2005

Using HSC data to give principals leverage

John DeCourcy
St Andrew's College (NSW)

Using HSC data to give principals leverage

John DeCourcy

St Andrews College, New South Wales

Dr John DeCourcy is College Principal of St Andrews College, a Catholic dual-campus secondary school of 1200 students in Western Sydney. His academic background is in theoretical structural chemistry, educational measurement and developmental psychology. Five years ago, seeking a better way of understanding and using achievement data from his own school, he began the project with the Catholic Education Commission (NSW) providing a multilevel analysis of HSC data described in this paper, and has gone on to provide this analysis to all Catholic secondary schools in the State. John is also currently president of the Australian College of Educators in his part of Sydney.

> 'Without data, I'm just another person with an opinion'
> (Barry McGaw, ACER Research Conference 2002)

What makes the difference in student achievement? What elements among 'what makes the difference' can a school principal influence? How does the principal influence these for the better? How do teachers best take account of the pedagogical information available in data? How can we use the data available to address these questions?

The topic of this conference is 'Using data to improve student learning'. Data will act to improve student learning broadly across a school only if the data become the principal's agenda, and the data will become the principal's agenda only if s/he sees the data as a useful lever to achieve worthwhile outcomes. So the question becomes: 'How do you get data to a form where it will provide the principal with leverage s/he can use and trust?' This paper draws on a five-year project involving over 120 secondary schools in New South Wales to outline what has been learned about the most effective ways to engage principals and teachers with a particular set of achievement data. The learnings from this project may well have applicability in other settings.

Principals and teachers can be reluctant to engage with data because their professional intuition leads them to be defensive about data analysis which purports to attribute large differences in achievement to schools or teachers, where the difference actually lies in factors beyond their control (O'Day, 2002). What is needed is a form of analysis that separates out the factors that do lie within the control of teachers, and gives a valid and easily interpreted analysis of these factors.

Visscher and Coe (2002) develop a heuristic for the interpretation of School Performance Feedback Systems (SPFS) which looks at the system in terms of its:

- design process
- features (the validity of the input information, the accessibility of the data, whether the output is standard or tailored to the school, the extent of support for use of the system, etc.)
- implementation process (the use of tailored user training, promotion of user participation, the monitoring of implementation, etc.)
- within-school organisational features (the school's and teachers' capacity to deal with innovation, the extent to which the system requires resources, the extent to which new skills must be developed, etc.).

Each of these four aspects of the system bear upon the fifth and critical aspect: the usage of the SPFS (whether it will be for instrumental, conceptual, symbolic, or strategic use). The choice of dominant usage pattern then affects the sixth characteristic of the system, its intended and unintended effects.

There are many examples of SPFS where failure to take adequate notice of the features, implementation, or organisational characteristics of the system leads to utterly inappropriate usage of the system, and undesirable unintended effects (Amrein & Berliner, 2003; Braun and Mislevy, 2005). The intention of the project described in this paper is to produce a usage pattern that is instrumental: the data becomes an instrument in the principal's and teachers' hands to monitor and improve pedagogy and students' performance.

As an instrument, the data is presented in a way that gives the principal leverage to support and effect innovation that has a positive effect on student achievement.

Context

For each of the last five years, the project has been a cooperative agreement, conducted under the auspices of the Catholic Education Commission (CECNSW) between the (now) 125 Catholic secondary schools of New South Wales to pool the results of their 14,000+ students in the Higher School Certificate (HSC) examinations to enable a multilevel analysis (Goldstein, 1995; Goldstein, Rasbash, Plewis, Draper, Browne, Yang, Woodhouse, & Healy, 1998) to be conducted across both the aggregate results and each of the 80,000+ results in individual subjects. The statistical methodology of the analysis is described in the Appendix to this paper.

The central concept of the project is 'comparative learning gain': what is the comparison in the performance of the students in this subject in this school with that of similar students in other schools, where 'similar' is taken as students of equivalent prior achievement two years earlier in the School Certificate, of the same gender and of the same socioeconomic status (SES). For the teacher in the HSC course, each of prior achievement, gender and SES is a given, each is liable to have a bearing on achievement, and each must be discounted if pedagogical effects are to be inferred.

A second important aspect of the analysis has been the inclusion of confidence intervals (uncertainties) in the graphical presentation of results. An apparent improvement of 2% in average achievement is not significant if the confidence intervals of the measurement are +/– 9%!

The product of the project as supplied to schools is an electronic file, consisting of five parts:

- The *Primary Analysis* of each subject, showing a comparative learning gain (with confidence intervals) of the mean result achieved in the subject with that achieved by similar students in other schools
- The *Secondary Analysis* of each subject, showing a comparison of the mean result achieved in this subject with firstly state average and secondly the average obtained in all of their other subjects by the students in this subject
- The *Trends Analysis* for each subject, showing the three measures from the primary and secondary analyses over the last six years, and showing any second-order effects for each year
- The *School Database* containing both the input data and the results of the analysis for each student in each subject, along with aggregations at the student, subject and school level. The database in particular allows for further investigation of the student- and class-level information
- The *Report* (DeCourcy, 2005b) to CECNSW on the performance of Catholic schools generally in the HSC, any issues arising from the analysis and a series of statistical appendices.

The process for delivery of the analysis to schools is centred on supporting the principal in his/her work with staff. Students and schools receive the results of the HSC in mid-December each year; the analysis of these results from the project is available for downloading before the start of the following school year, and the report on the overall HSC is available from June each year. The project is supported by a web site (http://stage.cecnsw.catholic.edu.au/hsca/) which has both a secure section where schools and systems can obtain their own data, and an open section containing the *Manual* (DeCourcy, 2005a) for the project, and a series of annotated PowerPoint files which can be used by principals and others in professional development activities with staff. Each year, a number of seminars on the use of the analysis are conducted under the auspices of CECNSW for those whose role it is to introduce the analysis to staff.

Initially, most principals met this project with a healthy degree of scepticism and suspicion; over the five years of the project, this has changed for most to insight and enthusiasm as they have seen the connection between the presentation of the data and their knowledge of their schools.

What we've learned

We've learned (Rowe, 2000, 2001, 2004a) that it's teachers who make the difference; whole-school effects are small compared to the effect of individual teachers. Multilevel analysis with all variables converted to normal-equivalent deviates as described in the appendix partitions the variance sources for student aggregate Tertiary Entrance Score (TES); a similar process can be undertaken for a subject such as Drama.

The contrast between the school effect in these two analyses is not surprising. For a TES, students will have experienced at least five different teachers, and usually

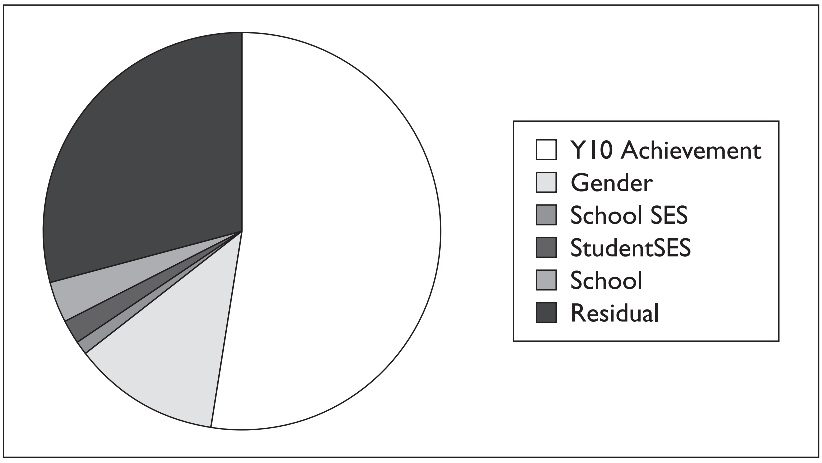

Figure 1 Sources of variance in 2004 TES

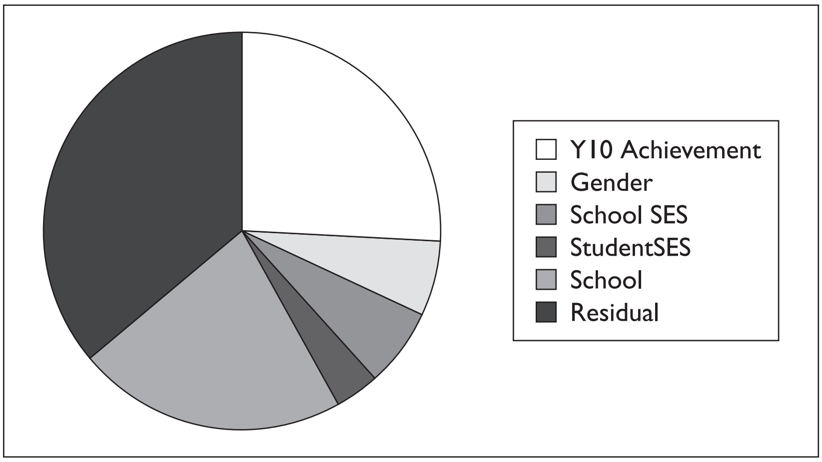

Figure 2 Sources of variance in 2004 HSC Drama

six or seven. The effect seen is an average across all of these subjects/teachers. Put differently, the data point to a consistent mix of teaching and pedagogy experienced by students. The point of leverage for principals is to see those subjects and teachers where the comparative learning gain is high and to build on these strengths; similarly, to see those where it is low and target appropriate interventions.

We've learned that in order to engage principals and teachers with data, you need to begin with the assumptions they make about data, and unpack these. When previously the only standard for comparison for schools was with state average, or with the school's previous results, there are predictable responses to results above average or those below average. Those above were greeted with, 'Haven't we done well!' Those below were dismissed with, 'They weren't a very good group this year'. Both of these responses rely on assumptions of the comparison of achievement with expectation. The 'Haven't we done well' response is a claim that compared to what might reasonably have been expected of this group of students, they have done better than expectation. 'They weren't a very good group' implies that expectations should have been low, and that achievement is in line with expectation. Both responses beg the question of an appropriate level of expectation, which can be addressed using multilevel modelling.

We've learned that most practitioners are engaged with the data not through a consideration of the analytic techniques as summarised in the Appendix, but through the use of a valid graphical presentation of the results of that analysis. For each subject in a school, the Primary Analysis is simply presented as a comparison of 'Achieved' with 'Expected', building on the unpacking of assumptions describing above, showing confidence intervals.

Learning to interpret a graph such as this is the focus of the seminar program and the manual. The diagonal line where achieved equals typical is the line of average comparative learning gain in this subject. The centre of the ellipse is the value that this subject in this school achieves as an average achieved standard score, against the average typical standard score as outlined in equation (13) in the Appendix. The axes of the ellipse are determined by the confidence intervals of the means, derived as outlined in the Appendix. When the ellipse is completely above the diagonal as in this case, the achieved result is above what students of the same prior achievement, gender and SES have achieved elsewhere. When the ellipse intersects the diagonal, it is 'in the range of expectation'. When it is completely below, it is 'below expectation'. In the case illustrated above, the principal and the teacher can indeed be confident that 'we have done well' even though the results may have been below state average.

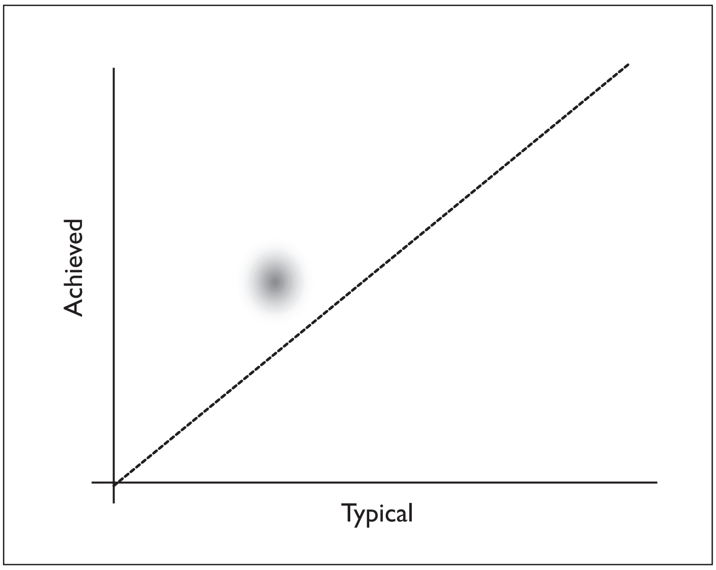

Figure 3 Typical primary analysis graph for a subject

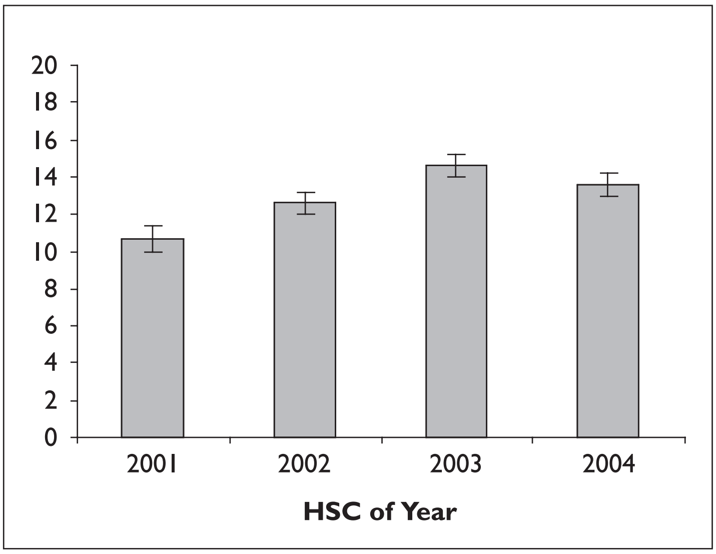

Figure 4 Percentage of variance in learning gain related to gender

We've learned that gender and SES do make a difference in results, but they are not variables which schools can change. The approach therefore has been to account for the variable, discount it (by factoring it into the typical or expected score as shown in equation (13)), and look at the pedagogy.

Gender is related to 10–14% of the variance in TES, favouring girls. The issues relating to appropriately differentiated pedagogy, enabling both boys and girls to engage with the curriculum at their point of need and learning style are considerable. In 2004, 31 subjects showed significant gender effects, with 30 of these favouring females. The size of the significant effects ranged from 2.3% (Mathematics) to 16.7% (Food Technology). The longitudinal data from the project give principals a basis for seeing whether the curriculum and pedagogical interventions they apply are having an effect.

We've learned that SES is related to only a tiny proportion of the variance in aggregate results (as shown in Fig.1), but it may be a bit larger in some individual subjects. There has been criticism (Marks, Rowe & Beavis, 2003; Rowe, 2004) of some analyses of achievement data which purport to show large SES effects but are in fact statistically invalid. The 2004 analysis in this project shows 1.1% of the TES variance related to variance in the school-level Farish index (Farish, 2004) and 1.9% related to variance in the postcode-average for the individual student (Australian Bureau of Statistics, 2004).

We've learned that the real test of the validity and utility of a data analysis for a principal lies in his/her ability to recognise in the graphical representation of the subject what s/he knows of what has happened within the school. For the first three years of the analysis, there was simply a single-year snapshot of data. When the data was summarised over time in a trends graph, principals began in a large way to engage with the data. Fig. 5 shows on particular school's trend on the primary (comparative learning gain) measure.

When the principal saw this, he immediately identified the reasons for the drop in 2001 from what he knew of what had happened in that subject in the school, and was convinced of the validity of the data analysis process.

We've learned that engagement with data is like peeling the layers of an onion: different audiences begin and end their engagement at different levels of the data. For district, diocesan or system officers, the beginning level of interest is

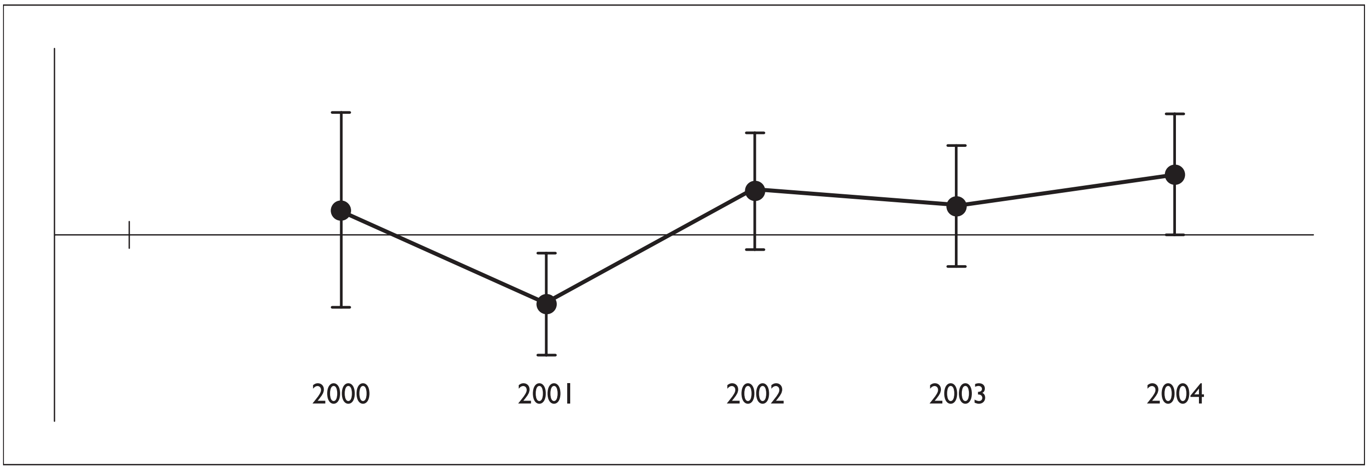

Figure 5 Example of a trend graph for primary analysis in one subject in a school

whole-state, leading down to system, then to individual schools and often stopping at that level. For a teacher, the initial point of interest is the department within the school, leading down to subject, then to class then to individual student. Teachers do not become engaged if they do not have student-level data, with each student identified by name. Hence the database that is provided as part of the package has the facility for schools to convert student ID numbers to names, and or any user to begin and end their consideration of the data at their points of interest.

We've learned that principals and teachers can be overwhelmed by a large dataset, but that you do need to provide the large dataset to enable each to follow his or her particular point of enquiry or interest. Hence, we have developed 'roadmaps' through the analysis package to give at least an initial way of logically engaging with the data. A typical roadmap for a principal takes him/her from the manual (DeCourcy, 2005a), to the trend graph 'Overall School Result', to Numeric Report 4, the 'school summary' from the database. This summary unpacks the overall school result to see the effect of each different subject, which can then be further investigated from the trends graph in that subject. If the second-order effect in the subject is significant, this is noted on the trends graph and can have (see below) significant utility in developing pedagogy. Roadmaps have been developed for the use of a number of other audiences for the analysis.

We've learned that a school performance feedback system like this has to be responsive to the needs of the users, as strongly stated by Visscher and Coe (2002). Many of the elements of the analysis, including the web site, the available PowerPoint files, the manual and the database have been provided following the expressed needs of those using the analysis.

We've learned that once the principal is engaged with the analysis, s/he will begin to use it as a lever to move the pedagogy and curriculum of the school.

The idea of leverage

The analysis gives principals and teachers an external point of reference for discussion about pedagogy and for attempts to improve both pedagogy and thereby student achievement. In Amrein and Berliner's (2003) terms, we aim for a low-stakes analysis so that teachers engage; if the analysis becomes a high-stakes accountability exercise, then the focus shifts to dealing with the analysis, rather than using the analysis to deal with the pedagogy. There are many methods of engagement between principals and teachers: some are outlined below as the levers a principal might use. The manual (DeCourcy, 2005a) gives more detail on most.

Lever 1 for the principal is to ask for the production of a brief report on each subject, addressing just four questions:

- What have you been doing, and why?
- How is it going?
- How do you know?
- What do you plan to do next?

The third question demands that the teacher engage with the analysis in order to substantiate their answer to the second question. The fourth question becomes the answer, the following year, to the first question. There is not room in this sort of analysis for blame-the-students responses, unless the teacher can hypothesise a distinctive characteristic of the particular group of students. If s/he can, then dealing with it becomes the answer to the fourth question.

Lever 2 is the Overall School Result report, which uses the layout of Fig. 5 above to plot over time the aggregate comparative learning gain for all students in the school. The single aim of an increasing comparative learning gain on this is a valid and stringent target for all in the school.

Lever 3 is the School Summary provided as part of the database, which ranks each subject from the highest to the lowest comparative learning gain. There is potential for misuse here if the idea behind confidence limits in measurement is not understood. The fact that the differences between the comparative learning gain in different subjects are small, particularly when compared to the uncertainties is illustrated by the relatively large size of the ellipses in the primary analysis (Fig. 3). Hence the School Summary lists subjects simply with a statement of whether the subject is above, within the range of, or below expectation. Its advantage is that one can easily see those subjects which are close to the boundary between these categories.

Lever 4 is to engage with teachers in terms of comparative learning gain. The focus of future planning is always around ways in which the comparative learning gain might be improved, as this is the most reliable way of improving outcomes. For this reason, subject trend graphs such as that shown in Fig. 5 are the focus of attention. The aim is to keep the graph going up. In dealing with these, the most powerful leverage comes from the simple questions. 'Can you just explain to me why this graph looks like this?'

Lever 5 is to engage teachers with second-order effects in the data. The comparative learning gain shown in Fig. 3 is a representation of the mean learning gain for the whole class, a first-order effect. Equation (3) in the appendix has a second-order statistic u_{1j} which represents the school-level residual of the slope of the line of best fit for each of the individual students within that subject within that school. Obviously, the line of best fit for one school may be parallel to, steeper than, or shallower than the typical line of best fit through the students in all schools. Where it is significantly steeper, the comparative learning gain of the students in the higher end of the distribution has been relatively better, a statistic summarised by a simple '+H'. Where it is significantly shallower, the comparative learning gain of the students in the lower end of the distribution has been relatively better, '+L'. For the principal and teachers, these second-order effects are recorded in the School Summary report, the Trends Analysis, and individual subject reports. A subject where the focus is on supporting struggling students and allowing the capable to fend for themselves will be identified by a string of +L results. A subject where the focus is on the achievement of the best students and the remainder are allowed to find their own level will gain +H results. For the discussion between principal and teacher, it is a valid aim if you have achieved a +H one year to strive to keep those gains and attempt a +L the following year, all the while keeping the first-order effect positive and increasing. Similarly, a +L one year can lead to an aim for a +H the following year. The pedagogical direction is towards differentiated instruction.

Lever 6 is in 'further factor' analysis. Explanations of why a particular result has been achieved in the primary analysis of comparative learning gain often come back to hypotheses at the individual-student level. For example, prior study of the subject in earlier years, class size, frequency with which some students arrived on the late bus and the differing effectiveness of different teachers in multi-class subjects might be hypothesised. A crude test of any of these can be simply performed, using the data supplied to the school. The database supplied to the school for each student in each subject includes a calculation (using equation (14) from the Appendix) of the Achieved and Typical results for that student in the subject. The mean of each of these gives the coordinates of the centre of the ellipse in the primary analysis. For a categorical hypothesis, such as the students who had previously studied the subject, it is straightforward to gain the means of the sub-groups, and then compare how they plot.

Lever 7 is in monitoring participation in different subjects, particularly in those which are most challenging. There is a temptation for able students experiencing their first taste of really having to struggle with a subject to drop to lower levels of the subject. Marsh (1991), Marsh, Chessor, Craven and Roche (1995), Marsh and Rowe (1996), Marsh, Hau and Craven (2003) and the data from this project show that the key to outstanding results in higher-level subjects lies in the combination of high participation by students, positive challenge from teachers, and appropriate pedagogy. We should be ensuring that more students take on challenging subjects, rather than seeking to advise students out of the subject.

Which lever or combination of levers a principal or teacher chooses to use is dependent on the school and the students. Together, they form a powerful set of tools to address pedagogical change.

Conclusion

The international research and the data from this project show that using data as an accountability mechanism, producing league tables which amplify tiny and statistically non-significant variations between schools into large differences in rank, is not effective in improving student performance. What is effective is valid analysis of data, presenting the results of the analysis in an engaging way, targeting professional development to support use of the analysis and then engaging teachers in professional development to support changes in pedagogy.

It is the teachers who make the difference.

Appendix: statistical methodology for the multilevel analysis

For the total sample, the School Certificate results in English-literacy, Mathematics, Science, Australian Geography and History were converted to standard scores x_e, x_m, x_s, x_h and x_g, based on the whole-of-state means and standard deviations in each test.

For each subject k in the Higher School Certificate, the x values from two years earlier were obtained for all students taking the subject k. Within the Higher School Certificate, each student is awarded a scaled exam mark and a (school-based) assessment mark that is moderated for each school against the examination mark. The mean of these two marks for each student is his/her 'HSC mark' in the subject. HSC marks for each subject were re-scaled to the mean and standard deviation of the x values, to give y_k values for each student.

Within each subject k the values of the mean for each school j of x and y_k were obtained. Since both x and y_k lie on the same scale, the comparison of the means $\overline{y_{jk}}$ and $\overline{x_{jk}}$ is then a crude comparison of achieved result with what might be expected from students of a similar level of performance two years earlier.

The standard error of the independent variable can be estimated in the usual way as $\sigma_{x_{jk}}/\sqrt{n}$, where n is the group size for school j in subject k. However, Goldstein (1995, p. 3) notes that such a method is likely to underestimate the standard error of the dependent variable, since it assumes a random sampling from the population and in this study we are specifically assessing non-random (school) effects on the groupings of the dependent variable.

To estimate the standard error in the dependent variable and to investigate any gender or SES effects, a model is fitted to the data using MLwiN multilevel modelling software (Goldstein et al., 1998) for each subject. This gives the value of for student i within school j studying subject k as:

A multilevel model was then fitted to the data for each subject allowing second-level variation in the β_l value and including gender,

$$y_{ijk} = \beta_{0ij}x_0 + \beta_{1j}x_{ij} + \beta_2 g_{ij} + \beta_3 S_j + \beta_4 s_{ij} \qquad (1)$$

where g_{ij} is the gender of student i in school j, S_j is the school-level measure of socioeconomic status (Farish, 2005), and s_{ij} is the student-level measure of socioeconomic status, taken as the postcode-average of the fourth SEIFA index (Australian Bureau of Statistics, 2004). In this equation, following Goldstein et al. (1998):

$$\beta_{0ij} = \beta_0 + u_{0j} + e_{0ij} \qquad (2)$$

and $\beta_{1j} = \beta_1 + u_{1j}$ (3)

Allowing for variation of β_l at the second level detects school effects where the extent of the impact of prior performance varies from school to school, and reveals second-order effects as described. Allowing for gender and SES effects detects that part of the variance in the HSC mark which can be attributed directly to gender or SES, and is not part of the school effect.

This allows an estimation of the expected mean result in school j in subject k to be given as :

$$\hat{y}_{jk} = \overline{\beta_0 + \beta_1 x_{ij} + \beta_2 g_{ij} + \beta_3 S_j + \beta_4 s_{ij}} \qquad (4)$$

and attributes school effects as being:

$$SchoolEffect = u_{0j} + u_{1j}x_{ij} \qquad (5)$$

There is often more than one class of a given subject in a school. It would be ideal to construct a three-level model for each subject, with students nested within classes within schools. However, the data as supplied do not include class designation, so this is not possible. Lever 6, described above and in the manual (DeCourcy, 2005a), gives a way in which this separation may be achieved by individual schools.

A second analysis is then performed with each set of data, using the separate SC results, giving models structured thus:

$$y_{ijk} = \beta_{0ij}x_0 + \beta_{1j}x_{Eij} + \beta_{2j}x_{Mij} + \beta_{3j}x_{Sij} + \beta_{4j}x_{Hij} + \beta_{5j}x_{Gij} + \beta_6 g_{ij} + \beta_7 S_j + \beta_8 s_{ij} \quad (6)$$

where

$$\beta_{0ij} = \beta_0 + u_{0j} + e_{0ij} \quad (7)$$

$$\beta_{1j} = \beta_1 + u_{1j} \quad (8)$$

$$\beta_{2j} = \beta_2 + u_{2j} \quad (9)$$

$$\beta_{3j} = \beta_3 + u_{3j} \quad (10)$$

$$\beta_{4j} = \beta_4 + u_{4j} \quad (11)$$

$$\beta_{5j} = \beta_5 + u_{5j} \quad (12)$$

giving

$$\hat{y}_{jk} = \overline{\beta_0 + \beta_1 x_{Eij} + \beta_2 x_{Mij} + \beta_3 x_{Sij} + \beta_4 x_{Hij} + \beta_5 x_{Gij} + \beta_6 g_{ij} + \beta_7 S_j + \beta_8 s_{ij}} \quad (13)$$

and

$$\hat{y}_{ijk} = \beta_0 + \beta_1 x_{Eij} + \beta_2 x_{Mij} + \beta_3 x_{Sij} + \beta_4 x_{Hij} + \beta_5 x_{Gij} + \beta_7 g_{ij} + \beta_8 S_j + \beta_8 s_{ij} \quad (14)$$

Finally, the process of equations (6) – (13) is repeated with all variables converted to normal equivalent deviates in order to obtain overall relationships between the variances in the dependent and independent variables.

Derivation of the second-order effect

In equation (3) above, the residual u_{1j} is significant in educational terms for schools. If the value is positive and significantly above β_l (MLwiN provides both the value and standard error of the residual) then the school has provided significantly greater learning gain for the higher-achieving end of the student distribution than is found in other schools in this subject. Such a result is depicted with the designation '+H' in the school report of the subject.

If the value is negative and significantly below β_l then the school has provided significantly greater learning gain for the lower-achieving end of the student distribution than is found in other schools in this subject. This is depicted with the designation '+L' in the school report.

In both cases, the effect inferred is relative. A '+H' is necessarily a '–L', and vice versa. Neither a '+H' nor a '+L' an inference of a deficiency in the teaching and learning: it is simply an observation of an effect.

Conversion to the Tertiary Entrance Score scale

The process of producing the University Admission Index ('UAI', Cooney, 2000) derives a measure of the student's performance compared with a whole-of-age group cohort. (A index similar to the UAI is produced in each state in Australia, the Equivalent National Tertiary Entrance Rank (ENTER).) In NSW the UAI is produced from a Tertiary Entrance Score (TES). The TES is a mark out of 500, consisting of the aggregate of the best 10 units of the student's re-scaled scores, including a minimum of 2 units of English. (Most subjects are 2-unit in value, giving a mark out of 100.)

The process compares subject with subject within the HSC using students common to pairs of subjects to derive a mapping of the Board of Studies (BOS) marks to a new 'UAImark' for each student in each subject. From these a 'UAI mean' for each HSC subject is derived and published (Cooney, 2005). These UAI mean then vary over a wide range, representing the relative performance of the cohort taking the particular subjects. (For example, Mathematics Extension 2 has a UAI mean of approx 44/50; at the other end of the scale, Construction has a mean of approx 16/50.)

Schools in NSW are given no information about individual student's UAI or TES. However, it is possible to take the individual student marks as provided by the BOS and to map them to gain reasonably accurate TES values using the published data of the Universities Admission Centre (Cooney, 2005). This is done by a simple linear mapping, such that a value t_k is gained as the TES equivalent of the BOS mark y_k where y_k lies between the mapping points b and b_d, where these points map to the UAI/TES scale u_c and u_d thus:

$$t_k = \left(\frac{y_k - b_c}{b_d - b_c}\right).(u_d - u_c) + u_c \quad (15)$$

Comparison with State average

As a part of the feedback to schools, one of the six presentations of the data that is provided is the comparison of school and state mean in the subject. To place the differences between school and state means in all subjects on the same scale, the difference that is reported (Δt_{km}) is the difference on the TES one-unit scale.

Comparison between this subject and all others in the school

Comparisons of the different means of subjects within the school carry little information when the subjects are on the BOS scale. Even when the marks are re-scaled as described above to the TES scale, the fact that different subjects attract candidatures of varying ability means that little can be gained by direct comparison of means.

However, if one uses the NSW DET method (Smith, 1999) a clear-cut comparison between subjects within the school can be obtained. Because the TES process places all marks on a common scale, the comparison of each student's mark in a subject t_k with the mean value of that student's results in each of his/her *other* subjects t_{im} gives a measure of the extent to which the individual student's performance in the subject is ahead of or behind other subjects. The mean of these individual values for all students in the subject t_{kjm} then gives a reliable comparison of subjects within the school.

$$t_{kjm} = \overline{t_k - t_{im}} \qquad (16)$$

References

Amrein, A. L., & Berliner, D. C. (2003). 'The effects of high-stakes testing on student motivation and learning', *Educational Leadership, 60*, 5, 32-38

Australian Bureau of Statistics (2004). SEIFA 2001: *The socioeconomic index for Areas*. Canberra: Self

Braun, H. I., & Mislevy, R. (2005). 'Intuitive Test Theory', *Phi Delta Kappan, 86*, 7, 489-497

Cooney, G. (2000). The Universities Admission Index (UAI). Sydney: Universities Admission Centre

Cooney, G., (2005). Report on the scaling of the 2004 NSW Higher School Certificate. Sydney: Universities Admission Centre; NSW Vice-Chancellors Conference Technical Committee on Scaling

DeCourcy, J. S. (2005a). *HSC data analysis: The manual*, Sydney: Catholic Education Commission.

DeCourcy, J. S. (2005b). *Report on HSC 2004*. Sydney: Catholic Education Commission.

Farish, S. (2004). Funding Arrangements for Non-government Schools 2005-2008: Recalculation of the Modified A Socioeconomic Status (SES) Indicator using 2001 Australian Bureau of Statistics Census Data Canberra: Department of Education, Science and Training.

Goldstein, H., (1995). Multilevel Statistical Models London: Arnold, 2nd edition.

Goldstein, H., Rasbash, J., Plewis, I., Draper, D., Browne, W., Yang, M., Woodhouse, G. & Healy, M., (1998). *A User's Guide to MlwiN*, University of London: Multilevel Models Project, Version 1.0

Marks, G. N., Rowe, K.J., & Beavis, A. (2003). Australian Schools not so 'Undemocratic', *Campus Review, June/July 34*.

Marsh, H. W., (1991). Failure of high-ability high schools to deliver academic benefits commensurate with their students' ability levels, *American Educational Research Journal, 28*, (2), 445-464.

Marsh, H. W. & Rowe, K. J., (1996). The negative effects of school average ability on academic self-concept: An application of multilevel modelling, *Australian Journal of Education, 40*, 65-87.

Marsh, H. W., Chessor, D., Craven, R. & Roche, L. (1995). The effects of gifted and talented programs on academic self-concept: the big fish strikes again, *American Educational Research Journal, 32*, (2), 285-319.

Marsh, H. W., Hau, K. T., & Craven, R. (2003). The big-fish-little-pond effect stands up to scrutiny, *American Psychologist, 59*, 4, 269-271.

O'Day, J. A. (2002). Complexity, accountability and school improvement, *Harvard Educational Review, 72*, 3, 293-329.

Rowe, K. J. (2000). Assessment, league tables and school effectiveness: Consider the issues and 'Let's Get Real', *Journal of Educational Enquiry, 1*, 1, 73-98.

Rowe, K. J., (2001). Equal and different? Yes, but what really matters?, Riverview Boys Education Conference, October 2001.

Rowe, K. J. (2004a). 'The Importance of Teaching: Ensuring better schooling by building teacher capacities that maximize the quality of teaching and learning provision – implications of findings from the international and Australian evidence-based research'; Paper presented at the 'Making Schools Better' Conference, Melbourne University, 26-27 August 2004.

Rowe, K. J. (2004b). '*Analysing and Reporting Performance Indicator Data: 'Caress' the data and user beware!*' Paper presented at the 2004 Public Sector Performance and Reporting Conference, Sydney 19-22 April 2004.

Smith, M. (1999). Methods for within-school comparison of subject results Pers. Comm.

Visscher, A. J. & Coe, R. (Eds), (2002). School improvement through performance feedback, Lisse, The Netherlands: Swets and Zeitlinger.