

Using national data for self-evaluation and school improvement

Ian Schagen explains how to find a clear path through the value-added jungle, and how to understand and use your performance and assessment (PANDA) output as part of school improvement.

Time was when the only data schools had to worry about was raw league tables, but we recognised those were deeply flawed and told us nothing about how schools were actually doing. But some of us well-meaning statisticians said long and loud that what we needed was value-added information, which would give schools a much better picture of how their pupils were doing, taking account of other factors including where they were starting from.

And behold, it was so (eventually), and now we have value-added information about schools in spades. It's in the league tables – sorry, Achievement and Attainment Tables – and in the new RAISEonline, and the Fischer Family Trust (FFT) produces loads of value-added information for schools and local authorities. But now we've got another problem: there's too much of this stuff for busy non-statistical school staff to wade through, and how are they to know what to ignore and how to understand and use the information that is useful? Like King Midas and his gold, we're in danger of drowning in too much of the good stuff.

In this article I will try to give some pointers to hard-pressed staff, which may help them to focus on the information that's really useful and allow them to understand it, and hopefully use it to drive improvements in teaching and learning.

What is value-added?

Value-added analysis is basically taking national data about pupils' performance linked to a set of other factors (prior attainment, sex, ethnicity, free school meals (FSM), etc.) and from this creating statistical relationships that can be used to define 'expected' performance for each pupil. If actual performance is better than this, then we give them a positive score (or 'positive residual' to use the jargon). If performance is lower than expected, then we give them a negative score (or 'residual'). Averaging up these positive and negative values for a whole school gives a school value-added score, which may be positive or negative – some agencies add a constant (e.g. 1000) to this because negative numbers make their brain hurt.

These are some important points to note about all this.

- The answers you get depend not only on the data but on the factors you take into account. Value-added scores using just prior attainment may be quite different from what you get taking measures of deprivation into account as well, or the fact that you are a grammar school. That is why you quite often get several different indicators presented, depending on what 'model' has been assumed.

- No value-added indicator is exact – they all have a built-in degree of uncertainty, based on the number of pupils included in the calculation and the statistical outcomes of the modelling process. These uncertainties should be made explicit, usually in the form of a ‘confidence interval’. If we say your value-added score is 1015 with a confidence interval from 1008 to 1022, this means we are 95 per cent sure that the ‘true’ value of your score is in the latter range. Such measures of uncertainty should always be presented.
- All such calculations take average performance as the norm, and in general there will be as many with value-added scores above average as below. Any rise in attainment across the board will be ignored by this methodology.
- Value-added analysis by its nature looks backwards. The pupils on which it is based have completed the key stage and achieved what they are going to achieve. If this information is to be useful, it must inform the school about areas of strength and weakness in its pupils’ performance, which can be translated into more effective teaching to improve next years’ results.

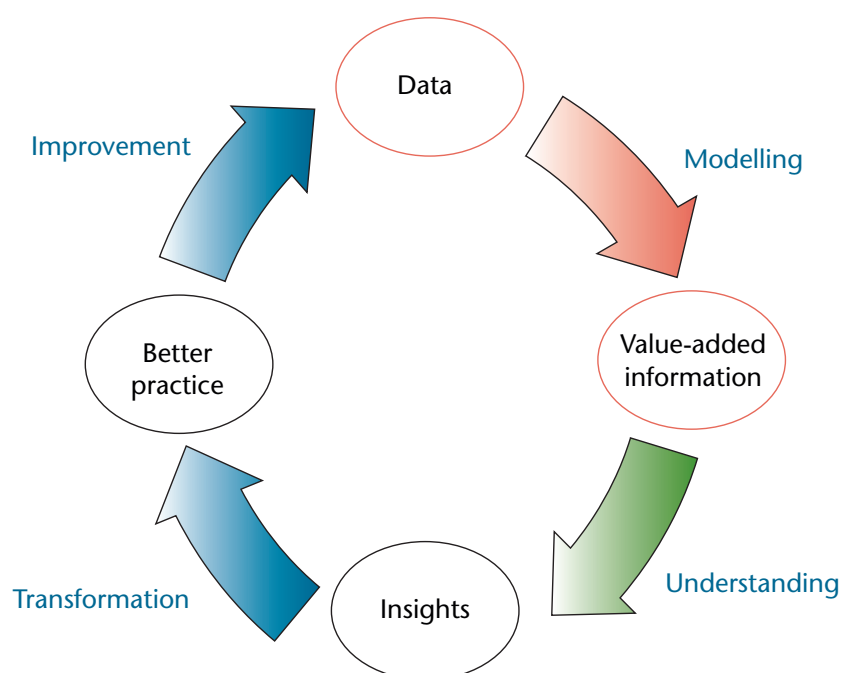
From the above, it is clear there is a chain of actions that need to take place in order for data to drive school improvement – this is illustrated in Figure 1. Red areas are the responsibility of the agency producing the value-added information. Blue areas are the responsibility of the school. The green arrow labelled ‘understanding’ is a shared responsibility – the agency needs to make value-added information clearly accessible, while the school needs to invest the effort it may take to understand it.

That’s the theory – but every link in this chain must function for the whole thing to be any more than a paper exercise in generating loads of numbers for their own sakes. If school staff cannot understand the value-added information, then the chain is broken and the information is worthless.

Understanding PANDA output

Performance and assessment (PANDA) reports for schools have been published by Ofsted for several years, based on the National Pupil Database (NPD) and in collaboration with the Department for Children, Schools and Families (DCSF). More

Figure 1 Chain of actions driving school improvement



recently, the same information is to be found on the interactive school performance website, RAISEonline. It is not possible to go through the available example reports in great detail in this article, but I will attempt to highlight some of the features that are

the answers you get depend not only on the data but on the factors you take into account

worth bearing in mind and some of the potentially useful analyses.

Ofsted produces training materials for understand-

ing the PANDA report. At the Ofsted website (www.ofsted.gov.uk) search for 'PANDA standards' and the first module of the training materials is the first result in the list of search results.

This includes examples of reports for both primary and secondary schools. In both cases there is an eclectic mix of information, presented in a variety of different styles and based on a range of different models and assumptions, which may be difficult for the inexperienced user to find their way through. I will try to highlight some of the most important features in order to encourage school staff to engage with this information critically.

The primary PANDA output is much shorter than the secondary equivalent, and comprises the following major elements:

- background information (including key stage 1 performance) by pupil characteristics such as sex, ethnicity, eligibility for FSM
- school characteristics such as percentage FSM compared with national percentage
- contextualised value-added measures (CVA)
- outcomes of core subjects over time at key stages 1 and 2.

Probably the most useful information is the contextualised value-added (CVA) data, based on national models of attainment controlling for prior attainment and other

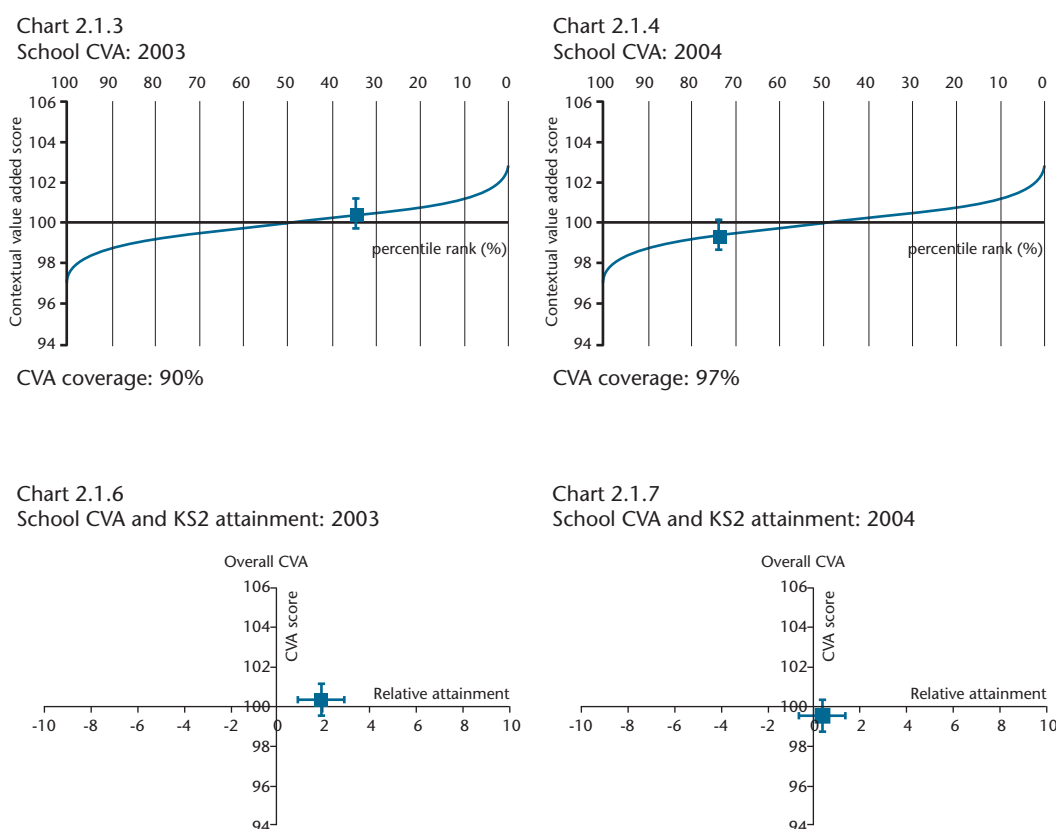
pupil and school characteristics. Further details of the CVA modelling are available on the DCFS website, by following this link: www.standards.dcsf.gov.uk/performance/1316367/CVAinPAT2005/?version=1

The CVA score is on a scale centred around 100 (1000 for secondary), and is presented with a 95 per cent confidence interval. Charts like those shown in Figure 2 illustrate the school's CVA measure compared with all schools in ascending order, and with the confidence interval marked. Below that, plots compare the CVA with the key stage 2 'relative attainment' (total key stage 2 score minus national average).

The top two charts show the CVA scores for all schools in England, in 2003 and 2004, plotted in ascending order. The vertical axis is CVA score and the horizontal axis is 'national percentile rank', which goes from 100 (lowest CVA) to 0 (highest CVA). The school of interest is plotted as a square blob in its position on the national S-shaped curve, with two little 'horns' above and below showing the 95 per cent confidence interval for its CVA score. If these 'horns' cross the horizontal line then the school is not significantly different from the national average in terms of CVA.

In 2003 the school was above average in CVA (above 100), but the confidence interval shows it was not significantly so. On the other hand, in 2004 the CVA was below average (less than 100) but again not quite significant. National percentile rank went from about 35 to about 75. This seems quite a large change, but it was produced by a relatively small shift in CVA from one year to the next (neither significantly different from the average) and is probably not a cause for serious concern unless a downward trend continues.

The bottom two charts also have CVA scores plotted on the vertical axis, but the horizontal axis is 'relative attainment' – essentially key stage 2 performance (not value-added) relative to the national average. In 2003 the

Figure 2 Sample charts showing a school's CVA measure compared with all schools

school was above average in absolute performance and also in CVA, but not significantly so for the latter. In 2004 it was not significantly different from the national average in either dimension.

Note that the exact scale on which the CVA measures are presented is not entirely clear, except that it is centred around 100/1000. The spread nationally is not specified, so these should not be confused with standardised scores, which have a predefined mean and standard deviation (e.g. 100 and 15).

The confidence intervals inspire less confidence when applied to very small sub-groups of pupils, as in the example shown in Figure 3.

There is one stated pupil, with a CVA of 99.4 and confidence interval of 1.9. In practice we can tell nothing from the results for a single pupil, and it is probably not sensible to look at results for groups of pupils

with fewer than about 10 cases. Otherwise, if you found a specific group (e.g. low-attaining girls) with sensible numbers and a CVA that was well below that for the school as a whole, you might conclude there was evidence that those pupils were under-performing relative to expectations.

The final bit of primary information is a set of plots of performance at key stages 1 and 2 over the last five cohorts and compared with national performance. These graphs are not value-added, and are more likely than anything else to tell you if your intake has changed over the last few years (see Figure 4).

Secondary PANDA output contains the following, in broad terms, covering both key stage 3 and key stage 4 (GCSE/GNVQ) results:

- school and pupil characteristics
- key stage 3 three-year summary, including CVA and attainment by pupil groups

Figure 3 Using confidence intervals with very small groups

Chart 2.1.15

CVA for groups within the school

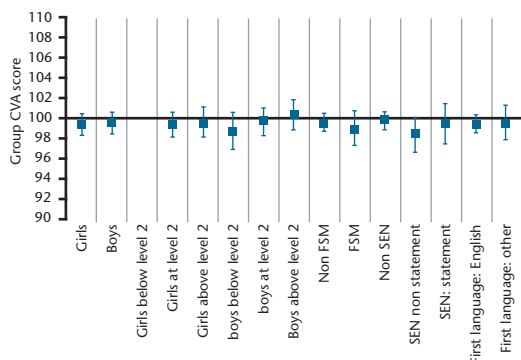


Chart 2.1.16

CVA for ethnicities within the school

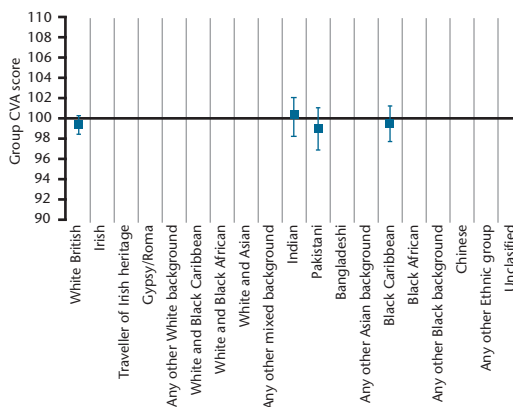


Table 2.1.17 Groups

	Girls	Boys	Girls < Level 2	Girls = Level 2	Girls > Level 2	Boys < Level 2	Boys = Level 2	Boys > Level 2	Non-FSM	FSM	Non-SEN	SEN: non-statement	SEN: statement	First Language: English	First Language: Other
Average. CVA score	99.4	99.5		99.4	99.7	98.7	99.6	100.4	99.5	99.0	99.8	98.4	99.4	99.5	99.5
Cohort	15	14		10	5	2	7	5	26	3	26	2	1	26	3
95% Confidence Interval	1.0	1.0		1.1	1.4	1.7	1.3	1.4	0.8	1.6	0.8	1.7	1.9	0.8	1.6
Significance															

Table 2.1.18 Ethnicities

	White British	Irish	Traveller of Irish heritage	Gypsy/Roma	Any other White background	White and Black Caribbean	White and Black African	White and Asian	Any other mixed background	Indian	Pakistani	Bangladeshi	Any other Asian background	Black Caribbean	Black African	Any other Black background	Chinese	Any other ethnic group	Unclassified
Average CVA score	99.5									100.2	99.1			99.6					
Cohort	23									2	1			3					
95% Confidence Interval	0.8									1.7	1.9			1.6					
Significance																			

- key stage 4 three-year summary, including VA/CVA for key stages 2–4 and key stages 3–4 and attainment by pupil groups
- post-16 three-year summary
- basic school characteristics
- attendance
- CVA plots for key stages 2–4 and 3–4
- results over time plots
- detailed GCSE results by subject area, with relative performance indicator (RPI) measures to show relative performance in different subjects.

Most of this stuff is the same as for primary schools (just more of it), although since group sizes tend to be larger there may be more significant results. The main extra information relates to GCSE subject areas, with tables produced which look like the one in Figure 5 (see page 32).

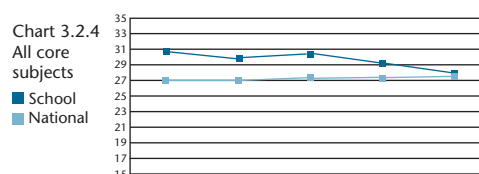
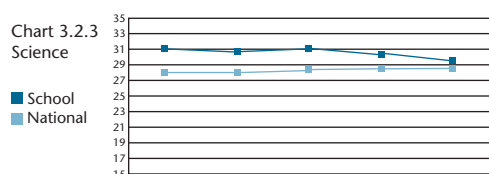
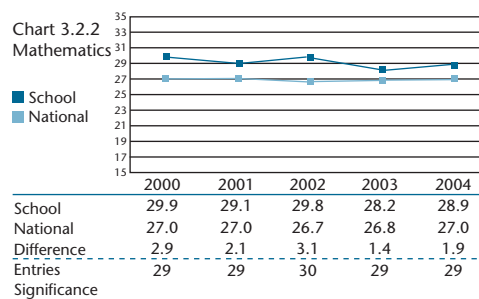
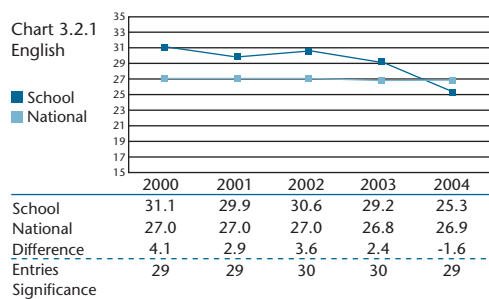
This shows where the percentages achieving various benchmark grades are significantly different from national values (but not taking account of prior attainment, etc.). It also gives the RPI value, which is computed by looking at the grades in all other subjects taken by the same students and comparing the difference between this average grade

Figure 4 Performance at key stages 1 and 2 compared with national performance

Anonymous Primary School: Attainment Summary, Key Stage 2

Average points score 2000–2004, by NC core subject

This page provides analysis of pupils' average points scores over the last five years in the National Curriculum core subjects. Significance tests have been performed on the data using a 95% confidence interval, and where the school value differs significantly from the corresponding national value **sig+** or **sig-** is shown below. Where the school value differs significantly to the previous year's, \uparrow or \downarrow is shown to the right.



and the subject grade with the same difference calculated nationally. A significant positive RPI (e.g. for PE/sports in Figure 5) could be interpreted as meaning that the schools' pupils did better in the subject than might have been expected given their grades in other subjects. This could give some information on the quality of teaching and learning, though any such conclusions would have to be qualified in the light of the many other factors that this particular calculation does not take into account.

Overall, there is a lot of information on the PANDA, based on a whole range of different models and ways of looking at the data. Some of this will be easier to digest than the rest, and it is probably best to focus initially on the elements that are most accessible and give you the greatest insights to help in your work.

Using value-added data for school improvement

So what are the general things you should look out for in any kind of value-added feedback, in order to identify aspects which

could be taken on board to improve teaching and learning?

- Look at CVA scores over the past year or so to identify your general position relative to expectations, but take account of confidence intervals. If any cohort is significantly different from usual, you may want to identify reasons for that.
- CVA for different subject areas and different groups may be more helpful, as it can help to identify where you are doing better and worse than expected. But be careful of looking at groups that are too small and placing too much reliance on their results.
- RPI may be helpful in secondary schools to identify departments that are doing well, and perhaps learning from their success.
- Trends over time in performance may not tell you whether you are improving or getting worse as a school, unless you are certain that the characteristics of each cohort are staying exactly the same. Trends in CVA over a few years may help to identify things that really are changing in the school rather than its intake.

Figure 5 GCSE results summary

Anonymous Secondary School: Summary of full GCSE results for 2004, All Pupils

Significance tests compare school data against the corresponding national data. Where sig+ or sig- is shown, the difference between the school value and the national is statistically significant with a 95% confidence interval. RPI provides a measure of whether attainment in a subject is significantly better or worse than attainment in other subjects taken by the same set of pupils, taking into account the national difference in attainment between subjects. These take use the new points scoring system.

Subject		Number of entries	Entry as a percentage of cohort	%A*-A	Sig	%A*-C	Sig	%A*-G	Sig	% Fail	Average points score	RPI	Sig
Information Technology	School	66	23.2	7.6		47.0		98.5		1.5	36.9		
	National	74,366	11.6	14.1		55.1		94.7		5.3	36.6		-0.4
Mathematics	School	26	9.1	23.1		69.2		100.0		0.0	41.2		
	National	560,246	7.2	23.3		65.9		95.0	sig+	5.0	40.2		-0.4
Music	School	26	9.1	23.1		69.2		100.0		0.0	41.2		
	National	46,246	7.2	23.3		65.9		95.0		5.0	40.2		0.2
Other languages	School	11	3.9	45.5		72.7		100.0		0.0	43.8		
	National	17,878	2.8	38.5		72.5		97.9		2.1	44.0		
Physical Education/Sports	School	49	17.2	38.8		77.6		100.0		0.0	44.3		
	National	116,617	18.2	15.6	sig+	57.0	sig+	99.4		0.6	39.4		5.0 sig+
Physics	School	62	21.8	51.6		100.0		100.0		0.0	47.5		
	National	28,336	4.4	39.4	sig+	88.6	sig+	99.6		0.4	47.7		1.2 sig+
Spanish	School	79	27.7	12.7		50.6		100.0		0.0	37.6		
	National	44,903	7.0	16.4		51.5		98.0		2.0	38.8		0.0
Statistics	School	2	0.7	100.0		100.0		100.0		0.0	52.0		
	National	33,066	5.2	17.5		69.6		97.0		3.0	39.9		
Total	School	2,343	n/a	15.9		61.3		99.1		0.9	n/a		
	National	4,230,644	n/a	14.8		56.3	sig+	97.1	sig+	2.9	n/a		

- Identify a few key measures in the output that you understand and feel are useful, and try to focus on those without trying to understand absolutely everything in the output. As you gain confidence, you may want to start looking at other things – but it may well be that not every piece of information is relevant, helpful or useful.
- Using past models to predict future results is something that should be done with care, especially for individual pupils. It may help to show a range of potential outcomes for individuals, and allow you to encourage them to aim for the top of that range, but beware of labelling pupils on the basis of such models.

Last words

This has been a whistle-stop tour through PANDA output as an example of value-added feedback for schools based on national data. One thing which is certain is that things are going to keep changing, as RAISEonline and other systems continue to develop. The important thing is to understand as much as is useful in the output in front of you, and

try to transfer that understanding to a new system by looking for similarities. In some ways understanding statistical output is a bit like learning a foreign language. To begin with, it all looks very daunting, but once you learn a few bits of it you can start to make sense of more and more – although you may never grasp the equivalent of the pluperfect subjunctive!

In the next issue of *practical research for education*, I'll explain how to understand and use the output from the Fischer Family Trust.

Further reading

Schagen, I. (2000). *Statistics for School Managers*, Westley: Courseware Publications. www.courseware-publications.co.uk/page15.html

Weblinks

Ofsted – for training materials www.ofsted.gov.uk

DCSF – for details of contextualised value-added modelling
www.dcsf.gov.uk/performance/tables/pilotks2_06/k5.shtml

About the author

Ian Schagen was Head of Statistics at NFER for some years, and is now Chief Research Analyst at the Ministry of Education in New Zealand.

Contact details

Ian Schagen
Ian.Schagen@minedu.govt.nz

Copying permitted

The NFER grants to educational institutions and interested bodies permission to reproduce this item in the interests of wider dissemination.