Data-informed research and practice: Evaluating student achievement in secondary schools

Carmel Richardson

Australian Council for Educational Research

Carmel Richardson is a Senior Research Fellow at the Australian Council for Educational Research. She taught in secondary schools for more than twenty years, mainly Maths, English and Special Education, before commencing doctoral studies at The University of Melbourne. Carmel's main research area is in analysing ability-adjusted student achievement data to provide schools and teachers with information to track the learning progress of each student. She analysed the data for the Hong Kong Quality Education Fund (1997–2000) as part of the project to develop Performance Indicators for Measuring Value-added Improvement of Primary and Secondary Students' Academic Performance. Carmel has conducted large-scale multilevel modelling analyses of Victorian final secondary assessments (1994–2000) for all secondary schools in Victoria, Australia. Carmel moved to ACER in 2003, and is now providing detailed 'within-school' multi-level analyses of Victorian Certificate of Education data at student and class levels for more than one hundred Victorian secondary schools. This work is being extended to primary schools as well as to the junior and middle years in secondary schools. Carmel is Project Director of ACER's Data Interpretation Service.

Abstract

The advantage of 'ability-adjusted' analyses of educational data is their capacity to provide fairer assessments of school and student achievement than reliance on raw scores alone. School performance evaluations based on students' unadjusted (raw) marks favour schools with higher intakes of bright and advantaged students. The learning gains of middle and lower ability students are overlooked, and the achievements of students and schools in disadvantaged areas are not valued, while focus is concentrated on those achieving the highest marks. With 'ability-adjusted' analyses of school data, any student who achieves higher marks than similar ability peers is acknowledged as having performed well. This paper describes findings from a series of 'ability-adjusted' analyses conducted within individual schools, where students' Victorian Certificate of Education (VCE) results were analysed at student and class levels. Staff members were assisted with verification and interpretation of their data to ensure its positive use within their school. This research led to a number of practitioners seeking 'ability-adjusted' analyses of their junior and/or middle-secondary students' achievements, as they recognised the benefits of this *data-informed* approach. The impact in terms of improving teaching and learning, and the on-going challenges inherent in designing each school's database, aligned with curriculum and assessment policy, are discussed.

Background and context

Data-driven quality assurance is a popular term used by system bureaucrats and researchers far removed from the heart of education – students and teachers in schools – whereas *data-informed* evaluation of student achievement and school performance is the term preferred by practitioners. This is not a pedantic wordplay – it highlights a key difference in the attitudes and practices of system personnel compared with those of school staff. At system-level, the focus is on a top-down, reform *driven* judgement of schools in terms of their students' achievements, with teachers typically assumed responsible when underperformance is identified. For school staff, the emphasis is on integrating their external and internal quantitative results with their qualitative data, to more comprehensively *inform* their monitoring of student achievement and their school's performance.

But what processes are required to ensure that both educational systems and schools accurately and fairly assess student and school achievement? Simplistic rankings place schools and classes with large numbers of bright students as the top performers, while real achievements in schools and classes with more disadvantaged and lower ability students are ignored. Clearly, as far as possible, all variables that affect student learning ought be taken into account, if genuine 'value-added' educational performance is to be recognised. Research in the School Effectiveness and School Improvement (SE&SI) tradition has consistently identified individual student ability and prior attainment as key factors associated with student achievement; and socioeconomic status is the most commonly debated contextual variable (Hattie, 2003; Hill & Richardson, 2001; Hill & Rowe, 1996, 1998; Mortimore et al., 1988; Schereens & Creemers, 1989; Teese & Polesel, 2003).

Teachers need ability and prior attainment data on each student at the start of the school year to monitor student progress effectively, and to provide parents with valid reports on their child's learning gains each semester. Schools are hampered from achieving these goals for many reasons, one being the negative attitude towards ability measures held by some educators, due in part, to the misuse of 'intelligence' tests throughout the last century. The major reason, however, is the lack of an ability measure, and system level failure to supply schools with developmentally appropriate attainment measures, scored on a common metric, longitudinal scale. Currently, school reports do not provide the next year's teacher with indicators of the standard that students have achieved in terms of clearly delineated skills and knowledge within each subject; nor do they give parents indicators of their child's achievement in relation to his or her potential. There are divergent views in the educational community about the merits and demerits of schools' reports that indicate student ability and achievement in relation to school-aged peers.

This paper discusses one approach where ability and achievement data were analysed, electronically displayed and comprehensively interpreted to assist school staff in monitoring student and school performance. This work was built on 'value-added' analyses from two large-scale research projects, which involved multi-level modelling of VCE results over the past ten years. Key findings included:

- The *value of an ability measure*, appropriately verified and comprehensively interpreted, for more accurate evaluation of school and student performance.
- The evidence regarding *the real gender effect*, illustrating the error in the general statement that 'boys are underperforming in relation to girls', based on overall patterns in aggregated data.
- The dangers associated with referring to class-level variance, that is, the class residual, as *the class/teacher effect*, or even more misleading, as *the teacher effect*.
- The need for each school to develop a 'within-school' database to enable *'ability-adjusted' monitoring of student and school performance*.
- The need for better resourcing to ensure *instructional effectiveness* within schools, and greater focus on monitoring *system-level effectiveness*.

The research background 1. 'Across-schools analyses'

For the seven years, 1994 to 2000, ability and attainment data for all VCE students in every government, Catholic and Independent school, were analysed at student and school levels, in a series of variance components models for each of the 20 largest VCE Studies.

The measure of ability for this 'across-schools' research was the General Achievement Test (GAT), based on general knowledge and skills in three domains: Written Communication (GAT c); Mathematics, Science and Technology (GAT m); and Humanities, the Arts and Social Sciences (GAT h). Each year, students are informed that they do not have to do any special preparation for the GAT, as the basic writing and reasoning skills being assessed have been developed in their earlier years of schooling, although they are advised to look at sample questions and past papers. The GAT is a component of the statistical moderation processes used by the Victorian Curriculum and Assessment Authority (VCAA) in monitoring school assessed work. The student's VCAA Study Score was the achievement measure for this research.

The methodology and modelling followed the process initiated by Hill & Turner in 1995 for 10 schools in the pilot version, and further developed by Rowe in 1998 for 50 schools in the trial VCE Data Project (Rowe, 1999). Effect sizes were calculated for the five explanatory variables – three student ability measures, school mean ability in each Study, and student gender, used to generate each student's expected VCE results. School residuals, representing the difference between the predicted and achieved VCE scores, were plotted for the 20 subjects each year. Patterns of consistency in the school's performance across the 20 Studies each year, and stability in each subject over time, were noted (Richardson, 2000a). The school residuals at this time were interpreted as indicators of the school's 'ability-adjusted' position amongst VCE providers across the State (Rowe, 1999).

Several concerns surfaced when these 'ability-adjusted' results were shared with schools. Practitioner-informed explanations for the patterns in their school's subject residuals were not always congruent with interpretations typically made by system-level bureaucrats and academic researchers, who rely too often on statistical analyses alone, and frequently 'got it wrong' (Richardson, 2001). In Victoria's League Tables, published from 1996 to 2000, there were blatant examples of schools being incorrectly highlighted by the media as 'top performers' or unfairly labelled as 'failing schools' based on system level data analyses with ability 'supposedly' taken into account. Detailed examination of these data revealed the problems with such gross school rankings, when, regardless of the standard achieved, half the schools in the State had to be below the median, by definition; and one school has to be 'bottom of the ladder' each year. Such 'so-called' accountability rating rarely affected elite schools in wealthy suburbs, but impacted most negatively on schools in poorer areas (Richardson, 2002).

The key difference between the system-level VCE data provision to schools since 1994 and this in-depth research project was the latter's inclusion of qualitative research undertaken with schools (Richardson, 2000b). Staff feedback and suggestions were integrated into subsequent data analyses each year, and macros and a software package were developed to facilitate data display and interpretation. The more informative, visual presentation of both raw and ability-adjusted data in this project was preferred by school staff to the residual plots provided at system level, because teachers could verify the raw data, and explain some of the patterns in the analysed results. In response to the positive feedback from schools, this doctoral research developed into an independent Data Interpretation Service, now operated by ACER.

The research background 2 'Within-school analyses'

The comprehensive verification of the data and review of the multi-level analyses indicated that in-depth, 'within-school' analyses had to precede 'across-schools' analyses, to ensure fair evaluation of school performance. In 2001, the 'within-school' analyses were developed and trialled. Over the next three years, 16 schools (2002 VCE data), 90 schools (2003 VCE data), and 105 schools (2004 VCE data) voluntarily participated in this research, that is, they effectively funded it.

One of the problems highlighted in the 'across-schools' research involved the use of the VCAA Study Score as the measure achievement. It is the student's rank, relative to all other VCE students within each Study. However, this rank is unsuitable for comparisons across an individual student's VCE Studies, and when comparing class and subject achievement within schools. In recognition of this, the Victorian Tertiary Admissions Centre (VTAC) transforms VCE Study Score ranks to scaled scores for calculation of students' ENTER (Equivalent National Tertiary Entrance Rank). As students and schools in Victoria are not given students' final VCE marks, only their Study Score ranks, the Scaling Guide that VTAC publishes to schools each year was used to calculate the student's VCE marks for this research. The difference between VCAA Study Scores (ranks) and VTAC Scaled Scores (marks) is illustrated and discussed further in the section on class level analyses below.

Another problem identified in the initial research was that year level was a key variable predicting student performance in some Studies, and differentially so in some schools. This factor needed to be included in the modelling. Students typically complete VCE Units 1 and 2 in Year 11, and VCE Units 3 and 4 in Year 12, although some students study one or more VCE Units 3 and 4 in Year 11.

For the two-level (students in classes) variance components modelling for the 'within-schools' analyses, the six explanatory variables used to predict the student's VCE marks were:

- Three student ability (GAT) measures: Written Communication (GAT c), Mathematics/Science/Technology (GAT m) and Arts/Humanities/Social Sciences (GAT h)
- Class mean ability (the mean ability for all students, with student ability calculated as the average (AvGAT), of the three GATs).
- Gender (males = 0, females = 1)
- Year level (Year 12 students = 0, Year 11 students = 1).

Effect sizes, and proportion of variance at student and class levels, were calculated for each school, which included government, Catholic, and Independent schools, small and large schools, single-sex and co-educational schools, and urban, regional and rural schools. Both student and class residuals were examined for both typical and atypical patterns. The residual is the difference between the predicted mark, based on student ability (3 GATs), gender, year level, and class mean ability (AvGAT), and the achieved VCE mark. Raw and ability-adjusted VCE results were summarised in graphs and tables, and provided to each school on a CD. To ensure that the analyses and the interpretations were statistically sound and educationally meaningful, the researcher and practitioners discussed and debated the results in a professional learning seminar. This allowed for the rich contextual knowledge available within the school to be taken into account in evaluating student, class and school performance.

Broad consistency in effect sizes and school patterns were found for the 2002-2004 'within-schools' research and the 1994-2000 'across-schools' population-level research, especially in terms of the average magnitude of variance explained – between 5-15% at the second level (classes for the 'within-schools' research; and schools for the 'across-schools' research). It is not possible in this paper to discuss all results of these two large-scale projects, however different aspects of the research have been presented in greater detail at seminars and national and international Conferences (Hong Kong, 2000a; Melbourne, 2000; Denmark, 2001; Sydney, 2003; Melbourne, 2003; Sydney, 2004). A series of seminars will be held after the Conference, where more in-depth displays and explanations of the research findings will be presented.

Student ability patterns

Some patterns from one school are outlined below to enable indicative patterns, and the depth of these 'within-school' analyses and data interpretations, to be examined. In all graphs, whether at student, class or school levels, individual student ability (AvGAT) was plotted on the *X*-axis, and achievement (VTAC mark) on the *Y*-axis. Drop-down menus beside each graph enabled results for particular students, classes or subjects to be highlighted as white circles, against the background of dark diamonds displaying the school's VCE results for the year. For example, in Figure 1, the diamonds pinpoint the ability and achievement scores for every student completing their VCE in this school in 2004. The white intersecting lines on the graph indicate State means (AvGAT = 20, VTAC Study mark = 30), and the diagonal line crossing the diamonds is the school's regression line. No student or class has been selected here (the blank space in the legend beside the white circle in the graph's title).

Noteworthy features are the general pattern of achievement increasing with student ability (higher GAT scores associated with higher VTAC marks, r = 0.6435); and the variation between students at each ability point. For example, around the State mean ability level (AvGAT ≈ 20), the range for individual students in this school in 2004 varied from 13 to 43 VTAC marks.

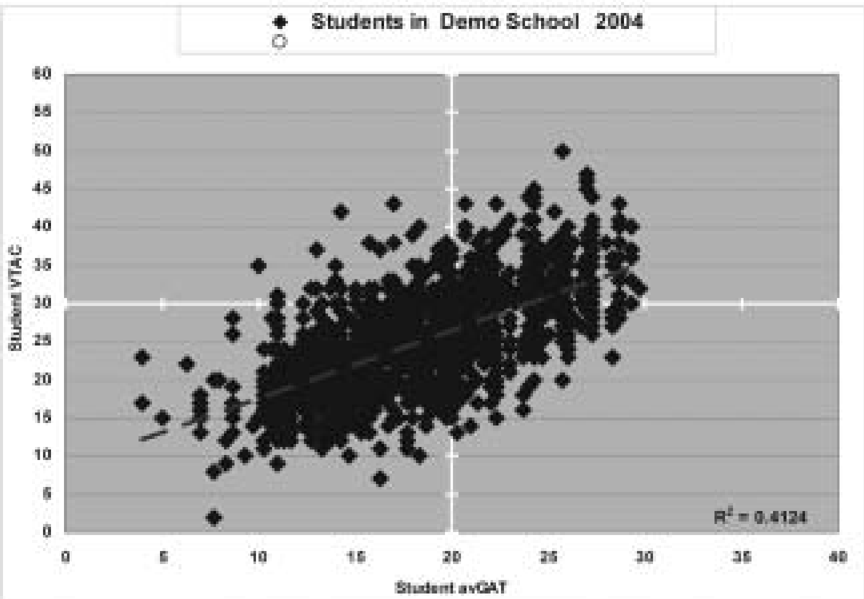
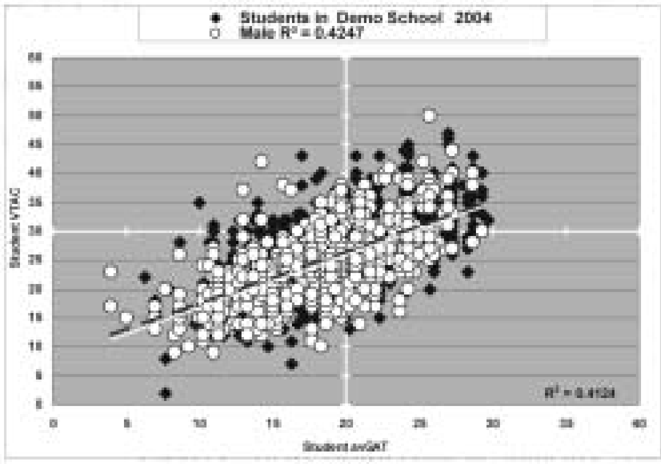


Figure 1 Scatterplot of Students' Mean GAT Score (Ability), Plotted against Students' VTAC Marks (Achievement).

This was the typical range across VCE ability and achievement data both within schools and across schools, except for schools with fee-paying overseas students, where a distinctly different pattern was evident. Further examination of these schools led to the understanding that two separate analyses were needed for such schools, to avoid the distortion that results when ability-adjusted data for overseas students are included in the VCE data analyses, and like schools reports.

Figure 1 demonstrates that while ability plays a large part in students' final academic achievement, the range in marks at every level of ability is considerable, thus ability alone does not determine final achievement. The multi-level modelling revealed that in this school in 2004, individual student ability as measured by performance on the three GATs accounted for 41% of the variation in student marks. Class mean ability, gender and year level explained a further 3% of student differences in VCE achievement. Thus, a total of 44% of the variation in student's scores was explained by the six factors modelled, with differences between classes accounting for 9% of the variance. The remaining 47% of unexplained variance in VCE results in this school was associated with factors not measured in these analyses. This unaccounted-for variation was what the discussions with staff in each school were intended to uncover, and were usually attributed to student effort, motivation and aspirations, teacher skill, school and home resources.



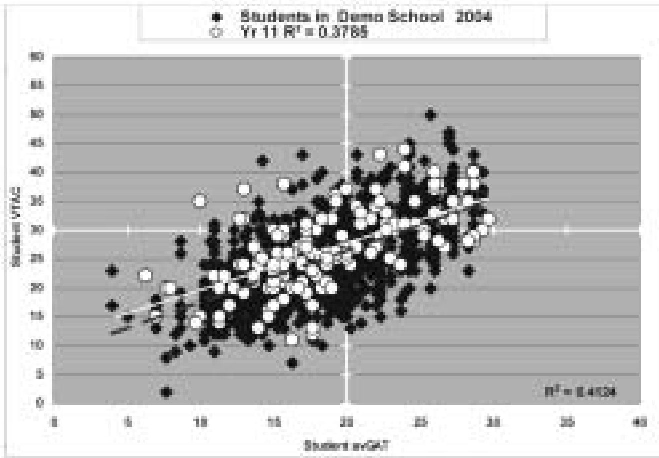


Figure 2 (left). Scatterplot of Students' Ability and Achievement, with Boys Highlighted (white circles).
Figure 3 (right). Scatterplot of Students' Ability and Achievement with Year 11 Students Highlighted.

From this graph of overall school VCE achievement (Figure 1), individual students could be identified using a 'Who is …?' button. This allowed obvious outliers in the data to be immediately identified, and their ability and achievement data further examined by reference to the tables and graph on the Student page (see Figure 8 below) of the CD. In addition, displays of school gender and year level patterns selected from the following eight options: Year 11 students, Year 12 students, males, females, Year 11 males, Year 11 females, Year 12 males, Year 12 females, could be highlighted on the Figure 1 graph.

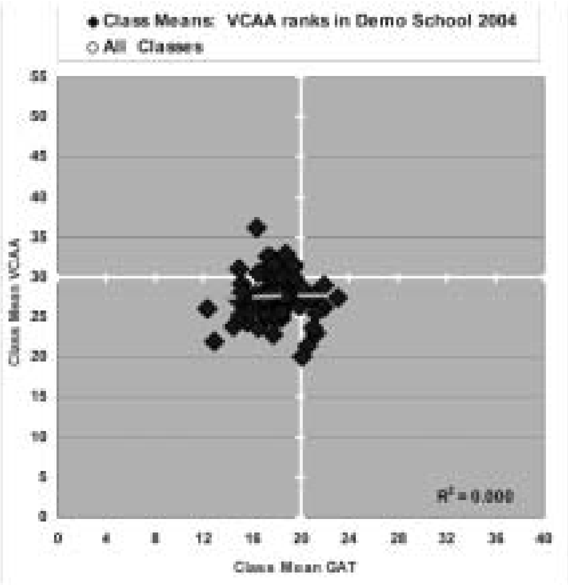
Gender and year level patterns

Similar patterns were evident in terms of the range for both ability and achievement data when gender and year level were examined in the 'within-school' analyses. The effect size for gender in the variance components modelling for this school's VCE 2004 scores was 1.2 and the effect size for year level was 1.7. That is, girls averaged just over one mark higher than boys (Figure 2), and Year 11 VCE students averaged almost two marks more than Year 12 VCE students (Figure 3). In these two graphs, the dashed line represents the school mean performance for boys (Figure 2), and the school mean achievement for Year 11 students (Figure 3).

At each ability level (Figure 2), there are boys achieving VCE marks higher (white circles above the regression line) and lower (white circles below the regression line) than the school's average across all Studies. This graph indicates how misleading the gross statement that 'girls are outperforming boys in VCE' is, given the range in marks for boys at each level of ability. The typical pattern was that bright boys achieved as well, if not better than bright girls, but more lower-ability boys performed worse than lower-ability girls. The mean gender effect at class and school levels, in both the 'within-schools' and 'across-schools' research, was due to the poor results of some of the lower ability boys, not because all boys are performing worse than all girls.

The educationally more informative questions in terms of gender ought be: 'Which boys are performing better than similar ability boys within the school?' and 'What factors are influencing some low ability boys to perform well, while other low ability boys do poorly?' These two questions can be re-worded for girls, as the same situation applies – at every ability level, there are groups of girls achieving both above and below their predicted score. The relevant focus for teachers and schools is 'which boys' and 'which girls' were under-performing, when evaluating performance. This research has more potential than the Federal government's response to the 'boys' under-performance' problem than male-only teaching scholarships, as it enables positive examples of low ability boys (and girls) who are performing well to be identified within each school, and the factors that contributed to their success can be evaluated and shared with all schools.

The same questions can be asked regarding year level and other variables known to affect student achievement: for example, 'Which Year 11 students (highlighted as white circles in Figure 3), performed better (or worse) than similar ability peers?' Careful



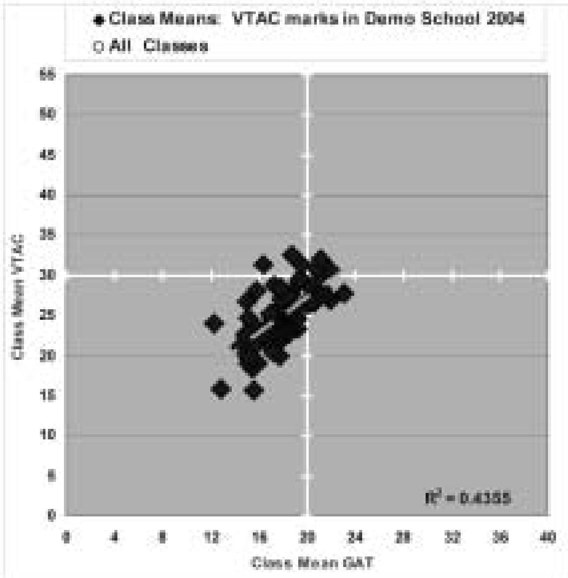


Figure 4 (left). Class Mean GAT and Class Mean VCAA Score.
Figure 5 (right). Class Mean GAT and Class Mean VTAC Score

examination of this graph, and discussions with students and teachers to accurately discern the reasons associated with each student's performance, enabled positive action to be taken where deemed necessary, for these students are the school's current Year 12's.

Class-level analyses

When class-level data were first examined in 2001, the need for the 'ability-adjusted' analyses to be conducted on the student's mark (VTAC Scaled Score), not the relative State rank (VCAA Study Score) within each Study, became evident. The following two graphs illustrate this at class level, where its effect is strongest; and were prepared to assist teachers to understand the difference between the VCAA Study Score and the VTAC Scaled Score. This pattern occurred also at individual student level, although to a variable degree, as students' marks are differentially affected by the impact of VTAC's scaling of Studies. In Figure 4, class mean GAT scores are plotted against class mean VCAA Study Scores; and in Figure 5, class mean GAT scores are plotted against class mean VTAC Scaled Scores, for the selected school.

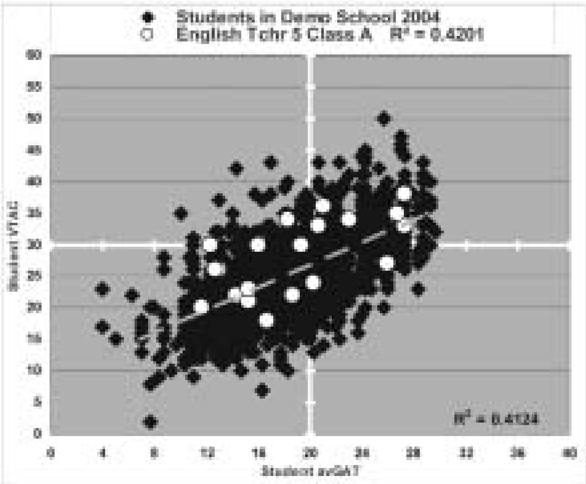
The pattern in the data in Figure 4 is of concern, with some low ability classes achieving higher class mean VCAA scores than some high ability classes in this school in 2004. The correlation between these two class level variables was zero, indicating no discernible relationship between class average GAT scores and class average VCAA Study Scores (ranks).

However, in Figure 5, when VCE marks (class mean VTAC Scaled Scores) were plotted against ability (class average GAT scores), the expected pattern for educational data was found (r = 0.660), with higher ability classes generally achieving higher marks than lower ability classes. In Figure 5, classes (diamonds) above the regression line are interpreted as performing better than expected within this school, while those below the line are not performing as well as predicted, based on the variables adjusted for in the modelling, and on the overall pattern in the school's data. A range of performance is evident at each ability level. Clearly, there are factors other than ability which influence class achievement, and hence the results for all students in each class were examined, along with the patterns within and across teachers for all VCE classes in the school.

Class and teacher patterns

The following two class graphs (Figures 6 & 7) provide examples of VCE results for two English classes (A and B), taught by the same teacher (Teacher 5) in the same school in the same year.

In teacher 5's first English class (A), more students were above than below the school's regression line, while in Class (B), the reverse pattern was



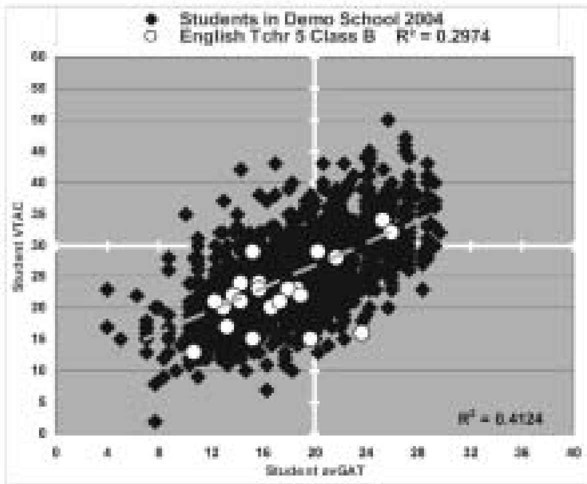


Figure 6 (left). English Teacher 5, Class A
Figure 7 (right). English Teacher 5, Class B

found. Yet in both classes, most students were on or around the school's mean line, that is, they performed as expected given their ability. In class B, however, two students were ten or more marks below the regression line, noticeably dropping below the majority of the school's 2004 VCE cohort (densely clustered diamonds). In particular, Student 429, identified as the third highest in terms of ability (AvGAT = 23.7), with a mark of only 16, strongly affected the mean achievement in Teacher 5's second English class.

Class A (Figure 6) had a positive class residual (1.6) and Class B (Figure 7) a negative class residual (–1.4). This is interpreted as the mean English mark for Class A was 1.6 marks higher than predicted, and Class B's mean was 1.4 marks lower than expected, given the six factors adjusted for in the multi-level modelling. Obvious questions are:

- 'Why the difference?'
- In what ways is the teacher responsible for the three mark difference in the two English class means?
- Who decides if this difference is educationally meaningful?
- What process determines where the line of acceptable ability-adjusted performance is drawn, and how is this authenticated?

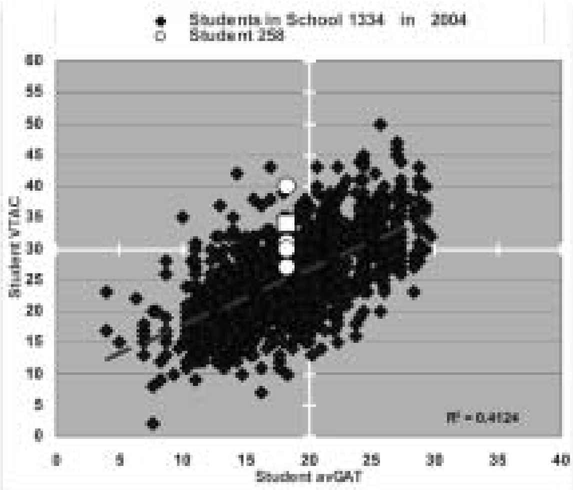
Detailed examination and informed discussion of the data were necessary when evaluating the factors believed to influence overall class achievement. More often, the effect of individuals or small groups of students within a class appeared to have a greater effect than an individual teacher on the class mean achievement. Frequently a student who achieved high marks in one class also scored positively in their other Studies, and vice versa, as Figure 8 & 9 illustrate.

Student level analyses

The white circles display an individual student's performance, set against the results for all students (diamonds). The English mark for each student is shown as a white square to assist with comparison and location of these students in their respective English class graphs (Figures 6 & 7). Data for a high performing student in Class A (Student 258) in Figure 8, and a low performing student in English Class B (Figure 9) are now examined. As can be seen, both students performed in similar fashion in English as they did in their other classes.

Student 258 (AvGAT = 18.3) in English Class A, with a mark of 34, contributed to Class A's positive residual (1.6), and generally achieved at or above expected level in all Studies. Student 429 (Figure 9) generally achieved less than expected, relative to other students of similar ability in this school in 2004, and contributed to the negative residual for his English class. This leads to the question: "To what extent can the English teacher be held responsible for the English marks of these two students?" These data, of themselves, do not and can not tell us whether Teacher 5 was a good, average or poor teacher of English in either class.

School staff generally attribute student effort and interest, or lack thereof, as the main explanatory factors in student performance. Students with positive residuals were described as those who



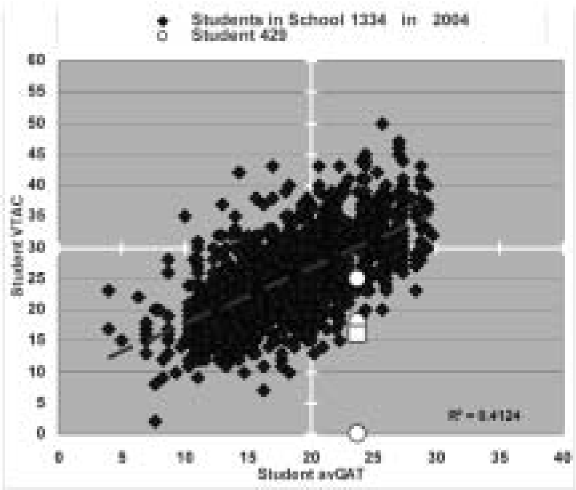


Figure 8 (left). VCE results for student 258, English Class A Teacher 5 (highlighted)
Figure 9 (right). VCE results for student 429, English Class B Teacher 5 (highlighted)

had high aspirations, gave appropriate time to the subject regularly throughout the year, and took notice of teacher feedback and instructions for improving their work. Students with negative residuals were usually said to have had low motivation, lack of home support for learning; and in some cases, illness and trauma were relevant factors.

Results of this research (Richardson, 2004b) indicate that the class residual ought not be referred to as the teacher residual. Even reference to variance at this level as the class/teacher residual, needs caution, given the unit of analysis is merely aggregated student level data, not specifically measured teacher or class variables. In multi-level modelling research, the class residual is simply the difference between the adjusted mean for all students in a class, compared with the adjusted mean for all other classes in the school. Yet in the vast majority of cases in this research, negative class-level residuals were clearly influenced by factors associated with a small group of students. Principals and senior staff in schools did not automatically associate the class residual with the measure of the teacher's effectiveness, as academics and system level staff tend to do.

The fact is that few researchers have actually analysed data that included valid measures on which teacher effects could be calculated. Until such measures are defined and gathered, claims of teacher effect sizes, calculated from multi-level models of students in schools, or even students in classes, must be more closely examined.

However, it is important to note that, in schools where such analyses were conducted over several years, teachers whose class residuals were strongly positive year after year were often the ones that colleagues named as 'high-performing' teachers. This was substantiated with detailed reference to the individual teacher's behaviours in terms of curriculum contribution, assessment practice, student feedback, and collaboration within the school. Other characteristics of 'top teachers' acknowledged by VCE staff in this research were openness to their own, ongoing learning, and capacity to acknowledge both 'good' lessons and 'difficult' lessons. These teachers were not paraded as 'perfect' teachers or persons, but as genuine educators, who loved learning, had strong discipline knowledge and love of their subject material, and were able to communicate well and sustain positive relationships with students. Note that no evidence was provided in the sense of these qualities being measured as they were merely observations of, and attributions made by, their peers within the school.

As a consequence of these data-informed discussions, many teachers independently selected areas of focus for themselves for their current VCE teaching – more examination practice, greater monitoring of student written work in class throughout the year. Examples of instances where a negative class residual was attributed to a 'poor' teacher were rare, but some class patterns did generate concern. Further investigation into their students' performance in their other Studies was undertaken in discussions with the teacher, as was consideration of contextual factors that may have accounted for the less-than-expected achievement. In some schools, additional support was given in terms of formal and informal mentoring. Some teachers were encouraged to develop contact with subject networks groups for improved access to curriculum and assessment resources and information.

School level analyses

In Figure 10, all class residuals (dark diamonds) for this school in 2005 are plotted in rank order from lowest to highest, with all English KLA classes (white diamonds) selected. The bounded line around each diamond indicates the 95% Confidence Interval for each class residual. As is the case in all schools, the majority of class residuals (vertical scale) in the school are within ±2 marks of their expected achievement. Residuals for Teacher 5's two English classes (Class A's residual 1.6, and Class B's residual – 1.4) are highlighted as grey squares (Figure 10). Because their respective confidence intervals do not overlap on this class residual plot, statisticians consider that there is a statistically significant difference between these two classes, and some then refer to this as the 'teacher effect'. However, detailed examination of individual students' results in Teacher 5's two English classes revealed that the difference was largely associated with performance of several students in each class.

The data and discussion associated with Figure 10 provides one example of the misinterpretation that can occur when statistical analyses alone are used to estimate school, subject and teacher performance.

At system-level, and in 'League Table' summaries, so-called 'failing schools' and 'top schools' are identified from such ranked residual plots, without any reference to the multi-variate, multi-level factors influencing these results, let alone acknowledgement of the unmeasured (and possibly unmeasurable) factors. Too often, negative subject or class residuals are misrepresented as the school or teacher effect, simply because the patterns across students are hidden. Only when lower level (student) data are examined is this problem avoided. A more detailed discussion and interpretation of patterns in residual plots will be presented in a series of seminars to be held at ACER later this year.

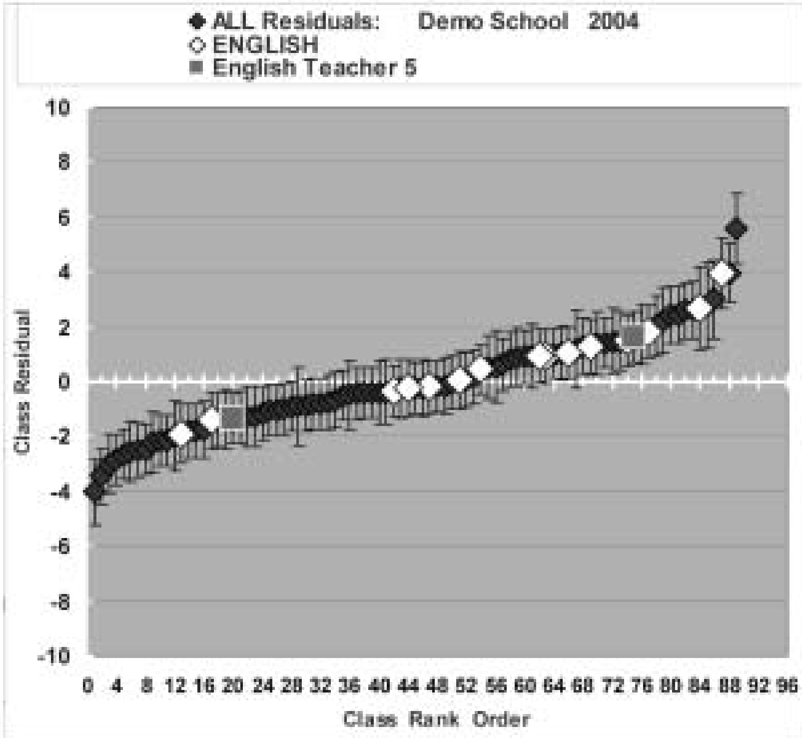


Figure 10 Ranked Class Residuals in this School, 2004 (dark diamonds), English Class Residuals (white diamonds), and Teacher 5's two English Classes (squares).

Principals and teachers preferred scatterplots (Figures 1–9) to residual plots (Figure 10), when examining their school's data, as the former better illustrated the meaning of student and class residuals. Within the school, staff could identify instances where the student and class residuals were inaccurate, and make appropriate adjustments in their evaluation of their school's performance.

Within schools research, years 7-11

School staff with access to this level of detailed student and class data quickly recognised what they described as 'the value of a good ability measure' to provide them with value-added information on their students' academic performance (Richardson, 2002, 2003a). In some schools, senior staff set about obtaining an independent measure of student ability at the key learning stages – entry to secondary school, and in Year 10 when there was a focus on work experience, careers advice and VCE subject choices.

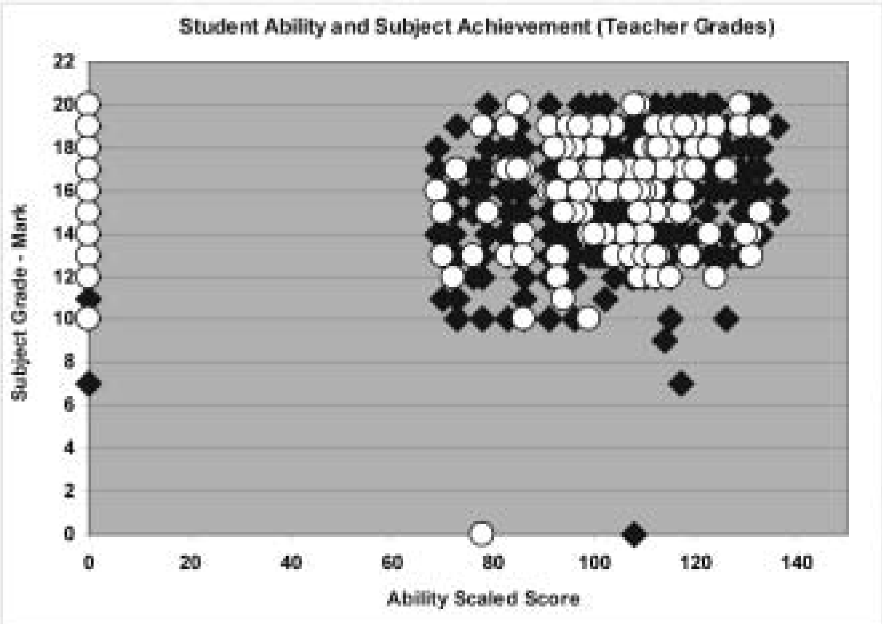


Figure 11 Typical Scatterplot of Student Ability and Achievement (based on teacher's semester grades/marks).

Concerns were often raised about the relationship between the results students receive on school reports, and their academic performance as measured on external assessments. Figure 11 illustrates the typical pattern found when internal school assessments (in this case, semester report grades) are plotted against external measures (in this case, an intake ability test). White circles represent one school's Year 8 Mathematics results for Semester 1, 2003, set against all Year 8 students' subject results (diamonds), with teachers' grades converted to marks (A+ = 20, A = 19, A– = 18, B+ = 17, etc.).

The diamonds and circles on the *Y*-axis (vertical line at zero ability score) represent students not assessed on the ability measure on entry to the school, and the missing data on the *X*-axis indicate students no longer at the school. Note the lack of correlation between ability and teacher grades, highlighting the reality that, when writing reports for students, teachers' grades are based on both observed behaviours and examined subject material over the semester. Some teachers give positive grades to 'reward' students for effort, and to encourage lower ability students. Figure 11 reflects the high variability amongst teachers when assessing student achievement, sometimes found even when moderation procedures are in place in the school.

Many schools are developing processes to support their teachers in monitoring and improving their assessment and reporting practices, and some schools have already begun this venture towards becoming a *data-informed* school (see poster displays at this Conference for examples). Sally Paterson now outlines the way her school embarked on the task of *'using data to support learning'*.

Research into practice

Urrbrae Agricultural High School (UAHS) is a specialist agricultural school located in suburban Adelaide. The school has 1000 students, all of whom are selected to enter the school. As with many schools, one of our goals is to achieve excellent learning outcomes, in particular, as expressed in our Strategic Plan: 'To achieve excellent learning outcomes which allow our graduates to be skilled contributors to our community'. This generated debate within our school about an operational definition of excellent learning outcomes. Subsequently, consideration was also given to the second strategic goal: 'To achieve growth of social capital for a community that is socially and environmentally sustainable'.

Defining excellence in terms of tertiary education entry scores was not appropriate for or relevant to many of our students, and also left us to work with data available only after students had left the school. We wanted the capacity to monitor progress of all students towards the goals as they moved through the school. Debate over the meaning of excellence led to a belief that, for us, it would be for the school to make a positive impact on student achievement. On an individual basis, excellence was defined in relation to the student's starting point. To monitor achievement in each learning area, we needed a measure of student ability. The proposal to collect baseline data was controversial in the school, with some fears raised from past memories or myths of IQ tests. However, staff members recognised that our school not only had goals relating to the quality of academic learning, but also to the development of social capital, as stated above, and gathering a multidimensional student profile would give us the opportunity to monitor all aspects of each student's development.

A commitment was made to establish a database with a comprehensive array of information gathered for each student, including a measure of ability in four domains – verbal reasoning, numerical reasoning, abstract reasoning and visual spatial reasoning, a measure of students' thinking style/learning preference, and their self-reported attitudes to learning and to the school's focus areas. These attitudes were expressed on a school-developed survey. Teachers also collected an example of student writing conducted in class throughout the first semester, to provide a baseline against which development in students' written expression could be mapped.

Research has shown that what the student brings to the learning situation predicts 50% of their achievement. The collection of baseline ability data gave the school the opportunity to identify the starting point for each student. Some research (Hattie, 2003) described the factors that predict student achievement as being individual ability (50%), the influence of school, home and peers (20%), and the quality of teaching (30%). To allow us to focus on the impact we can have, as school staff, we need to be clear about the factors we cannot influence, such as the student's ability on entry to the school.

To monitor our progress in achieving excellent outcomes, it was necessary to ensure that teachers were using at least some common assessment tasks, completed individually and under supervision. Teachers were involved in moderating these assessments. In some learning areas, new assessment tasks were devised and in other areas, existing ones fitted the appropriate criteria. In all cases, the assessments were referenced to Level 4 outcomes of the State curriculum framework. Discussion of the need for, and the structure of, these assessment tasks led to an interesting professional debate. In some situations, teachers raised questions about the increased workload this change in assessment required. In most cases, however, it was not that assessment tasks had to be created and marked which was new, but the requirement for the content and tasks to be common between classes studying the same course material. There was some additional work for the teachers to participate in moderation of the results.

The Design and Technology faculty drew up a task that Year 8 students completed at the end of their semester of study in this area. It was designed to assess all of the desired outcomes of the course and was completed across a number of lessons. This model promoted discussion among the curriculum coordinator group and inspired interest within other learning areas in devising similar tasks. Mathematics and Science already used common tests and these results were correlated against the ability data.

This project is still at a very early stage; however, an examination of Year 8 Mathematics data in Figure 12 ($r = 0.656$) in this school in 2005

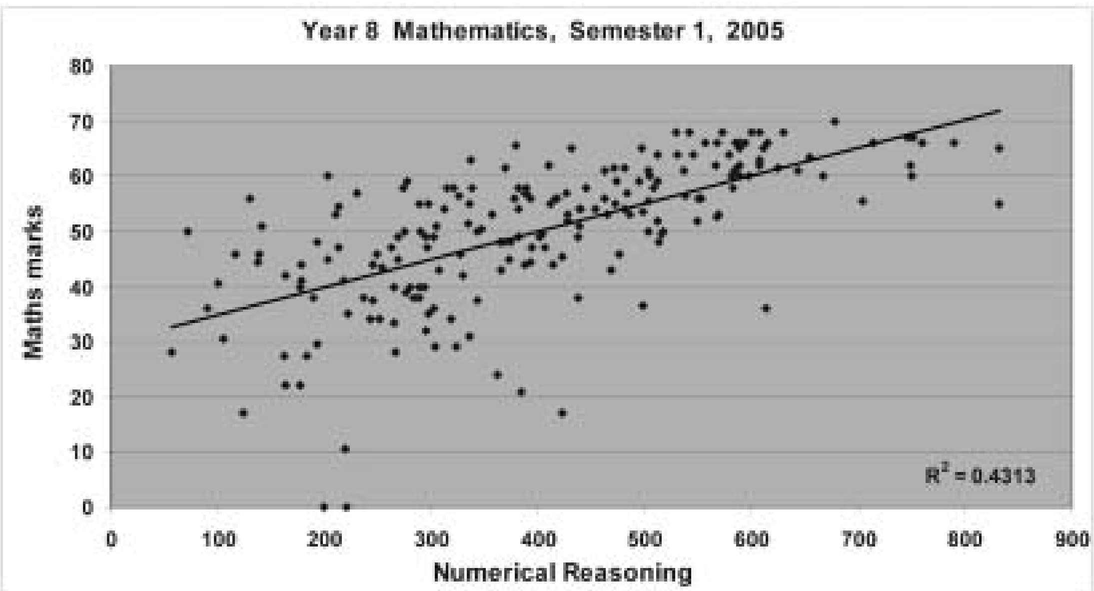


Figure 12 Scatterplot of Student Ability and Year Level Common Subject Assessment, 2005

indicates a pattern closer to the expected relationship between ability and achievement data, than for example that shown in Figure 11 (Mathematics, Year 8, 2003).

Once data from each learning area is correlated against the baseline data, teachers are asked to examine and reflect upon individual and group variations from the ability-predicted results. In some cases, there may be clear, non-school explanations for under- or over-achievement. Most research indicates that the more likely explanation for those variations is a teacher effect. Teachers are being supported to develop the expertise to analyse the data and work towards finding and addressing the possible reasons for the variation in performance, from that expected given their ability.

As an example, some of our teachers see one cause of under-achievement in our senior school as student participation in vocational education programs, which take the students out of their normally scheduled classes. This issue will not be able to be investigated with our new database for some years (until this year's Year 8 students reach Year 11 and 12); however it is an example of the enquiry that will be possible, as a result of our commitment to developing a longitudinal database.

Another issue of debate has been the reporting of the correlated data (achievement to ability). From the outset, parents were informed that the correlation of their child's achievement with their child's ability would be reported to them. The intention was to do this at the end of each semester. There was never any intention of reporting the raw data from the student profile. For the ability section of the profile, we did not want to foster views such as: 'My child is top (or bottom) of the class'. The discussion we want to have with parents and students is about how well the student is achieving in relation to their own ability-predicted achievement. For the attitudinal data, we have offered to report this to parents in a face-to-face discussion with a staff member to fully flesh out the implications of this information.

The intention was to ensure consistency in achievement reporting across the year level, as well as to build capacity for monitoring each student's progress throughout their time in the school in ability-adjusted terms, commonly referred to as the value-added contribution of the school to student academic achievement.

We did not meet our goal of reporting to parents at the end of the semester. This was because it became clear that considerable professional discussion still needed to occur for the teaching staff to feel comfortable with their capacity to answer questions from parents and students regarding these data. There are other sensitivities as well. For example, if all results are available to staff, how will performance of individual learning areas be seen by staff of other learning areas? If we continue to send home student grades as well as the correlated ability and achievement data, will parents question results that may appear anomalous? Some teachers are finding the stated expectation that they can influence the quality of students' learning outcomes to be, at least to some degree, quite confronting.

As we proceed with this project, opportunities and questions continue to arise. Our commitment is to run this project for five years at least. In 2005, it is only our Year 8 students who are involved. As they move through the school, achievement data will continue to be correlated against the intake ability data. There are several questions as yet unanswered: How will we best represent a student's Year 9 achievement levels against their Year 8 levels? Is there a meaningful way to do that? Will the results of each learning area be correlated against each ability strand separately, or only against the general reasoning or some combination of these?

The point in conducting this research is to lead to our teaching processes being data-informed and as a consequence, more effective. We believe the focus on the student's own real learning progress will contribute to improved relationships between the teacher, student and parents.

Conclusion

Schools have extensive data – as student records and reports, in staff offices, administration areas and archives, and of course, the vital information carried in teachers' heads. Some secondary schools have intake data or scholarship results, but few schools have gathered the comprehensive data required for effective monitoring of student achievement, as identified in the research discussed above.

A measure of student ability, against which to evaluate student attainment via common tasks and moderated subject assessment for each year level, is essential to provide schools with the capacity for ability-adjusted monitoring of each student's learning progress, at regular intervals. The explicit purpose for developing a school database is to support learning at all levels within the school, so that trends over time can be identified, with early detection of issues leading to remediation and extension

However, managing the school's academic database so that appropriate information can be readily accessed when needed requires time, skill and financial commitment. Planning is essential to ensure that all data are formatted and integrated, as students, teachers, year level co-ordinators, curriculum and welfare staff, administration and management all require different analyses, report formats, and levels of access. In addition, security, regular updating and archiving of information also require attention.

To begin the task of value-added monitoring of student and school performance, the following steps are recommended:

Step 1. Enter all current educational data available in the school in relevant spreadsheets in the school's academic database.

Step 2. Arrange for appropriate analyses to be conducted, with output formatted to ensure user-friendly access to, and interpretation of, all tables, graphs and summary information.

Step 3. Use this school database to initiate *informed* discussion and debate around the following questions:

- Do these data provide us with answers to the questions we have been asking?
- What questions remain unanswered, and what further data are needed to respond to these?
- What new questions have emerged?

The challenge for school staff when reviewing their data is to identify the factors that affect student learning, both positively and negatively then adjusting their practice accordingly. This research indicated that school leaders and teachers need considerable time to examine their value-added data. The graphs and tables of data provided to each school on their CD encouraged staff to reflect on the student and class patterns, and to discuss and debate their attributions for the factors impacting student and class scores. The capacity to highlight each student within the school, examine performance in depth at individual student, class and subject levels, as developed in this research, was new to, and positively received by, school leadership teams. It is hoped that this research can be extended in the future to include student feedback, as it would be of interest to record the factors students considered were major influences on their results, and whether they believed they had achieved to their potential.

However, the extensive, multi-level factors that affect student learning have yet to be definitively identified and modelled. For this to occur, a re-allocation of resource provision at system level is needed, so that valid and reliable curriculum measures and assessment protocols are available in all schools. Clearly defined subject knowledge and skills, in appropriate developmental stages with common metric assessment scales, would enable teachers to report valid learning gains for each student.

When contextual information and data-informed interpretation are lacking at the level at which the data were gathered, class and subject residuals are often misrepresented as evidence of teacher performance. Principals and teachers are rightfully concerned about their performance being judged by the type of data analyses and displays of VCE results similar to that displayed in Figure 10, and currently used at system level in Victoria.

Improved collaboration between researchers and practitioners can lead to more truly *data-informed* analyses, if the voices of *all* stakeholders are represented, and not dominated by system-level statistical analyses that are not independently verified. It is possible to conduct more equitable evaluation of student and school performance, both across and within schools (Richardson, 2004a). This VCE research found support for Rowe's (2003) statement that 'All too frequently *systems, schools and teachers* (my emphasis) have lacked credible information regarding the magnitude of their relative contributions to performance and effectiveness'. Greater effort needs to be focused on research within schools where it is possible to validly identify the factors influencing students' achievements for both boys and girls and for low, average and high ability students.

School effectiveness research (SER) and system-level analyses still over-emphasise teacher effectiveness, and fail to take into account the multi-level structure within which teaching and learning operate. Student responsibility for learning (at senior secondary) and system-level accountability need appropriate attention so that resources are diverted to research that has the potential to identify and verify sources of variation at student, class and teacher levels within and across schools.

We do not yet have appropriate measures of the verifiable teacher behaviours explicitly linked to student achievement that can be validly reported in terms of effect size. Claims regarding the proportion of variance explained at the so-called 'teacher' or 'class/teacher' level were not supported in this detailed, evidence-based research (Richardson, 2003b, 2004b), where interpretations were validated in discussions with senior staff within schools. This does not mean that teachers and the quality of teaching are not vitally important influences on student achievement. Just as we rarely have measures of student motivation and aspiration, time on task and degree of private tutoring, illness and personal trauma, all of which affect student performance, so too, we do not yet have the comprehensive data needed to identify the teacher behaviours and attitudes that positively impact student performance across all ability levels.

For more authentic evaluation of teacher performance, and calculation of *genuine teacher effects*, valid measures of teaching knowledge, skills and behaviours demonstrated to make a positive difference to student achievement, are needed. While some research quotes 'characteristics of effective teachers' (Sammons, 1999) no definitive studies have measured these variables over time. The Hay-McBer (2000) research on teacher effectiveness provides one way of conceptualising a matrix of factors that could be modelled to further our understanding in this area.

Many research reports aggregate one or more student measures to create second-level variables, then discuss this aggregated group variance, be it class/subject or school-level, in terms of the 'class/teacher effect', sometimes the 'teacher effect'. Results from 'within-school' analyses in this research indicated that even when student data are aggregated to class level (or subject level) it is misleading to name this as the 'teacher effect'.

The value added by the school is usually estimated in terms of student and group performance above that of their peers. Yet it is rare for all academic characteristics such as ability, past performance in the subject area, teaching and learning strategies, and contextual variables such as gender and SES at student and school levels to be comprehensively measured. This level of data is just not available yet in Australia.

Of all the States in Australia, because of the ability and achievement measures collected for the VCE, Victoria has the greatest potential to take the lead in developing research to identify positive teaching and student learning effects (Richardson, 2004a). One way that this could be achieved is, for example, if the Hay-McBer (2000) Teacher Effectiveness variables were measured and integrated with the type of 'within-school' analyses described above. More than 100 Victorian schools have already demonstrated their commitment to 'using data to improve learning'. It is now time for both the Federal and State Governments to collegially support and extend this research.

References

Hattie, J. A. (2003). *Teachers make a difference: What is the research evidence?* The Australian Council for Educational Research, (ACER) Conference, 2003, Building Teacher Quality: What does the research tell us? Melbourne, October, 19-21, 2003.

Hay-McBer. (2000). Research into Teacher Effectiveness. A Model of Teacher Effectiveness. Research Report, No 216. DfEE, U.K.

Hill, P.W. & Richardson, C. A. (2001). *The role of contextual effects in predicting the academic progress of students in Hong Kong secondary schools.* Presentation at the 14th International Congress for School Effectiveness and Improvement (ICSEI) Conference, Copenhagen, January 3-6, 2001.

Hill, P.W. & Rowe, K. J. (1996). Multilevel modelling in school effectiveness research. *School Effectiveness and School Improvement,* 7 (1), 1-34.

Hill, P.W. & Rowe, K. J. (1998). Modelling student progress in studies of educational effectiveness. *School Effectiveness and School Improvement, 9* (3), 310-333.

Mortimore, P., Sammons, P., Stoll, L. Lewis, D. & Ecob, R. (1988). School matters: the junior years. In P. Mortimore, *Effective schools: current impact and future potential*. Director's Inaugural Lecture, Institute of Education, University of London. February 7, 1995.

Richardson, C.A. (2000a). *Stability and consistency in students' VCE achievements*, 1994-1998. Presentation at the 13th International Congress for School Effectiveness and Improvement (ICSEI): "Global Networking for Quality Education", Hong Kong, January 4-8, 2000.

Richardson, C.A. (2000b). *Positive teacher effects at VCE level: Evidence, observation and reflections.* Presentation at the Annual Postgraduate Students' Conference, Faculty of Education, The University of Melbourne, December 2, 2000.

Richardson, C.A. (2001). *The Role of the General Achievement Test (GAT) in VCE results, 1994-2000*. Presentation at the Annual Postgraduate Students' Conference, Faculty of Education, The University of Melbourne, December 2, 2001.

Richardson, C.A. (2002). *The value of value-added analyses of VCE results*.

Presentation at IARTV Invitational Seminar, Melbourne.

Richardson, C.A. (2003a). *The data-informed school: Positive use of student information for teachers*. Presentation at the Australian Council for Educational Research (ACER), November, 2003.

Richardson, C.A. (2003b). *Fairer methods of assessment: Value-added analyses of final secondary results*. Presentation at the 15th International Congress for School Effectiveness and Improvement(ICSEI), Sydney, Australia, January, 2003.

Richardson, C.A. (2004a). *Effectiveness and Quality Assurance in Education*. Presentation to the Board, Victorian Schools Innovation Commission, 22nd July, 2004 Melbourne.

Richardson, C.A. (2004b). *Do teachers make a difference at VCE? Exploring empirical evidence and reflecting on findings*. Presentation to the 9th Annual Roundtable: Assessing Assessment, 7-9 November, 2004, Crowne Plaza Sydney.

Rowe, K.J. (1999). *VCE Data Project: 1994-1998. Concepts, Issues, Directions and Specifications.* Melbourne. Victorian Board of Studies.

Rowe, K. J. (2003). The importance of *Teacher Quality* as a key determinant of students' experiences and outcomes of schooling. ACER Conference, Building Teacher Quality: What does the research tell us? Melbourne, 19-21 October, 2003.

Sammons, P. (1999). School Effectiveness. Coming of age in the twenty-first century. Swets and Zeitlinger, B. V., Lisse, The Netherlands.

Scheerens, J. & Creemers, B.P.M. (1989). Conceptualizing school effectiveness. In Developments in school effectiveness research [Special Issue]. *International Journal of Educational Research*, 13, 691-706.

Teese, R., & Polesel, J. (2003). Undemocratic schooling; Equity and equality in mass secondary education in Australia. Melbourne, VIC: Melbourne University Press.