot'; this is a graph which plots each school's CVA point estimate, along with error bars representing upper and lower confidence bounds. Figure 4 provides an example of this for schools in one Local Authority.

Figure 4 Key Stage 2-4 CVA scores for schools in one Local Authority



The chart shows the overall CVA measure for all schools in this LA, relative to the national mean of 1000. Where the confidence interval does not cross the national average line, the school value differs significantly from the national average. It would be possible to produce this plot for all maintained mainstream schools in England, but with several thousand schools the error bars are not easily legible so we more frequently use a "snake" plot showing only the point estimates. It is then possible to draw lines showing bands for where the results of the top 5%, 25% and bottom 5% and 25% of schools lie, for example.

Users are told that value added measures represent a better estimate of school effectiveness than the raw results that take no account of prior attainment. However, the Department recognises that test results and no single measure of performance can not tell the whole story about a school's effectiveness. This is why the guidance

encourages parents not to refer to CVA, or indeed the attainment and achievement tables, in isolation but also to seek out Ofsted reports, school prospectuses etc and, above all else, to visit the school before making choices.

Use of value added for school improvement and inspection

Schools and Local Authorities use a variety of data sources for school improvement, from simple Excel spreadsheets to the analyses of specialists in value added like the Fischer Family Trust (Kirkup et al, 2005). However, the Department and Ofsted now provide all schools with a free online system that allows them to analyse their own data and compare it against national patterns or the results and value added achieved by high performing schools. It provides a more extensive range of data than the Performance Tables, including value added for a wider range of outcome measures and for subgroups of pupils within the school.

RAISEonline aims to provide a common set of analyses for schools, Local Authorities, inspectors and School Improvement Partners. Schools use RAISEonline as part of the self-evaluation and target setting process that they undertake with the help of School Improvement Partners. The data is also available to Ofsted's inspectors for use in assessing the extent to which the school is improving or has the capacity to improve before making an inspection visit.

Features of RAISEonline include:

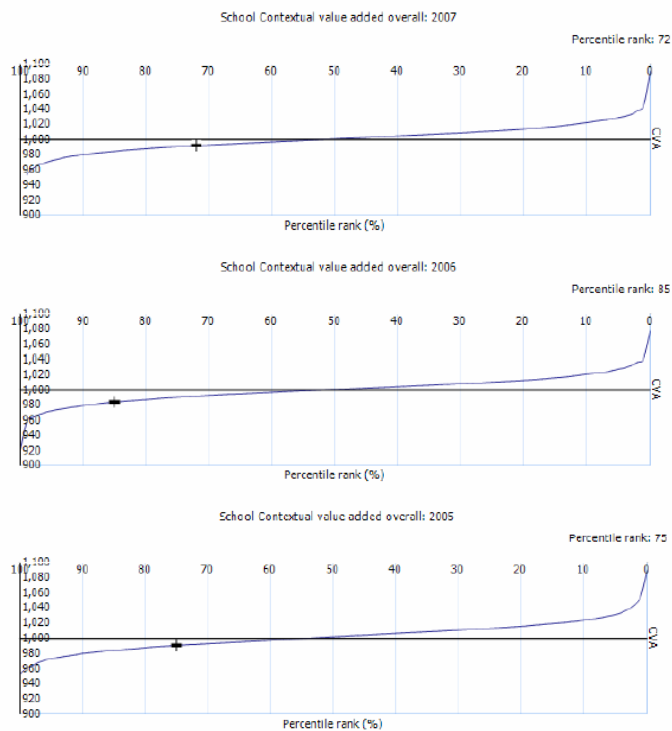
- The ability to produce reports and analysis covering the attainment and progress of pupils in Key Stage 1, 2, 3 and 4, with interactive features to allow exploration of pupil performance;
- Contextual information about the school, including comparisons to schools nationally;
- Question-level analyses, allowing schools to investigate the performance

of pupils in specific curriculum areas;

- Target-setting to support schools in the process of monitoring, challenging and supporting pupil performance;
- A data management facility providing the ability to import and edit pupil-level data and create school-defined fields and teaching groups.

Figure 3 is taken from a RAISE report and shows the overall CVA measure for an example school relative to the national mean of 1000. The school is placed within the national distribution in order to illustrate the range of CVA scores attained by other maintained mainstream schools and a 95% confidence interval is shown for the school in question. Where this does not cross the national average line, the school value differs significantly from the national average.

Figure 3 Snake plots to show school CVA for Key Stage 2 to 4



The chart shows how the school's percentile rank has changed over 3 years: 2005-2007. Although the school has a CVA which is significantly below the national average in 2006, the school's CVA in 2007 and 2005 cannot be said to differ significantly from the national average.

Other uses of value added indicators

Value added is not used directly in funding schools or teachers. However, there is an indirect link with pay since many individual classroom teachers have progressed to a higher pay scale through an accreditation process. For this, they were required to provide evidence that as a result of their teaching, "their pupils achieve well relative to the pupils' prior attainment, making progress as good or better than similar pupils nationally".

Value added has been used within policy initiatives as a way of selecting out particular schools for extra attention. For example, under-performing schools have been given additional consultant support and high performing schools asked to share expertise. However, selection of schools has not been on the basis of value added alone: other performance data and information are taken into account.

The availability of value added results also provides useful information for monitoring progress in groups of schools subject to specific policies or administrative arrangements. Care needs to be taken that differences in value added are not misinterpreted: selection bias may mean that it is generally ineffective schools, for example, that are subject to a policy initiative, and so we would expect the value added for pupils in these schools to be relatively low. *Changes* in value added may be more useful, which would show the relative progress of pupils subject to the policy compared to pupils generally, before and after the introduction of a policy, for

example. However, as the value added measures are relatively recent and policy interventions may take some time to have a measurable impact, many of the policy analyses carried out so far using a value added approach are subject to important caveats.

The value added models themselves also have a policy use, such as to demonstrate the relative importance of factors which impact on pupils' progress, and to show that the school effects are in fact a relatively small part of the total variation. Therefore, tackling achievement gaps must not simply focus on the worst performing schools but also needs to consider the underperformance *within* schools.

References

DfEE (1998a) *The Autumn Package*, London: Department for Education and Employment

DfEE (1998b) *1998 Value Added Pilot: Supplement to the Secondary School Performance Tables*, London: Department for Education and Employment

DfES (2004), 'Variation in Pupil Progress 2003', *Statistics of Education Bulletin 02/04*. London: Department for Education and Skills

DfES (2007), 'Schools and Pupils in England, January 2007 (Final)', *Statistical First Release: SFR 30/2007*

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

Goldstein, H. (2003) *Multilevel Statistical Models: Third Edition*. London: Arnold

- Goldstein, H, Burgess, S and McConnell, B (2007) Modelling the effect of pupil mobility on school differences in educational achievement, *Journal of the Royal Statistical Society, Series A*, 170(4): 941-954.
- Goldstein, H., Rasbash, J., Yang, M., Woodhouse, G., Pan, H., Nuttall, D. & Thomas, S. (1993) 'A Multilevel Analysis of School Examination Results', *Oxford Review of Education* 19(4), 425-433.
- Goldstein, H and Sammons, P (1997) The influence of secondary and junior schools on sixteen year examination performance: a cross-classified multilevel analysis, *School Effectiveness and School Improvement*, 8: 219-230.
- Kirkup, C., Sizmur, J., Sturman, L. and Lewis, K. (2005) *Schools' Use of Data in Teaching and Learning*, DfES Research Report 671. London: Department for Education and Skills
- Mortimore, P., Sammons, P., Stoll, L., Lewis, D. and Ecob, R. (1988) *School Matters: The Junior Years*. Wells: Open Books.
- National Audit Office (2003) *Making a Difference: Performance of maintained secondary schools in England*. London: The Stationery Office
- Sammons, P., Thomas, S., Mortimore, P., Owen, C. & Pennell, H. (1994) *Assessing School Effectiveness: Developing Measures to put School Performance in Context*, London: Office for Standards in Education
- Saunders, L. (2000) 'Understanding schools' use of 'value added' data: the psychology and sociology of numbers', *Research Papers in Education* 15(3), pp241-258
- SCAA (1994) *Value Added Performance Indicators for Schools*. London: School

Curriculum and Assessment Authority

Schagen, I. & Hutchison, D. (2003) 'Adding Value in Educational Research – the marriage of data and analytical power', *British Educational Research Journal*, Vol. 29, No. 5, October 2003

Teddlie, C. & Reynolds, D. (Eds.) (2000) *The international handbook of school effectiveness research* London: Falmer Press.

Thomas, S., Peng, W. J., Gray, J. (2007) 'Value added trends in English secondary school performance over ten years', *Oxford Review of Education*, 33 (3), July 2007 , pp. 261-295

Tymms, P. & Coe, R. (2003) 'Celebration of the Success of Distributed Research with Schools: The CEM Centre, Durham', *British Educational Research Journal*, Vol. 29, No. 5, October 2003

Webster, W. and Mendro, R. (1997). The Dallas value-added accountability system. In J. Millman (Ed.), *Grading teachers, grading schools: Is student achievement a valid evaluation measure?* Thousand Oaks, CA: Corwin Press, Inc., 81-99.

Wilson, D. and Piebalga, A. (2008). Accurate Performance Measure but Meaningless Ranking Exercise? An Analysis of the English School League Tables, *CMPO, Working Paper No. 07/176*.

DCSF Achievement and Attainment Tables website:

<http://www.dcsf.gov.uk/performance/tables/>

Teachernet Guidance

<http://www.teachernet.gov.uk/management/payandperformance/threshold/archives/round3/applicationpacks/>