The Principal Axis Approach to Value Added Calculation

Qingping He

(*Office of Qualifications and Examinations Regulation, Coventry CV5 6UB, UK*)

Peter Tymms

(*School of Education, Durham University, Durham DH1 1TA, UK*)

Abstract

The assessment of the achievement of students and the quality of schools have drawn increasing attention from educational researchers, policy makers and practitioners. Various test-based accountability and feedback systems involving the use of value added techniques have been developed for evaluating the effectiveness of individual teaching professionals and schools. A variety of models have been employed for calculating value added measures, including the use of linear regression models which link students' present and prior achievements. One of the limitations associated with the use of the conventional linear regression methods in value added calculation is that the value added measures are likely to be overestimated for students with higher prior achievements while underestimated for students with low prior achievements. This study explores an alternative approach, the principal axis approach, to calculating value added measures which can eliminate some of the limitations associated with the conventional linear regression methods.

Keywords: Accountability, value added, large scale assessment, school effectiveness, linear regression models, principal axis

Introduction

Educational assessments can be used as a means to identify the effectiveness of specific teaching and learning strategies with regard to their intended effects, in addition to providing

important information about what students know and are able to do in particular subject areas (Wiliam, 2010). To help school improve its performance, various accountability and feedback systems involving the use of results from large scale assessments as performance indicators have been developed in many parts of the world for evaluating the effectiveness of schools and teachers (e.g. Raudenbush and Willms, 1991; Willms, 1992; Fitz-Gibbon, 1996, 1997; Sanders et al., 1997; Yang et al., 1999; Hamilton et al., 2002; Tymms, 2002; Demie, 2003; Schagen and Schagen, 2003, 2005; Harker and Tymms, 2004; McCaffrey et al., 2004; Wilson, 2004; Ray, 2006; Gorard, 2006; Lissitz, 2006; Taylor and Nguyen, 2006; Ray, 2006; DfE, 2011; Hout and Elliott, 2011; Isenberg and Hock, 2011; Timmermans et al., 2011; Wei et al., 2012; Coe, 2003). It has long been recognized that test scores only reflect the present achievement that a student has attained at the time of testing and will not provide a complete picture of the effect of the school on the student over the period of study in the school (see, for example, Crane, 2002). As a results, most test based accountability systems use value added measures which involve the comparison of students' present or outcome achievement with their prior achievement to assess the effect of schooling for the specified period of study in the school (see Fitz-Gibbon, 1996, 1997; Sanders et al., 1997; Tymms, 2002; McCaffrey et al., 2004).

Test scores reflect the combined influences of a number of factors such as the learning environment in the school, the social-economical background of the students, the student's attitudes towards study, the academic achievement attained before entering the school, and many others. When calculating value added measures, the present performance measure of a student is normally partitioned into two components, with one related to the prior achievement of the student and the other related to the effect of the school and other factors that may have influences on the progress made by students. This latter component supposedly represents contributions from the school attended and the other factors. As the value added is

assumed in part to reflect the effect of the school, its accurate calculation is important in order to provide an objective assessment of the progression made by the student. However, it is difficult to provide accurate estimate of the contribution to a student's present performance from his/her prior achievement, and various theoretical and empirical models, including the simple single level linear regression models and more complex multilevel models with fixed effects or random effects have been proposed for value added calculation (e.g. Bryk et al., 1996; Fitz-Gibbon, 1996; Sanders and Rivers 1996; Goldstein, 1997, 2003; Ray, 2006; McCaffrey et al., 2004; Lissitz, 2006; DfE, 2010, 2011; Wei et al, 2012).

In England, value added calculated based on the simple linear regression models was used as school performance measures before 2006. From 2006 to 2010, value added measures based on both the simple linear regression models and the complex multilevel models were used. The multilevel models used in England are also termed the contextual value added (CVA) models which take into consideration the contribution from the differing characteristics of students and schools to students' outcome performance measures (DfE, 2010, 2011). However, as demonstrated by Fitz-Gibbon (1997), students' characteristics, including socioeconomic status (SES), account for only a small proportion of the variance in outcome measures at the student level and have limited influence on the value added measures at individual student level for the English education system. From 2011, the UK government ceased to use CAV models for deriving school value added performance measures in England due to the difficulty in and confusion caused by interpreting their results. The linear regression models have however been retained due to their simplicity in terms of data collection and analysis and interpretation of results. However, as will be argued, one of the limitations associated with the use of the conventional linear regression method in value added calculation is that the value added is likely to be overestimated for students with higher prior achievements while underestimated for students with low prior achievements. The

present study investigates such limitations and proposes an alternative approach which represents a more appropriate, simpler and fairer approach and overcomes some of the limitations.

The OLS Regression Approach to Value Added Calculation

As indicated earlier, the conventional or Ordinary Least Squares (OLS) regression models have been used for value added analysis in England since the introduction of value added school performance measures (Fitz-Gibbon, 1996, 1997; Ray 2006; DfE, 2010, 2011). The method used for value added calculation involving OLS regression models can be summarised as follows. The relationship between the present performance measures (Y_i) of student i and his/her prior performance measures (X_i) is represented using a linear regression model like the one shown below:

$$Y_i = kX_i + Y_0 + R_i \tag{1}$$

where k and Y_0 are model parameters. R_i is the residual representing the departure of the present achievement from the model predicted value. Model parameters are estimated by minimising the squares of the vertical residuals (i.e. the least squares method). It is to be noted that linear regression analysis is appropriate when the purpose of the investigation is mainly about predicting variable Y (predicted variable) from observations on variable X (the predictor) or investigating the association between the two variables. Furthermore, in the linear regression model represented by Equation (1), the explanatory variable X is taken to be fixed (i.e. without measurement and sampling errors) and that only the response variable Y is a random variable (see Krzanowski, 1998). In the context of value added analysis, the term $kX_i + Y_0$ in Equation (1) represents the average present performance measure of students with

similar prior performance measures and can be interpreted as reflecting the influence of the prior achievement on the present achievement for the population under study as a whole. If there is a perfect linear relationship between present achievement Y and the prior achievement X (i.e. the square of the correlation between the two variables R^2 is 1.0, or the fitted line runs through all data points) then the influence of the school will be the same for all schools if the data used to derive the relationship were collected for students from a group of schools. In such a case, the variation in the present performance measures among the students can be solely explained by variation in their prior performance measures, and the residual R_i will be 0. The regression line, or the line of best fit, $Y = kX + Y_0$ can also be interpreted as representing the average expected present student's performance measure based on the prior performance measure. For an individual student i, the residual R_i is defined as:

$$R_i = Y_i - (kX_i + Y_0) \tag{2}$$

The residual R_i is also termed value added and can be interpreted as measuring the relative magnitude of progress made by the student in the time span between the present and prior performance measurements in relation to the average present performance for students with similar prior academic abilities. This value added therefore reflects the influence of factors other than the level of achievement attained by the student at the beginning of his/her study on the outcome achievement. The value added is assumed to reflect, to a certain extent, the influence of the school on the individuals concerned over the period of study. If the value added is positive (i.e. the data point is above the regression line), the student performed better than the average performance of students with similar prior performance measures. On the other hand, if the value added is negative (i.e. the data point is below the regression line), the student performed less well than the average performance of students with similar prior

achievement. Undoubtedly, all schools will aim to raise their students' academic achievement level, but they vary in achieving their objectives. Some schools perform better than others due to the existence of differences in the learning environment between schools. Students studying in better-performing schools will have an advantage over students studying in other schools in terms of learning. The average aggregated value added for all students in a school can be defined as:

$$VA_{school} = \frac{1}{N} \sum_{i=1}^{N} R_i \tag{3}$$

where N is the number of students in the school. This average aggregated value added will reflect the overall performance of the school in comparison with the average performance of other schools. A positive average aggregated value added will indicate that the school performed better than the average performance of all schools, while a negative aggregated value added will indicate that the school performed less well than the average performance of all schools.

An important assumption of the linear regression models outlined above (regressing Y on X) is that there is no error associated with the prior performance measure (X) and only the present performance measure (Y) is the random variable and has errors. However, as with the present performance measures, the prior achievement measures will also have errors. The regression approach to value added analysis requires that the present and prior performance measures are positively correlated. Furthermore, regression models are normally used for the purpose of prediction. However, value added analysis is not simply about using prior achievement to predict present achievement, and linear regression is only one of the approaches that are used to describe the relationship between the two measures. For example, it is possible to undertake an inverse regression (regressing X on Y involving minimising the

squares of the horizontal residuals) to derive a relationship between the two variables *X* and *Y* similar to that represented by Equation (1) but with different values for the two model parameters. Visually, in the case of a bivariate dataset that is normally distributed, the density of the data in the *x-y* plane will have contours that are concentric ellipses, and the regression line for *Y* on *X* will run through the ellipses where they are tangent to the vertical line, while that for *X* on *Y* will pass the ellipse where they are tangent to the horizontal line and will therefore be different for the line for regressing *Y* on *X* (see Friendly et al., 2013). The shape of the ellipses, as characterised by the values of the semi-major and semi-minor axes, will be determined by the covariance or correlation of the two variables. Figure 1 shows the two regression lines derived for a bivariate dataset using the approach outlined above. One ellipse is also shown in the graph for illustrative purpose. One of the symmetry lines, which the extension of the semi-major axis of the ellipse and lies between the two regression lines, for the dataset is also plotted in Figure 1. All three lines pass through the central point of the dataset ($\overline{X},\overline{Y}$) (where $\overline{X}$ is the average of the prior performance measure and $\overline{Y}$ is the average present performance measures). Although both regression lines are normally used to represent the relationship between two variables for a bivariate dataset, it will be argued that an important disadvantage of the use of Equations (1) and (2) for value added analysis is that the value added thus calculated will be overestimated for students with higher prior achievements but underestimated for students with low prior achievements as a result of minimising the squares of the residuals (the departures of the present performance measure *Y* from the average of all students with similar prior performance measures) under the assumption of no error on *X*. If the inverse regression is used, the opposite effect will occur. That is the value added will be underestimated for students with higher prior achievements but overestimated for students with low prior achievements. This issue associated with the use of Equations (1) and (2) to calculate value added has been raised by Harker and Tymms

(2004) who have noted how the positively biased value added towards groups of high ability students can lead to a compositional effect (see also Hauser 1970, 1974; Bryk and Raudenbush 1992).

<Figure 1>

The present study proposes an alternative approach that alleviates some of the problems associated with the conventional regression method used for value calculated calculation discussed above.

The Principal Axis Approach to Value Added Calculation

As indicated above, an important assumption made in the use of a regression model like Equation (1) for value added calculation is that the independent variable representing the prior achievement X is measured without error. Only the dependent variable representing the present achievement is assumed to be a random variable and have an error component involving both sampling error and measurement error. Since both the present performance and the prior performance measures are generally represented by test scores, they will have measurement errors as a result of unreliability of the tests used (see He and Opposs, 2012; He et al., 2013). Both variables should be treated equally as random variables in the analysis (A more sophisticated approach might weigh the line in proportion to the ratio of errors on the two measures. Any lines that lie between the two regression lines would be equally acceptable for describing the relationship between the two variables effectively. However, they measure slightly different characteristics of the data. It is clear from Figure 1 that there are situations where a data point can be above one fitted line but below another line and verse versa. A residual for each data point with respect to each of the three lines can be defined using an

equation similar to Equation (2). The extent of the difference in residuals for the different lines will depend on the magnitude of the covariance or correlation between the two variables or the ratio between the two axes of an equal density ellipse. Once we accept that both X and Y should be treated equally as random variables in the analysis, the line of best fit that can describe the structure of the data most accurately and effectively would be the semi-major axis of the concentric ellipses, not the conventional regression lines based on regressing Y on X or X on Y. The semi-axis of the concentric ellipses is a line of symmetry of the data points. In the case of a bivariate dataset like that described above, the major axis or the semi-major axis of the concentric ellipses corresponds to the first or major principal axis (PA) from the Principal Component Analysis (PCA). The major principal axis produces the maximum variances for the dataset along the direction of the line. PA is particularly appropriate for describing the bivariate scatter of the data in situations where both dependent and independent variables are measured with similar magnitude of errors and are treated symmetrically as random variables.

In the present study, we propose to use the major principal axis to represent the relationship between the student's present performance measure and his/her prior performance measure. For a bivariate dataset, the equation of the principal axis derived using the PCA technique or any other methods can be expressed as:

$$Y_i = k_{pa} X_i + Y_{0,pa} + R_{i,pa} \tag{4}$$

where k_{pa} is the slope of the principal axis and $Y_{0,pa}$ is the intercept. The slope k_{pa} and $Y_{0,pa}$ can be expressed as:

$$k_{pa} = \tan \theta$$
$$Y_{0,pa} = \bar{Y} - k_{pa} \bar{X} \tag{5}$$

where $\bar{Y}$ and $\bar{X}$ are the means of the present performance measure and prior performance measure respectively, and θ is the principal angle (the angle between the major principal axis and the x-axis in the two dimensional x-y coordinate system, see Figure 1) and can be expressed as (see Preisendorfer, 1988):

$$\theta = \frac{1}{2}\arctan\frac{2\sum_{j=1}^{N}(X_j - \bar{X})(Y_j - \bar{Y})}{\sum_{j=1}^{N}(X_j - \bar{X})^2 - \sum_{j=1}^{N}(Y_j - \bar{Y})^2} + \frac{n\pi}{2} \tag{6}$$

In Equation (6), n is 0 or an integer, and N is the sample size or number of data points. If the two variables are standardised to have the same deviation or variance and the same mean value, then $\theta = 45^o$, and the principal axis in this case is termed standardised principal axis or standardised major axis which would be particularly preferred when the two variables are not measured on comparable scales (see Warton et al., 2006). This would be suitable for the present study as the current and prior performance measures are reported using different scales (see later discussion). As the regression lines and the major axis pass the central point of the dataset, the line of regressing Y and X represents a clockwise rotation of the major principal axis, while the line of regressing X on Y an anticlockwise rotation. The magnitude of the rotation depends on the strength of the correlation between the two variables, the strong the correlation, the small the rotation. All the three lines will overlap each other when the relationship between the two variables is strictly linear. The relationship between the dependent variable Y and independent variable X derived using the PCA approach is the same as that derived by minimising the Total Least Squares (TLS, i.e. the squares of the perpendicular distance of the data points to the fitted line) which are normally termed as the

Orthogonal Regression (OR) or Orthogonal Distance Regression (ODR) by some researchers (see, for example, Isobe et al., 1990; Boggs and Rogers, 1990; Warton et al., 2006).

In the case of $\theta = 45^o$, the two model parameters, the slope and the intercept of the major axis, $k_{pa} = 1.0$ and $Y_{0,pa} = 0$. Equation (4) of the major principal axis becomes:

$$Y_i = X_i + R_{i,pa} \tag{7}$$

That is the major principal axis in this case is simply the identity line $y = x$. The line $y = x$ therefore represents the average present performance measure or the average expected standardised scores based on the prior performance measure. The residual or value added for an individual student i calculated using the PA method simply becomes:

$$R_{i,pa} = Y_i - Y_{pa}(X_i) = Y_i - X_i \tag{8}$$

Equation (8) indicates that the residual calculated using the PA method is simply the difference between the present performance measure and the prior performance measure for a student if the values of both measures are standardised in the same way. Data points above the line $y = x$ will have positive value added, while data points below the line $y = x$ will have negative value added. An intuitive interpretation of the validity of the use of this method in value added calculation is that as both test scores have been standardised in a similar fashion, they become comparable, and the definition for the residual using Equation (8) removes the influence of the level of prior achievement of the individual from the present performance measure directly. In other words, the difference between the present performance measure and the prior performance measure in this case represents the relative progress made after taking into consideration the prior ability of the student and therefore only reflects the effect

associated with the external factors, which include the influence of the school on the progress made by the student. It is to be noted that in the conventional PCA approach to Factor Analysis (FA), a new reference frame involving the rotation of the principal components (axes) is frequently used to represent the relationships between variables for the purpose of grouping the observed variables and facilitating data interpretation (particularly for dataset with more than two variables), although such interpretation can be complex (see Jolliffe, 2002). In the case of two observed variables, the reference frame based on the principal axes will be the most appropriate for data interpretation and any further rotation will not be needed. Moreover, for the purpose of the present study, the data is still presented in the x-y frame rather than in the reference frame of the two principal axes. This is because in the principal axes frame the residuals will just be the projected values of the transformed data points on the minor principal axis, which will be the same as the perpendicular distances of the data points to the major principal axis. Although such residuals will be proportional to the residuals calculated using Equation (8) based in the x-y frame, their interpretation is difficult as they will involve residuals related to both the x value (the prior performance measure) and the y value (the present performance measure) of the data points.

The CEM Centre Yellis Value Added Project Case Study

The Centre for Evaluation and Monitoring (CEM) at Durham University, UK, has been providing value added information for schools for self-evaluation and management through a number of performance indicator systems which involve the comparison of results from baseline assessments provided by the CEM centre and outcome measures from official national tests and public examinations both in the UK and abroad (see http://www.cemcentre.org; Fitz-Gibbon 1996; Tymms and Coe, 2003).

The GCSE examinations in England

In England, the General Certificate of Secondary Education (GCSE) is the main school-leaving qualification taken by students aged 16. GCSEs are available for over 50 subjects, including mathematics, English, sciences and many others. GCSEs are assessed by varying amounts of internal or school-based assessment as well as external assessment. Students taking GCSEs are classified into eight performance categories known as grades, from A* to G with A* the highest grade and G the lowest grade. Students who fail to achieve one of the grades are unclassified (grade "U"). In addition to certification and selection for further study or training for individual learners, GCSEs are also used as school performance measures for accountability purpose either in the form of raw results such as percentages of students achieving certain grades set by the government or used as outcome measures for deriving value added performance measures (see DfE, 2010, 2011; West, 2010, West et al., 2011).

The Yellis value added project

One of the systems provided by the CEM centre to schools is the Year 11 Information System (Yellis) for students aged from 14-16. Yellis has been used by over 1,200 secondary schools in the UK, involving about 200,000 students every year (see http://www.cemcentre.org/yellis). It provides an innovative baseline test for students aged 14 in schools which is used as the prior achievement measure for all students involved in the Yellis project. The Yellis baseline test was made as curriculum free as possible in order to provide an aptitude measure rather than the academic attainment measure of the students and included two main sections, vocabulary, and mathematics. The Yellis vocabulary test was XXX minutes long and had a maximum of XXX marks, and the mathematics test was XXX minutes long and had a maximum of XXX marks. These tests were taken under examination conditions. The Yellis baseline score is defined as the average score of the vocabulary section

score and the mathematics section score. The Yellis baseline assessment was designed for the full ability spectrum of the whole population.

After the students have taken their GCSE exams (results from individual subjects are used as the outcome measures or present performance measures) at the age of 16, value added measures for individual students are derived by comparing their GCSE results from individual subjects with the Yellis baseline test scores obtained two years previously using the conventional regression method outlined above. The Yellis scores and GCSE subject results are highly correlated, with values of correlation varying from XXX for subject XXXX to XXX for subject XXX (see Coe ??, also see discussion below). Individual level, group/class level, school level and subject level and departmental level value added analysis can be performed using a software system supplied to the schools by the CEM Centre.

Comparison of value added between the regression model and the principal axis model

In this study, the Yellis baseline test data collected for nearly 170,000 students from 1093 secondary schools in the UK in 2000 and their subsequent GCSE Maths grades data collected in 2002 have been used for analysis to demonstrate the differences in value added calculation using the PA approach and the conventional regression approach discussed above. Figure 2 shows the frequency distributions of the stdeunts' Yellis scores and their GCSE mathematics grades. As can be seen, the Yellis scores are distributed symmetrically, while the GCSE mathematics grades are slightly left-skewed. The reliabilities of the Yellis vocabulary test, the mathematics test and the overall test are relatively high, with values of Cronbach's alpha being XXX, XXX and XXX respectively. Although the reliability measures of the GCSE mathematics scores and grades were not available for this study, they were expected to be similar to those of the Yellis test.

<Figure 2>

To make the value added calculation simple, the GCSE grades are converted into a point scale with values varying from 0 (corresponding to GCSE's U grade) to 8 (corresponding to GCSE's A* grade). As Yellis results and GCSE results are reported using different scales, both the Yellis scores and the GCSE Maths point scores have been standardised to have a mean of 100 and a deviation of 15 in this study. Since the standardisation represents a linear transformation, the relationship between the two variables will not be affected. As indicated earlier, the important advantage of using standardised test scores for analysis is that the two measures can be compared directly.

<Figure 3>

Figure 3 shows the distribution of the standardised GCSE Maths point scores obtained in 2002 against the standardised Yellis baseline scores for students collected in 2000. Correlation analysis indicates a coefficient of 0.76, suggesting that about 58% of the variation in one variable is associated with the variation of the other. It could therefore be assumed that 58% of the variation in the GCSE maths point scores can be explained by the variation in the Yellis baseline scores and the remaining 42% of the variation could be assumed to be related to the influence of the schools and other external factors.

A linear regression of the standardised GCSE Maths point score ($Y_{GCSE,Maths}$) on the standardised Yellis score (X_{Yellis}) suggests the following relationship between the two test scores:

$$Y_{GCSE,Maths} = 0.756 X_{Yellis} + 24.4 \tag{9}$$

With R^2=0.58. However, if an inverse regression (i.e. regressing X_{Yellis} on $Y_{GCSE,Maths}$) is conducted, we will have:

$$X_{Yellis} = 0.756Y_{GCSE,Maths} + 24.4 \tag{10}$$

Therefore, in this case, the relationship between the GCSE Maths point scores ($Y'_{GCSE,Maths}$) and Yellis scores (X'_{Yellis}) becomes:

$$Y'_{GCSE,Maths} = 1.322X'_{Yellis} - 32.3 \tag{11}$$

The two regression lines are also plotted in Figure 2. It is clear that the slopes as well as the intercepts for the two regression lines are different. The residuals calculated from Equations 9 and 11 will therefore be different for each individual students and the average of aggregated residuals for a class or for a school. Although the two regression lines pass through the centre point of the dataset, they do not represent the symmetric lines of the dataset. As can be seen, the choice of different conventional regression models results in different value added, and this has important implication in terms of providing objective and fair evaluation of the performance of individual students, classes/groups and schools.

To compare the conventional OLS regression (regressing GCSE Maths scores *Y* on Yellis scores *X*) and PA value added calculation methods, the residuals for an individual student *i* are calculated as follows:

$$\begin{aligned} R_i &= Y_i - (0.756X_i + 24.4) \quad &\text{for OLS regression} \\ R_{i.pa} &= Y_i - X_i \quad &\text{for \quad PA} \end{aligned} \tag{12}$$

<Figure 4>

Figure 4 depicts the distribution of the standardised residuals (with a mean of 0 and a deviation of 1.0) calculated using the two approaches for individual students against their standardised Yellis scores. It is clear from the figure that the residuals for students with Yellis scores close to the average Yellis score for the population (100) are similar for both methods, but those for students with high Yellis scores have been reduced while those for students with low Yellis scores have been raised by the PA method in relation to the values calculated using the regression method. The magnitudes of the increase or reduction depend on the difference between the Yellis scores and the average Yellis score of 100 or the departure from the Yellis average. The maximum change in residuals between the two methods occurs in the regions of relatively low or high Yellis scores and is close to 1.0 which is the standard deviation of the standardised residuals derived from the individual methods. As can be seen from Figure 4, the PA derived residuals represent a clockwise rotation of the regression derived residuals around the mean of the standardised Yellis scores, reflecting the fact that the regression line represents a clockwise rotation of the major principal axis discussed previously.

The difference in residuals for individuals between the two methods will also affect the average aggregated value added for a group/class or the entire school. In some schools in England, students are placed in different sets according to their academic abilities (Ireson et al., 2002). The conventional regression method will therefore produce higher average value added for groups with high ability and lower average value added for groups with low ability. This increase in value added has been attributed to educational mechanisms by some researchers (see, for example, Harker and Tymms, 2004 for an exploration of different interpretations). For example, some suggest that students in a high ability group will influence

each other positively in their study and will therefore make greater progress than students in lower ability groups. As can be seen the compositional effect could be an artefact, resulting from the use of the conventional regression method, and the PA method can therefore remove or at least reduce its magnitude.

Figure 5 further illustrates the relationship between the residuals for individual students derived using the two methods. Although the two methods give different results, the values are still highly correlated (R^2=0.88). Overall, 13.7% of the students have the sign of their regression-based residuals changed in relation to PA-based residuals. About 8.0% of these students have their positive regression-based residuals changed to negative PA-based residuals, while 5.7% of them have their negative regression-based residuals changed to positive PA-based residuals.

<Figure 5>

Figure 6 shows the distributions of the average standardised residuals for 1093 schools used in this study estimated using the two methods with their corresponding school average Yellis scores (calculated as the mean of all the students taking both the Yellis test and the GCSE exams). Again, it is clear that the standardised residuals are similar for schools with school average Yellis scores near the population Yellis average score of 100. However, the average aggregated standardised residuals calculated using the PA method for schools with high average Yellis scores have been reduced relative to the residuals calculated from the regression method, while those for schools with low average Yellis scores have been increased. As can be seen from Figure 6, the average standardised residuals for schools derived using the PA method represents again a clockwise rotation of those derived using the conventional regression method around the mean standardised Yellis score. However, the size

of the rotation is smaller in comparison with that for individual students shown in Figure 5 as the school value added is calculated as the mean of the value added for individual students. The value added for a proportion of the schools with low Yellis scores and negative residuals in the regression method has changed to positive residuals in the PA method, while the value added for a proportion of schools with high Yellis scores and positive residuals has changed from positive to negative. The maximum change in residuals between the two methods is about 0.67, which is about two thirds of the standard deviation of the standardised residuals from the two methods. Figure 7 further depicts the relationship between the average standardised residuals for these schools estimated using the two different approaches. A correlation analysis suggests a coefficient value of 0.86 between the residuals derived using the two different approaches. Overall, about 15.5% of the schools have the sign of their regression-based average school residuals changed in comparison with the PA-based average residuals. About 8.0% of the schools have their negative regression-based average school value added changed to positive PA-based average value added, while 7.5% of them have their positive regression-based average value added changed to negative PA-based average value added.

<Figure 6>

<Figure 7>

Figure 7 shows the distribution of the standardised GCSE maths point scores for students against their standardised Yellis scores in a specific school selected for this study. Both the ordinary regression line and the PA line for the overall population are superimposed in the figure. Figure 8 shows the corresponding distribution of the standardised residuals for individual students from the two methods against their standardised Yellis scores. Again, the

results from PA represent a clockwise rotation of the results from the conventional regression method. The value added calculated using PA for students with high Yellis scores is substantially reduced in comparison with that calculated using the regression method, and a proportion of these students have their positive regression derived residuals changed to negative PA derived residuals. In contrast, the value added calculated using PA for those with low Yellis score is considerably increased in comparison with the regression derived value added. This has important implications in terms of assessing class teachers using the overall aggregated residuals of individual students in a class if the students are grouped into sets of different academic abilities. The regression method will favour groups with higher abilities while disadvantage groups with low abilities.

<Figure 7>

<Figure 8>

Discussion and Conclusion

Many test-based accountability systems use value added analysis technique involving the comparison of students' prior achievement and outcome achievement to produce performance measures that are used to judge the quality of learning of individual students and the quality of teaching of class teachers and schools. These value added performance measures should be estimated as accurate as possible. Specifically, the contribution to the progress made by a student associated with the level of achievement before entering the school has to be accurately estimated.

The OLS regression models have been frequently used to represent the relationship between students' present and prior performance measures and used to calculate value added. It has been demonstrated in this study that the use of the OLS regression method to calculate

value added is likely to disadvantage students with lower prior performance while give advantages to students with higher prior performance. This is primarily due to minimising the squares of the departures of the present performance measures from the population averages when estimating the regression model parameters. The OLS regression method assumes that only the present performance measures have errors and are treated as a random variable. However, such assumption is not likely to be met by any test data. The value added calculated based on the conventional regression model will be overestimated for students with high prior achievement but underestimated for students with low prior achievement. At group or class level, teachers teaching groups of low ability students will be disadvantaged while those teaching groups of high ability students will gain unfair advantage. At school level, the conventional OLS regression method will favour schools with intakes represented by high prior performance measures and disadvantage schools with intakes represented by low prior performance measures.

The principal axis (PA) approach on the other hand, treats both present and prior performance measures equally as random variables and uses the major principal axis (the major symmetrical line of the dataset) to represent the relationship between the two performance measures. This would be a more appropriate representation of the relationship between the two variables as both performance measures are likely to have similar magnitude of errors and there is no need to assume a cause-effect relationship between the two variables since both the prior and current performances reflect the level of the underlying ability of the student. The PA method increases the value added for low prior performing students while decreases the value added for high prior performing students in comparison with the OLS regression approach. It therefore removes the limitations associated with the OLS regression method that unfairly disadvantage students with low prior achievement and gives advantages to students with high prior achievement. At group or class level, the PA approach will also

provide fairer value added for teachers of all ability groups than the regression method which gives unfair advantage to groups of high ability students. At school level, the PA approach will also produce fairer value added for schools than the OLS regression approach which unfairly disadvantage schools with intakes of low prior performing students while give advantages to schools with intakes of high prior performing students.

Findings from the present study clearly indicated that the value added measures estimated using the PA approach can be considerably different from those estimated based on the conventional regression approach, particular at both the bottom and top of the prior achievement range. Results from this study has important implications in terms of conducting fairer comparison of the relative learning progression between individual students, the quality of teaching of class teachers and schools when using value added as a performance indicator. The appropriateness of the model used for calculating value added performance measures to a large extent depends on the various assumptions of the model that are met by the observed data. It has been argued in this paper that the PA approach is more appropriate than the conventional regression model when comparing test scores and is likely to produce fairer value added for students, teachers and schools.

Conceptually, the PA approach is more acceptable and easier to understand than the conventional regression approach as it uses the major symmetry line to represent the relationship between the prior and outcome achievement measures. Technically, when both prior and present achievement measures are standardised in the same fashion, the PA approach is also simpler to implement and would be easier to understand by teachers.

It is recognised that in this study, it has been assumed that the value added calculated using both the conventional regression method and the PA approach is school-related and the contribution from other factors such as students' demographics has been ignored. Work by Fitz-Gibbon (1997) suggested that variation in students' social indicators (including SES)

only accounts for a small proportion of variation in achievement measures and contributes little to the prediction of value added at individual student level. She suggests that students' potential cannot be accurately judged by their home background. However, it is suggested that further work is required to compare the PA-derived value added measures with those derived from models that incorporate other factors such as students' SES for individual students, classes and schools in order to explore further the impact of different models on the calculation of test score based value added performance measures.

Disclaimer The views expressed in this paper are those of the authors and are not to be taken as the views of the Office of Qualifications and Examinations Regulation (Ofqual).

Notes on Contributors

Dr Qingping He is a Principal Researcher at the Office of Qualifications and Examinations Regulation (Ofqual). His research interests include assessment development, assessment reliability and validity, test equating, item banking, computer-based testing (including computer adaptive testing), and qualification standards.

Professor Peter Tymms is Head of Department and Chair of the Board of Studies in the School of Education at Durham University. His main research interests include monitoring, assessment, performance indicators, ADHD, reading and research methodology. He devised the PIPS project, which is designed to monitor the affective and cognitive progress of children through primary schools starting with a computer adaptive on-entry baseline assessment. He was Director of the CEM Centre at Durham University until 2011.

References

Boggs, P. and Rogers, J. (1990) Orthogonal Distance Regression. In *Contemporary Mathematics, Volume 112: Statistical Analysis of Measurement Error Models and their Applications* (ed. P. J. Brown and W. A. Fuller), 183-194. American Mathematical Society, Providence: Rhode Island.

Bryk, A. S. and Raudenbush, S. W. (1992) *Hierarchical Linear Models*. Sage: Beverly Hills, CA.

Bryk, A., Raudenbush, S. and Congdon, R. (1996) *Hierarchical Linear and Nonliear Modelling with HLM/2L and HLM/3L Programs*. SSI Scientific Software International: Chicago.

Coe, R. (2013) Improving education: A triumph of hope over experience. Inaugural lecture, Durham University. Available online at: http://www.cem.org/attachments/publications/ImprovingEducation2013.pdf.

Crane, J. (2002) The promise of value-added testing. Progressive Policy Institute Publications, Policy Report (November 2002). Available online at: http://www.ppionline.org/ppi_ci.cfm?knlgAreaID=110&subsecID=136&contentID=251035.

Demie, F. (2003) Using Value-added Data for School Self-evaluation: a case study of practice in inner-city schools. *School Leadership & Management* 23, 445–467.

Depart for Education (DfE) (2010) Developing data-literate school leaders RAISEonline: an essential tool for school improvement Delegate's handbook – Secondary. Available online at: http://dera.ioe.ac.uk/718/1/d33459b52a557dc676ae77764994a778.pdf

Depart for Education (DfE) (2011) Value added key stage 2 to 4 in 2011 school & college performance tables & raise online. Available online at: http://www.education.gov.uk/schools/performance/secondary_11/KS2-4_General_VA_Guide_2011_FINAL_AMENDED.pdf

Fitz-Gibbon, C.T. (1996) *Monitoring Education – Indicators, Quality and Effectiveness*. Continuum: London.

Fitz-Gibbon, C.T. (1997) *The Value Added National Project: Final Report. Feasibility studies for a national system of value added indicator*. School Curriculum and Assessment Authority: London.

Friendly, M., Monette, G. and Fox, J. (2013) Elliptical insights: understanding statistical methods through elliptical geometry. *Statistical Science* 28, 140.

Goldstein, H. (1997) Methods in school effectiveness research. *School Effectiveness and School Improvement* 8, 369-395

Goldstein, H. (2003) *Multilevel Statistical Models* (3rd edition). Arnold: London.

Gorard, S. (2006) Value-added is of little value. *Journal of Education Policy* 21, 235-243.

Hamilton, L., Stecher, B. and Klein, S. (editors) (2002) *Making Sense of Test-Based Accountability in Education*. Rand: Santa Monica, CA.

Harker, R. and Tymms, P. (2004) The Effects of Student Composition on School Outcomes. *School Effectiveness and School Improvement* 15, 177-199.

Hauser, R. M. (1970) Context and consex: a cautionary tale. *American Journal of Sociology* 75, 645-664.

Hauser, R. M. (1974) Contextual analysis revisited. *Sociological Methods & Research* 2, 365-375.

Hout, M. and Elliott, S. (2011) Incentives and Test-Based Accountability in Education. Washington, DC: national Research Council and the National Academies Press. Available online at: http://jrnetsolserver.shorensteincente.netdna-cdn.com/wp-content/uploads/2011/07/Test-Incentives-Education-NAS-report1.pdf

He, Q., Hayes, M. and Wiliam, D. (2013) Classification accuracy in Key Stage 2 National Curriculum tests in England. *Research Papers in Education* 28, 22-42.

He, Q. and Opposs, D. (2012) The reliability of results from national tests, public examinations and vocational qualifications in England. *Educational Research and Evaluation* 18, 779-799.

Ireson, J., Clark, H. and Hallam, S. (2002) Constructing ability groups in the secondary school: issues in practice. *School Leadership and Management* 22, 163-176.

Isenberg, I. and Hock, H. (2011) Design of Value-Added Models for IMPACT and TEAM in DC Public Schools, 2010–2011 School Year: Final Report. Available online at: http://dcps.dc.gov/DCPS/Files/downloads/In-the-Classroom/Design%20of%20Value-Added%20Models%20for%20DCPS%202010-2011.pdf

Isobe, T., Feigelson, E., Akritas, G. and Babu, G. (1990) Linear regression in astronomy. I. *The Astrophysical Journal* 364, 104-113.

Jolliffe, I. (2002) Principal Component Analysis (2nd ed). Springer: New York.

Krzanowski, W. (1998) *An introduction to statistical modeling*. Arnold: New York.

Lissitz, R. (2006) *Longitudinal and Value Added Models of Student Performance*. JAM Press: Maple Grove, MN.

McCaffrey, D., Lockwood, J. Koretz, D. and Hamilton, L. (2004) *Evaluating Value-Added Models for Teacher Accountability*. RAND Corporation: Pittsburgh, PA.

Preisendorfer, R. (1988) *Principal Component Analysis in Meteorology and Oceanography*. Elsevier: Amsterdam, The Netherlands.

Raudenbush, S. W. and J. D. Willms (1991) *Schools, Classrooms, and Pupils: international studies of schooling from a multilevel perspective*. Academic Press Inc: New York.

Ray, A. (2006) *School Value Added Measures in England: A Paper for the OECD Project on the Development of Value-Added Models in Education Systems*. London: Department for Education and Skills.

Sanders, W., Saxton, A. and Horn, S. (1997) The Tennessee Value-Added Assessment System (TVAAS): A quantitative, outcome-based approach to educational assessment. In: *Grading Teachers, Grading Schools. Is Student Achievement a Valid Evaluation Measure*? (ed. J. Millman), 137-162. Corwin Press: CA, USA.

Sanders, W. L., and Rivers, J. C. (1996) *Cumulative and Residual Effects of Teachers on Future Student Academic Achievement*. Knoxville: University of Tennessee Value-Added Research and Assessment Center: TN, USA.

Schagen, I. and Schagen, S. (2003) Analysis of National Value-added Datasets to Estimate the Impact of Specialist Schools on Pupil Performance. *Educational Studies* 29, 3-18.

Schagen, I. and Schagen S. (2005) Combining multilevel analysis with national value-added data sets-a case study to explore the effects of school diversity. *British Educational Research Journal* 31, 309-328.

Taylor, J. and Nguyen, A. (2006) An Analysis of the Value Added by Secondary Schools in England: Is the Value Added Indicator of Any Value? *Oxford Bulletin of Economics and Statistics* 68, 204-224.

Timmermans, A., Doolaard, S. and de Wolf, I. (2011) Conceptual and empirical differences among various value-added models for accountability, *School Effectiveness and School Improvement* 22, 393-413.

Tymms, P.B. (2002) *Baseline Assessment and Monitoring in Primary Schools – Achievements, Attitudes and Value-added Indicators*. David Fulton Publishers: London, England.

Tymms, P. and Coe, R. (2003) Celebration of the Success of Distributed Research with Schools: the CEM Centre, Durham. *British Educational Research Journal* 29, 639-653.

Warton, D., Wright, I., Falster, D. and Westoby, M. (2006) Bivariate line-fitting methods for allometry. *Biological Reviews* 81,259–291.

Wei, H., Hembry, T., Murphy, D. and McBride, Y. (2012) Value-Added Models in the Evaluation of Teacher Effectiveness: A Comparison of Models and Outcomes. Available online at: http://www.pearsonassessments.com/hai/images/tmrs/ComparisonofValue-AddedModelsandOutcomes.pdf

West, A. (2010). High stakes testing, accountability, incentives and consequences in English schools. *Policy & Politics* 38, 23-39.

West, A., P. Mattei; & J. Roberts (2011). Accountability and Sanctions in English Schools, *British Journal of Educational Studies* 59: 1, 41 – 62.

Wilson, D. (2004) Which Ranking? The Impact of a 'Value Added' Measure of Secondary School Performance. *Public Money and Management* 24, 37-45.

Wiliam, D. (2010) Standardized testing and accountability. *Educational Psychologist* 45, 107-122.

Willms, J. D. (1992) *Monitoring School Performance: A Guide for Educators*. Lewes, Falmer Press.

Yang, M., Goldstein, H., Rath, T. and Hill, N. (1999) The use of assessment data for school improvement purposes. *Oxford Review of Education* 25, 469-483.
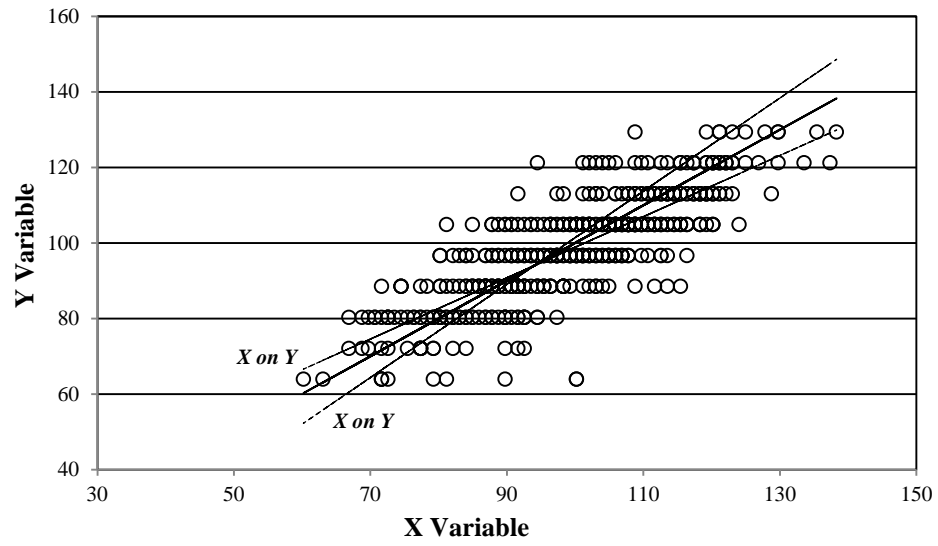
Figure Captions

Figure 1 The regression lines, a line of symmetry and an ellipse for a bivariate dataset.
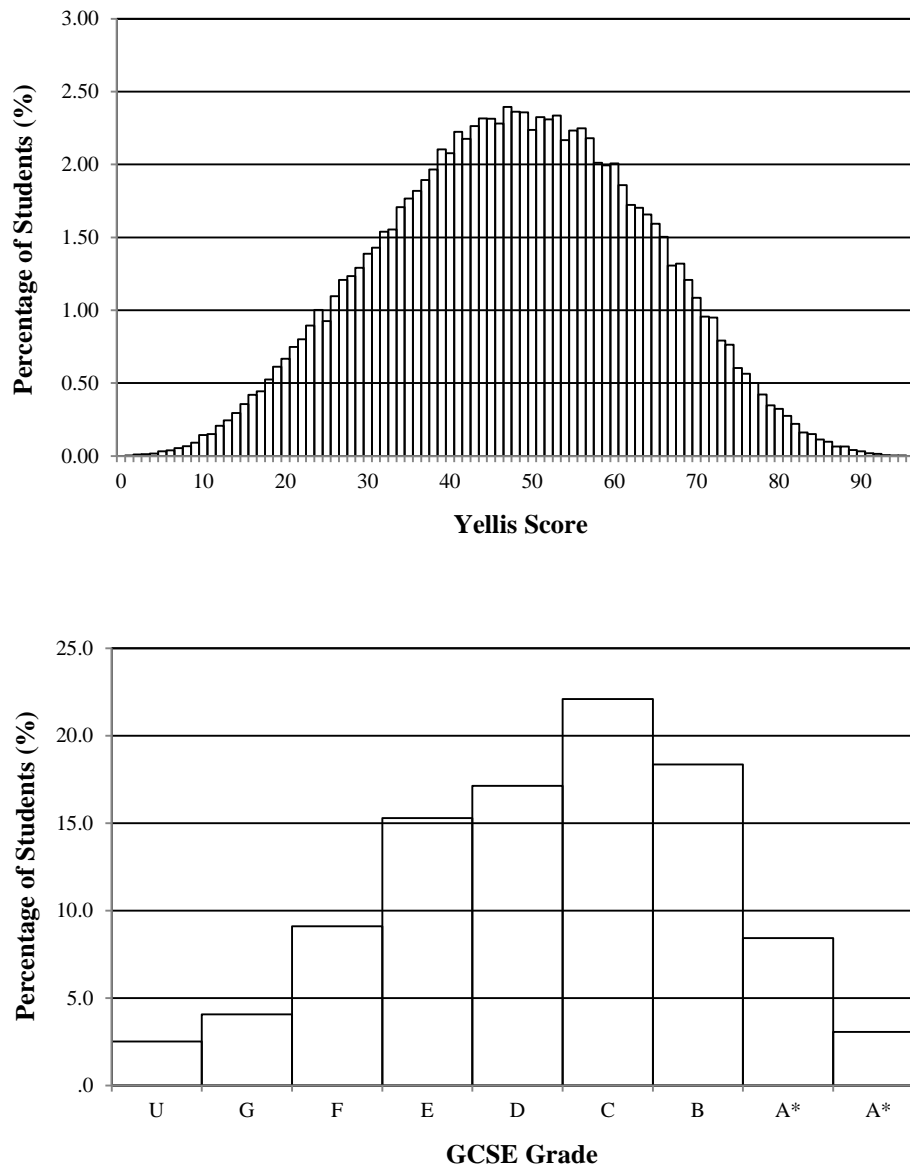
Figure 2 The distributions of Yellis scores (top) in 2000 for students involved in the Yellis project and their GCSE Maths grades in 2002.
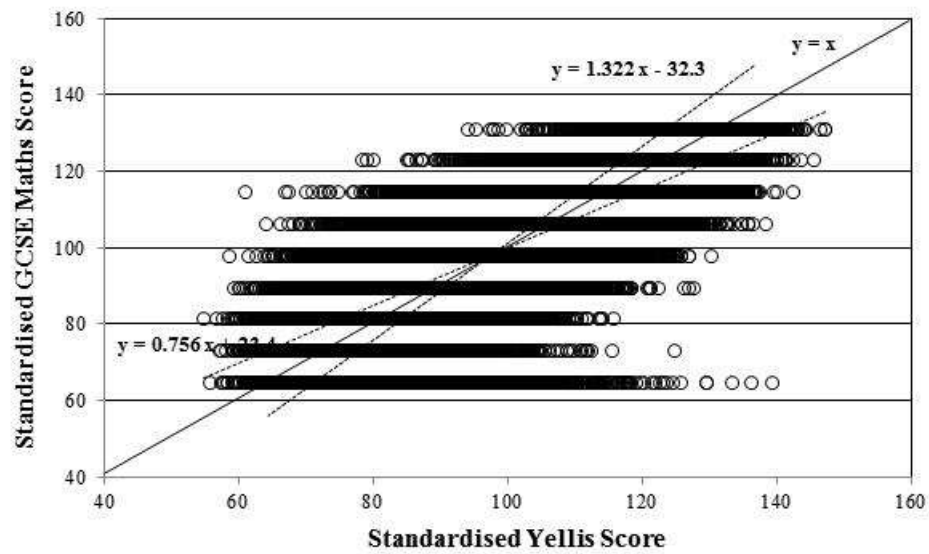
Figure 3 Relationship between standardised GCSE Maths point scores and Yellis baseline test scores.
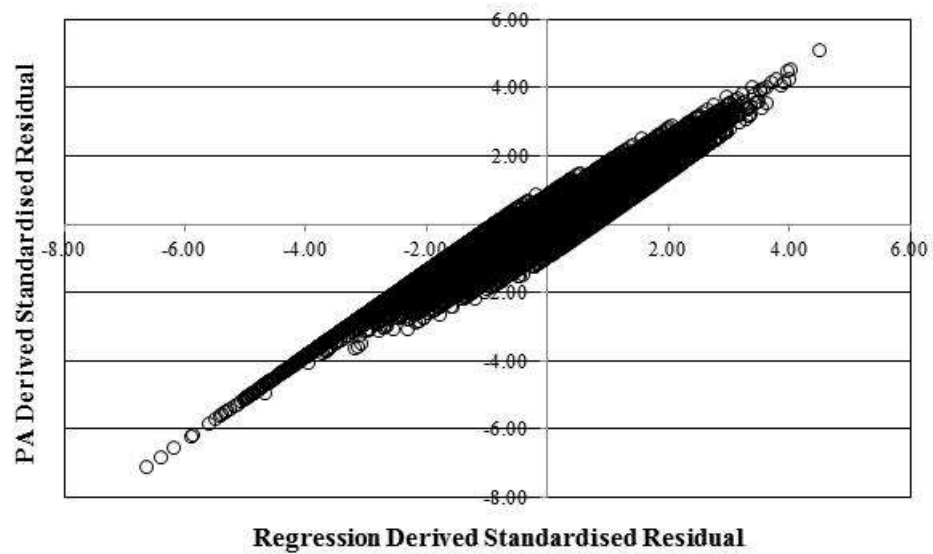
Figure 4 Distribution of standardised residuals for individual students calculated using the conventional regression method and the PA method against Yellis scores.
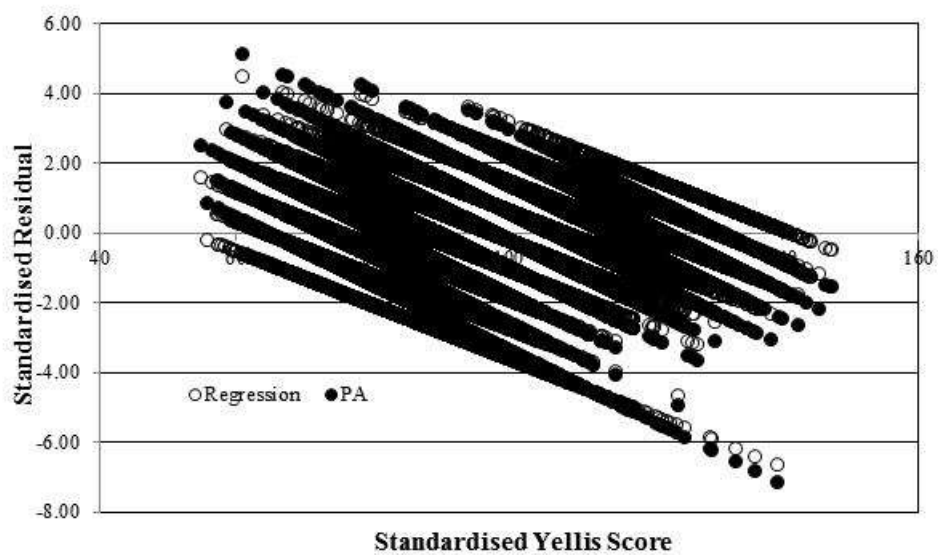
Figure 5 Relationship between standardised residuals for individual students calculated using the conventional regression method and the PA method.

Figure 6 Distribution of standardised residuals for schools calculated using the conventional regression method and the PA method against Yellis scores.

Figure 7 Relationship between standardised residuals for schools calculated using the conventional regression method and the PA method.

Figure 8 Distribution of standardised GCSE Maths scores and Yellis scores for students in a specific school.

Figure 9 Distribution of standardised residuals for students in a specific school calculated using the conventional regression method and the PA method against Yellis scores.

Figure 1 The regression lines, a line of symmetry and an ellipse for a bivariate dataset.

Figure 2 The distributions of Yellis scores (top) in 2000 for students involved in the Yellis project and their GCSE Maths grades in 2002.

Figure 3 Relationship between standardised GCSE Maths point scores and Yellis baseline test scores.

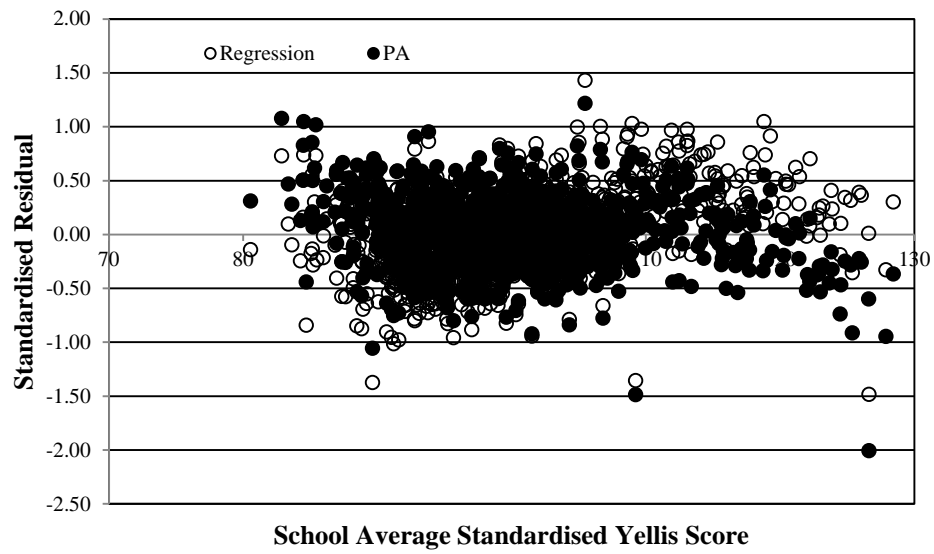**Figure 4** Distribution of standardised residuals for individual students calculated using conventional regression method and PA method against Yellis scores.

Figure 5 Relationship between standardised residuals for individual students calculated using conventional regression method and PA method.

Figure 6 Distribution of standardised residuals for schools calculated using conventional regression method and PA method against Yellis scores.
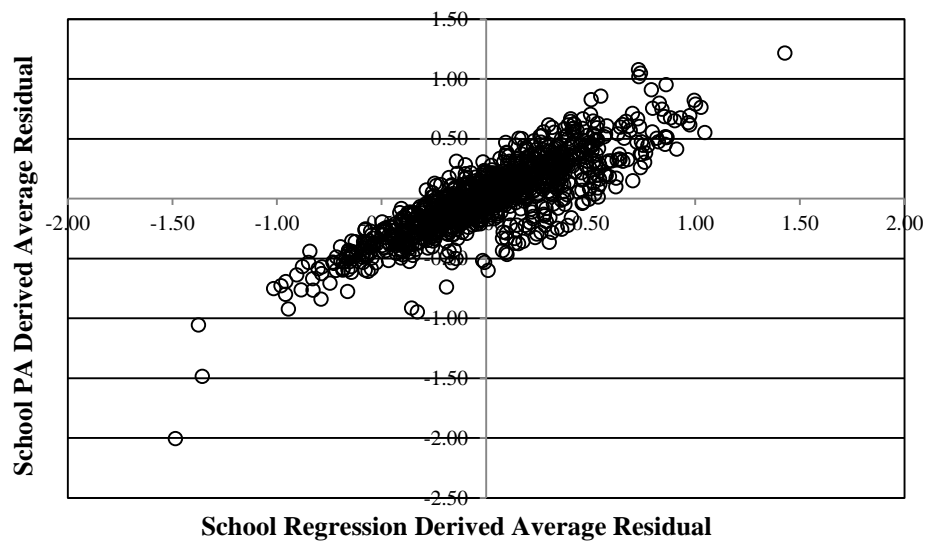

Figure 7 Relationship between standardised residuals for schools calculated using conventional regression method and PA method.
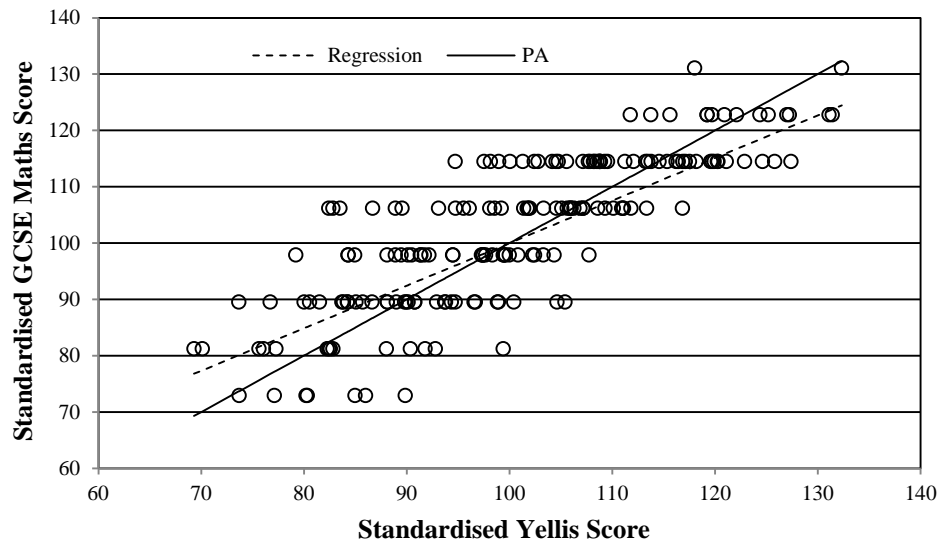
Figure 8 Distribution of standardised GCSE Maths scores and Yellis scores for students in a specific school.
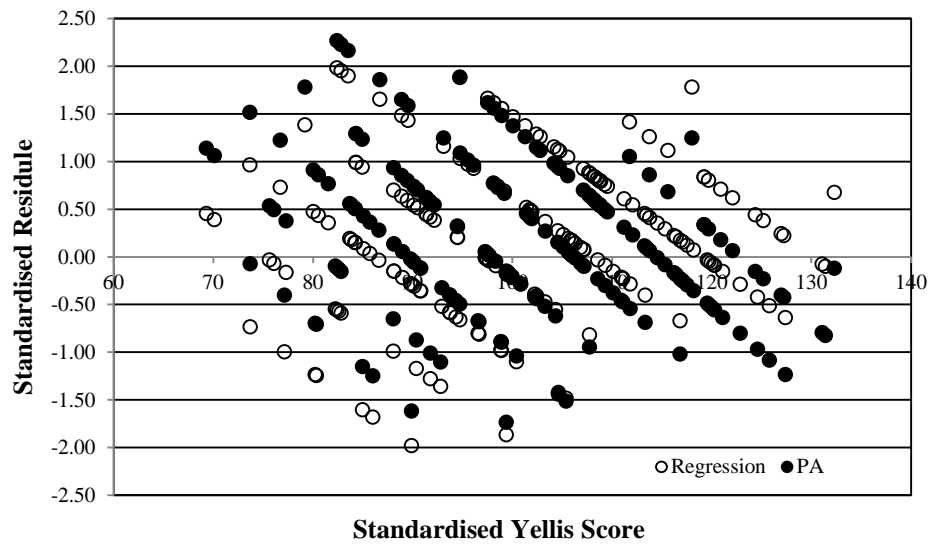

Figure 9 Distribution of standardised residuals for students in a specific school calculated using conventional regression method and PA method against Yellis scores.