Reporting Value Added measures using statistical process control charts

Paul Trower & Carol T. Fitz-Gibbon
C.E.M. Centre. University of Newcastle upon Tyne

Introduction

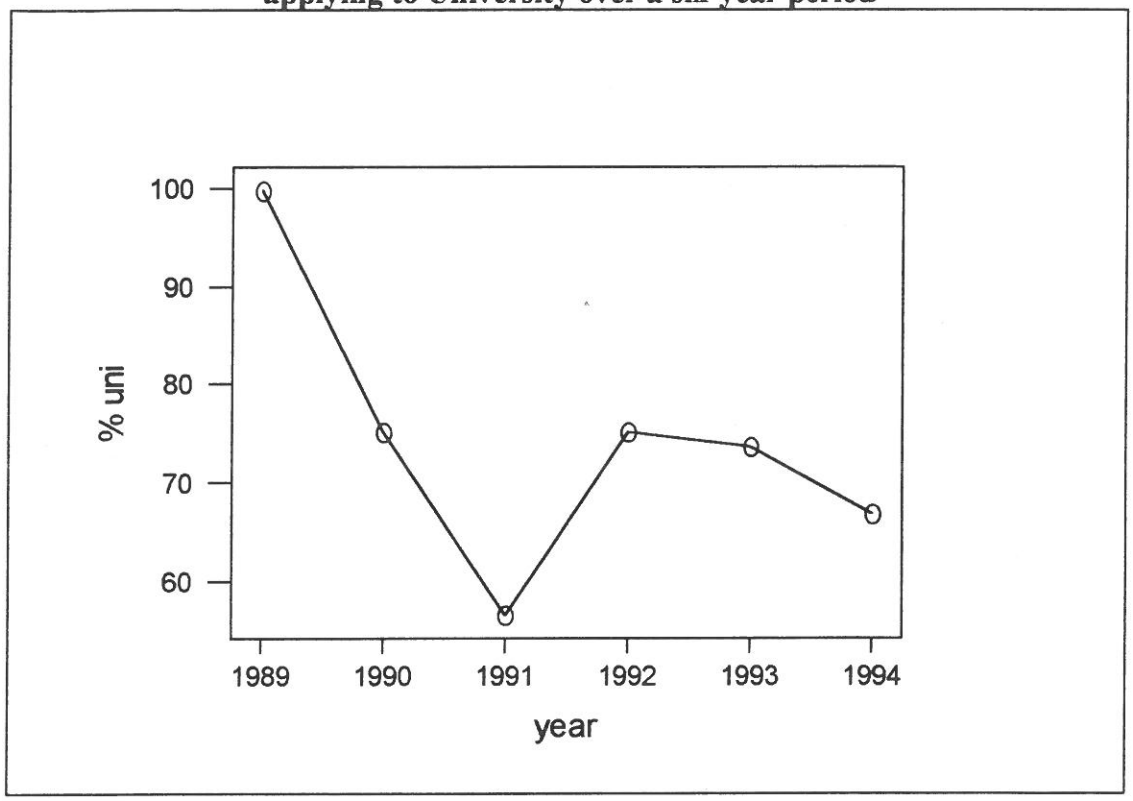
This paper is not designed to be an overview of control charts and their uses but to suggest some ways in which they may be adapted to, and applied in, educational settings. A good introduction to the subject can be found in DeVor et al. (1992) and some useful warnings about the misuse of control charts are contained in Alwan & Roberts (1995) and Caulcutt (1995). There are those who baulk at any technique designed for industry being used in education but while the monitoring of the 'end product' in education is more rigorous than in most industries in that the major assessments are external to the organisation, the goals are similar.

Control charts were introduced during the 1920s when Walter Shewart of Bell Laboratories suggested a statistical approach to the scrutiny of variation in outcome measures based on continual on-line monitoring. Judgements about the process were to be made on the basis of the patterns in the outcome measurements presented in diagrammatic form as a control chart. The baton was taken up by W. Edwards Demming in the 1950s. He projected Shewart's ideas on an institutional basis and central to his approach was the emphasis on the responsibilities and obligations of senior management.

Measurements can be made regularly on a process - frequently in some processes (e.g. daily pupil absences), rarely in others (e.g. annual examination results). The results of these measurements can be plotted on a chart with time on the horizontal axis. Such diagrams are useful in themselves in that trends can become apparent and some idea is given of the degree to which the readings vary. Informal but informative judgements can be made from the distribution of points on such a chart; a weekly pattern in the numbers of pupils absent may become clear, a gradual improvement in the examination pass rate is easier to detect visually than from a table of figures when the sheer volume of information can be overwhelming. Figure 1 shows an example of such a diagram where the percentage of students in a school who applied to University is plotted for six years.

Figure 1

Percent of students in a school applying to University over a six-year period



Shewart's proposal was that there be some method of indicating formal significance to aid decision making as opposed to the informal judgements made from the points alone. This would be done by superimposing on the time-series of data points three lines - the centre line at an appropriate target value and the outer, or control, lines set at a suitable distance above and below the centre line. If the plotted points fall between the control lines and are distributed randomly about the centre line, the system can be considered to be stable (in control). If there are points which fall outside the control lines or there is a distinct pattern among the points, then the system has been under some particular influence (out of control).

Setting up Control Charts

To set up a control chart, the positions of the central, 'target' line and the control lines must be determined. In some cases, a target value is available but these will be rare in education and more commonly the central line will be positioned at some historical average (the mean weekly absence rate in the school for the past term) or at a national average figure (the national mean value-added). This central line ought to be selected on the basis of making comparisons of like with like. To set the 'target' GCSE pass rate, for example, at the national average would not be reasonable for a selective school or for a school whose pupils generally have very low prior attainment.

The method used in deriving the control lines will differ if the quantity of interest is a measurement on each individual (A-level French score) rather than an attribute (present or absent on a particular day). In this paper we concentrate on charts for measurements, details of constructing charts for attributes can be found in DeVor et al.

Control Charts for measurements

The crucial requirement in positioning the control lines on a measurement chart is an estimate of the variability (usually the standard deviation) of the measurements when the process is stable. If there is naturally a large variability, the lines will need to be far apart to indicate this; if the measurements tend to be consistent, the lines can be close together.

One approach is to gather data from the process within the school and, when enough has accumulated, to use the measurements to estimate the standard deviation of the process. When the information is gathered frequently - as it would be for pupil absence, this is relatively simple to do. There will be no long waiting period and the process will, with luck, remain stable during this initial setting-up period.

If the data are produced infrequently, as with annual examination results or value-added, this approach is untenable. The time required to gather enough information to estimate the variability would be prohibitive - twenty five observations is the recommended minimum - and the system would have altered during the period so rendering any findings out of date. An alternative approach will be needed in this case.

It may be possible to estimate the standard deviation of the measurements using concurrent data and then there will be no time lag. If we take examination results as a paradigm, a school could calculate the standard deviation of its own results in a particular year and use this as the basis for setting the control lines. It would be comparing itself with itself in each separate year and would conclude that the system was always stable. A more valuable approach is to base the estimate on data pooled from many schools if such information is available. The larger sample will provide a more reliable estimate as long as the sample is chosen so that like is compared with like. For example, to pool the raw examination results for all schools would produce an inflated estimate of the standard deviation. If, however, the raw results are adjusted for prior achievement as in value-added analyses, pooling the data from all schools would lead to realistic estimates of the standard deviation.

Once an estimate of the standard deviation of the measurements, $\hat{\sigma}$, has been produced, the control lines can be plotted on the chart. If there are n measurements at a particular time and they are

combined as an average (e.g. the mean value-added for a department) the lines will be placed at a distance $\frac{3\hat{\sigma}}{\sqrt{n}}$ above and below the central line.

Example from the ALIS project

The ALIS project produces value-added indicators from GCSE to A-level for students in subscribing schools and colleges. The project (Fitz-Gibbon 19##) began in 1982/83 and in 1994/5 included over one-third of all A-level candidates. For each student, schools provide the GCSE scores which are matched to the subsequent A-level performances together with responses from a fairly detailed questionnaire on attitudes and learning processes.

The measure of prior achievement which best predicts A-level performance is the *average* GCSE score - the arithmetic mean of the GCSE scores (A*=8, A=7, B=6, . . . , G=1) for all subjects attempted by a student. For each A-level subject separately, with grades coded in the standard UCAS scale (A=10, B=6, . . . , N=0 together with U= -2), the results are pooled and the Ordinary Least-Squares (OLS) regression line of A-level score on Average GCSE computed. A student's residual from this line is the value-added score for that student in that subject in that year. This, essentially, is a measure of how well the student performed in that subject as compared with other students in that subject of similar prior achievement. As such, it can provide the basis for fair comparisons of student performance.

The student subject residuals can be averaged for all students in a department to provide a department residual.

In order that comparisons can be made across subjects and across years, the residuals are scaled to have standard deviation of unity across all students in any given subject. A consequence of using OLS regression is that the average residual across all students taking a subject will be zero. Both the raw and the standardised residuals are reported to schools.

Table 1 and Figure 2 give the departmental standardised residuals (value-added scores) and the control charts for two French departments over a six-year period. The standardised residuals are used so the 'target' line will be at zero, the average for all French departments in schools and colleges subscribing to the project. The standard deviation is unity and the control lines will therefore be at $\pm\frac{3}{\sqrt{n}}$ where n is the size of the department A-level entry.

On the chart for school A the control lines have moved towards the centre line over the years indicating that the French department is steadily increasing in size. The points representing the

department value-added scores have swung from positive to negative with no apparent pattern over the six years and for the first five fell between the control lines. In 1993 the residual was positive and above the upper control line signalling an exceptionally fine performance by the students. It is important to note that, while the results for 1989 and 1992 were below what might have been expected, they were quite consistent with that expectation allowing for the variation in results that naturally occur with such small numbers of candidates. To search for reasons why the results were not as good as expected in such circumstances would be fruitless. This is not, of course, to say that the results could not be improved, merely that they are not out of line with what might reasonably be expected given the prior achievement of the students in the group.

The department will have particular information that could throw light on the 1993 results; there may have been a new initiative two years' earlier, a move to better buildings, a change of head of department. The chart indicates that something special has happened and that the circumstances are worth investigating so that lessons, perhaps, can be learned.

The chart for school B shows that there has been a small increase in the French department A-level entry over the years and that the department value-added indicator has been pretty much in line with expectation. All points are within the control lines and, while there is no strong trend, the results for the past five years have shown a slight but steady improvement.

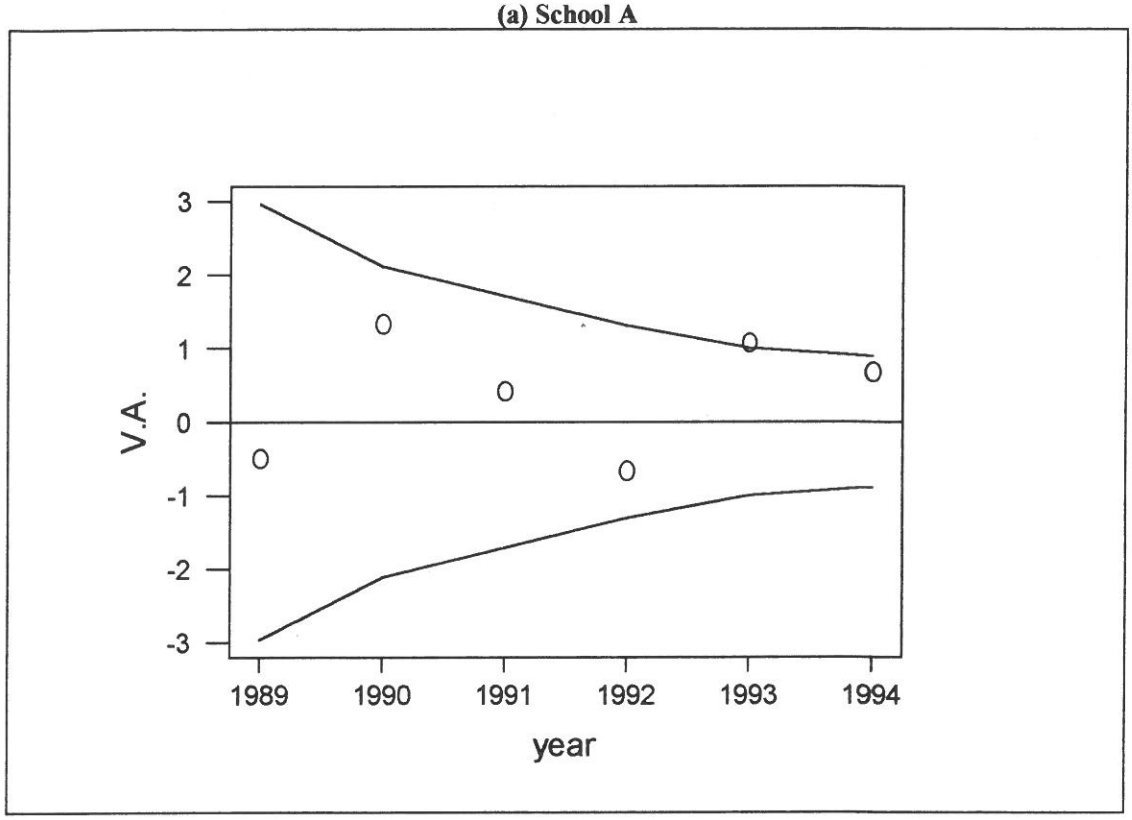
Table 1
French department standardised residuals (and cohort sizes)
for two schools over six years

School	1989	1990	1991	1992	1993	1994
A	-0.49 (1)	1.36 (2)	0.42 (3)	-0.66 (5)	1.09 (9)	0.70 (11)
B	0.10 (7)	-0.23 (8)	-0.11 (17)	-0.14 (17)	0.13 (19)	0.33 (21)

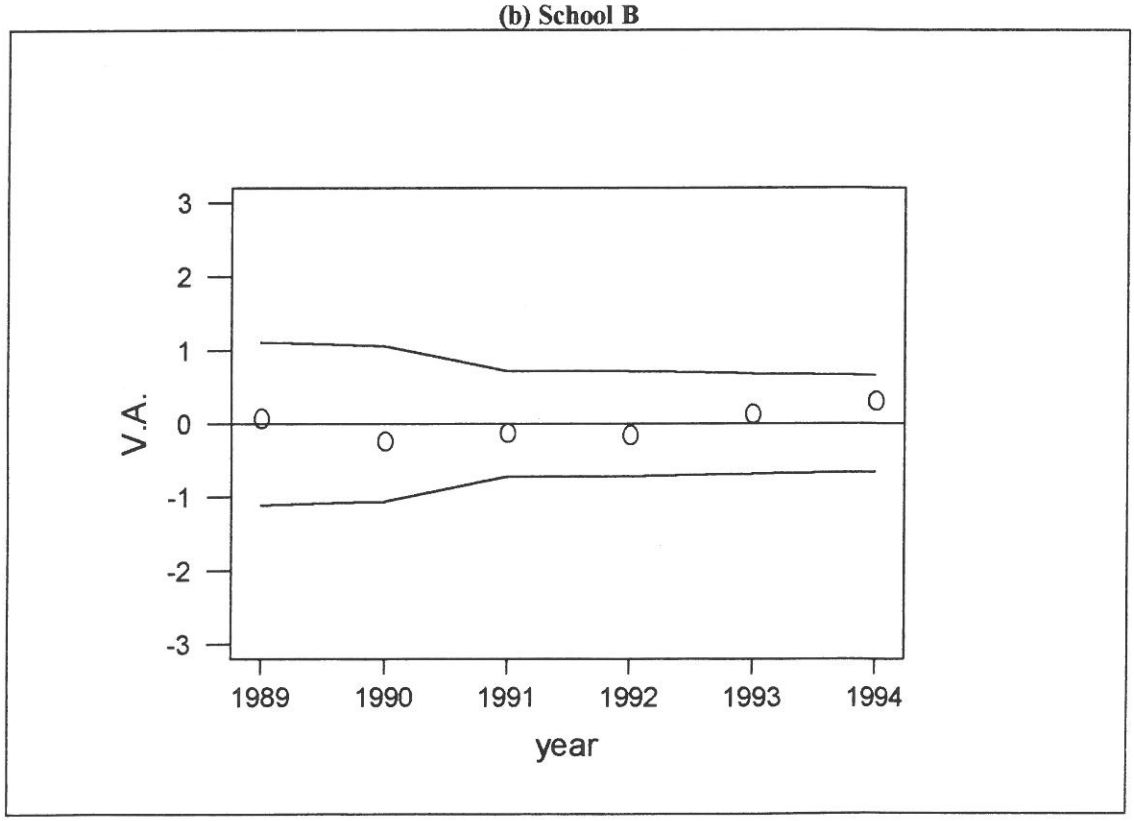
Figure 2

Control Charts for Value-added in two French departments

(a) School A



(b) School B



Benefits of Control charts for Schools

We have for some time encouraged institutions in the ALIS project to adopt control-chart approach to monitoring their value-added results and have produced software to create the charts. The feedback has been most enthusiastic and positive.

Some of the benefits to the use of control charts in school and colleges are reported to be:

- They help to curb the over-zealous interpretation of one year's (or short time-span) results and encourage a more measured, longer-term view.

- The use of, and familiarity with, control charts can lead to the development of an understanding of variation - a skill which Demming claimed to be essential to all managers.

- They indicate when especially noteworthy events have occurred and highlight possible trends in a process. In doing so, they can stimulate a search for the reasons behind such events and from the findings, lessons may be learned. There can be a tendency among some managers to respond to every deviation from the 'expected' - to attempt constantly to 'correct' results which are simply the product of natural variability. This tends to lead to *more* variable results and greater instability in the system. Control charts can check the urge to tinker unnecessarily.

References

Alwan, L.C. and Roberts, H.V. (1995) The problem of misplaced control limits. *Journal of the Royal Statistical Society (C)*, **44,** 279-288

Caulcutt, R. (1995) The rights and wrongs of control charts. *Journal of the Royal Statistical Society (C)*, **44,** 279-288

DeVor, R.E., Chang, T., Sutherland, J.W. (1992) *Statistical Quality Design and Control.* MacMillan

Fitz-Gibbon, C.T.