Constructing and Using Educational Performance Indicators[1]

Dr Ken Rowe
Principal Research Fellow, Australian Council for Educational Research
and
Professor Denise Lievesley
Director, UNESCO Institute for Statistics, Montreal, Canada

Background paper for Day 1 of the inaugural
Asia-Pacific Educational Research Association (APERA) regional conference

**Educational Performance Indicators:
Their construction and use**

ACER, Melbourne April 16-19, 2002

Abstract:

Within the context of a growing international movement toward the adoption of 'outcomes-based' modes of educational provision and governance, this paper focuses on the context, nature, purpose and construction of educational performance indicators. Key issues related to types and sources of indicators, essential features of useful indicators, constructing and using performance indicators to advantage, and key elements of Education Management Information Systems (EMIS) are presented and discussed.

---

[1] Correspondence related to this paper should be directed to Dr Ken Rowe, Research Director (Learning Processes & Contexts), Australian Council for Educational Research, 19 Prospect Hill Road (Private Bag 55), Camberwell, Victoria 3124, Australia; Tel: +61 3 9277 5584; Fax: +61 3 9277 5500; Email: rowek@acer.edu.au; Web: www.acer.edu.au

THE NATURE AND PURPOSES OF EDUCATIONAL PERFORMANCE INDICATORS: International and national perspectives

What are performance indicators?

In general, *performance indicators* (PIs) are defined as *data indices of information* by which the functional quality of institutions or systems may be measured and evaluated. Typically, within the context of specified goals and objectives, PI data are 'measures' of various operational and functional aspects of organizations and/or systems, and provide evidential bases for determining the extent to which such goals and objectives have and are being met. PIs serve various purposes, the most notable of which are for **monitoring**, **policy formulation**, **target-setting, evaluating** and **reforming**. Although the essential features of educational PIs are consistent with their counterparts in other government and corporate enterprises, they also have unique characteristics – key aspects of which are highlighted in this paper. At the outset, however, it is helpful to note the importance of educational PIs in the prevailing international context.

The importance of PIs in an international context

There is a growing body of opinion throughout the international community that a crucial task of national governments, including that of local and international support agencies, is the collection of data to construct key educational PIs for monitoring purposes. Whereas the high level political significance attached to the 'declaration' of goals established at World Education Conferences such as *Education for All* is to be welcomed, it is important that they be turned to advantage – particularly to the advantage of developing countries. Since the specified indicators are dependent upon achieving responses from as many countries as possible, with keen attention being paid to the poorest countries, the gaps in data can be highlighted and resources sought to assist those countries to collect the relevant data. Such data are powerful for advocacy purposes and enable the international community to co-ordinate appropriate actions; with the targets being useful for accountability purposes.[1]

However, through the mechanism of world conferences, goals chosen without reference to their measurability or other logistic attributes, are often endorsed by countries before there has been an opportunity to consider characteristics such as:

- Are they realistic?
- Can they be measured?
- Will they show sufficient change over time to be valuable as yardsticks?
- Have relevant benchmarks been specified?
- Are the goals universal, and do they make sense in different parts of the world?
- Are they coherent with what has been used in the past?

Following such 'declarations', goals cannot be changed easily. Typically, the translation of goals into indicators is viewed merely as a technical process. Too rarely is there recognition that both conceptual development and fieldwork are needed to develop reliable indicators. Since indicators are often expressed in terms of change over time (e.g. halving adult illiteracy by a specified date), the unquestioned assumption of the availability of adequate baseline data can be problematic.

Nevertheless, the selection of appropriate goals is critical since ideally, such goals should be realistic and achievable. It has been proposed that countries should be stretched

to achieve them, but if they are too difficult to reach this can be detrimental since it perpetuates a sense of failure. It is not easy to see how this can be managed with a set of identical indicators for all countries of the world, which are of course at very different stages of development. Skeptics of education indicator programmes often cite the fact that indicators produced from goals 'declared' at World Conferences are unrealistic for many of the developing countries that endorse them, and that all specified targets have been missed consistently. For interesting perspectives on these issues, see Scheerens (2001a,b).[2]

Because of the need to collect the same data for all countries (or a large number of them) the indicators must reflect the lowest common denominator. Nevertheless, inertia in the system, together with anxiety that no changes should be made to the agreed set of education indicators, have unfortunate consequences of placing constraints on their potential for growth. Moreover, it can dissuade countries from developing new indicators and entrench the selection made at a particular point in time. Incidentally changing an indicator is often confused with 'changing the goal posts' even though the indicator may have been, in the first instance, an inadequate proxy for the goal.

Whilst recognizing the value of indicators they should not be regarded as encapsulating the whole contribution to education policy development. The dangers are twofold: First indicators should not be imbued with more meaning than is justified by their content. Second, to concentrate on PIs for monitoring and advocacy alone is to ignore the broader value of PI information in formulating evidence-based policies.

Work on indicators should therefore be complemented by support to develop rich databases about the state of education. A common and fallacious impression is that the priority indicators are the only statistical outputs a country needs – even for its own governance. A further problem, and perhaps one of the most difficult, is the dilemma between open accountability and national sovereignty in relation to what data are collected, the methods used to collect and analyze them, and who is to have access to the results.

The nature and purpose of educational PIs

During the last decade, education systems throughout the world have been subject to considerable reform and change – all justified on the grounds of **improving** the *quality* of school education. A key feature of this change has been the frequent revisions of style and policy focus, especially in the area of PIs, with major emphases being placed on the assessment and monitoring of student learning outcomes. Indeed, current policy activities related to 'outcomes-based' educational PIs and their links with growing demands for *accountability*, *standards monitoring*, *benchmarking, school effectiveness* and *reform* are widespread and well established in many developed countries.[3]

Such emphases are aptly illustrated in the reported proceedings of a recent meeting under the auspices of the *Summit of the Americas*,[4], which states:

> Although it is now part of daily life in schools and in debates between specialists, education assessment has recently become a relevant topic for governments and society, especially because of the economic crisis and the acceleration of the globalization process, which made investments in education a strategic point while the resources available for the sector have shrunk.

In many developed countries, much of this activity has been (and continues to be) directed away from concerns about *inputs* and *processes* of educational systems (e.g., physical resources and curriculum provision) to *outputs* (e.g., improvements in student achievement outcomes, as well as in school and system performance). A major effect of such activity has been to signal shifts in government policy intention to:

- encourage system accountability to ensure both efficient and effective utilization of resources, and
- bring the delivery of educational services into public sector accounting, underscored by a concern to ensure that such services represent 'value for money'.

Since schooling accounts for a significant proportion of public and private expenditure, as well as generating a substantial quantity of paid employment for teachers and administrators, the enduring interest by governments in the relative *performance* of school education is not surprising. This is an especially sensitive issue at the present time given the level of consensus regarding the importance of school education as an element of micro-economic reform and in meeting the constantly changing demands of the modern workplace in the context of increasing globalization of the world economy.[5]

Whereas the provision of quality education is critical to the development of all countries, it is especially the case for developing countries where there is considerable pressure to increase access to education, but not at the expense of quality. Hence, the demand is to ensure that PIs do not provide a partial, and thus potentially misleading picture of either *quality* or *effectiveness*, as has often been the case in the past.

Despite the difficulties entailed in defining *educational effectiveness* at the school or system level, and reaching consensus on the relevant criteria, a good deal of discussion has focused on what is meant by *quality schooling*, and how it might be measured and improved. Although the term *quality* is likewise problematic, the "...measurement of the *quality* of schooling is of critical importance at a time when so much school reform in so many parts of the world is being undertaken."[6] In fact, concerns about the *quality* of school education and its monitoring have long been high priority policy issues in all OECD countries.[7]

An illustration of this priority is evident in recent developments by Nepal. Following its endorsement of the 1992-2000 plan of action at the Jomtien Conference in 1991 to implement the *World Declaration on Education for All* (EFA) and *The Framework of Action to Meet Basic Learning Need*s, Nepal prepared its own National Plan of Action (NPA) in order to achieve the EFA goals. These goals were: universal primary education and literacy, policies to promote basic and primary education, and attainment of education for all. As a prelude to specifying targets of the NPA, the Nepalese state:[8]

> Learning achievement is one of the most important measures of the quality of education. It is also intricately linked to school efficiency because the promotion and repetition rates are directly related to the learning achievements of the students, to which in turn school drop-out can be attributed.

In a similar vein, Manno[9] asserts:

> When judging educational quality, either we focus on what schools spend – or one of its many variants – or we focus on what students achieve, what they know and can do. Those who advocate a focus on outcomes in judging educational quality hold one common belief: we must specify what we expect all children to learn, and we must assess them to determine whether they have learned it.

While measures of student learning outcomes are prime PIs of education systems and the services they provide and for which they are responsible, there are many others (including both *inputs, processes* and *outputs*) that constitute useful bases for informed planning and decision-making, followed by implementation and reform (see below). If decisions for improvement are to be **informed** rather than based on political whim or ideology, then useful, dependable and timely information on indicators is required. Indeed, such bases constitute key purposes of specifying, gathering and using PIs for educational change and reform. In particular, PI information allows systems and their constituent

organizational elements to: (1) formulate strategic policy priorities and their related targets, (2) specify achievable objectives, (3) implement them, and (4) evaluate the extent to which those target objectives have been attained.

Thus, as indicated earlier, *performance indicators* (PIs) are usefully defined as *data indices of information* by which the functional quality of institutions or systems may be measured and evaluated, providing evidential bases for determining the extent to which specified goals and targets are being achieved. PIs serve various purposes, the most notable of which are for **monitoring**, **policy formulation**, **target-setting, evaluating** and **reforming.**

Types and sources of performance indicators

The types of input-output PIs are many and varied. Among the major educational PIs that may be collected include:

- Indicators of resource provision and funding, specified against stipulated targets;
- Access rates at: pre-school, elementary, secondary, vocational and tertiary levels – per capita of age/stage cohort population, and inequities in access to education;
- Participation rates in education at all levels, barriers to participation, characteristics of children out of school;
- Repetition rates and completion of five years of schooling;
- Percentage of GDP devoted to education;
- Per capita costs at each of these levels;
- Class sizes; teacher:student ratios;
- Provision of teacher training and participation in in-service professional development;
- Measures of cross-sectional, age/grade-level, cohort and student achievement outcomes in core curricular;
- Longitudinal achievement progress indicators and measures of factors affecting students' progress rates;
- Measures of impact of strategic interventions for students with special needs and those from disadvantaged backgrounds.

The types and range of educational PIs that many countries collect are well illustrated in the UNESCO *Education For All* (EFA) *Year 2000 Assessment* program web site at: http://www2.unesco.org/efa/wef/countryreports/country.html.

Leading up to the World Education Conference in Dakar in April 2000, the 18 Core EFA indicators and related targets for **improvement** by the Year 2000 were specified as follows:

Indicator 1: Gross enrollment in early childhood development programs, including public, private, and community programs

Indicator 2: Percentage of new entrants to primary Grade 1 who have attended some form of organized early childhood development program

Indicator 3: Apparent (gross) intake rate in Grade 1 as a percentage of the population of official entry age

Indicator 4: New entrants to primary Grade 1 who are of the official primary school entrance age as a percentage of the corresponding population

Indicator 5: Gross enrollment rate (Grades 1-5 total)

Indicator 6: Net enrollment rate (Grades 1-5 total)

Indicator 7: Public current expenditure on primary education as: a percentage of GNP; and per pupil as percentage of GNP per-capita

Indicator 8: Public expenditure on primary education as a percentage of total public expenditure on education

Indicator 9: Percentage of primary school teachers having the required academic qualifications

Indicator 10: Percentage of primary school teachers who are certified to teach according to national regulations

Indicator 11: Teacher:pupil: ratios

Indicator 12: Repetition rates at Grade 1 and 5

Indicator 13: Survival rate to Grade 5

Indicator 14: Coefficient of efficiency to Grade 5

Indicator 15: Achievement Test Scores and Basic Learning Competencies in Language, Mathematics and Social Studies

Indicator 16: Literacy rate of population 15-24 years old

Indicator 17: Literacy rate of population 15+ years old

Indicator 18: Gender parity index (female to male literacy rate)

The following six goals were established at the *World Education Conference 2000* in Dakar:

(i) expanding and improving comprehensive early childhood care and education, especially for the most vulnerable and disadvantaged children;

(ii) ensuring that by 2015 all children, particularly girls, children in difficult circumstances and those belonging to ethnic minorities, have access to and complete free and compulsory primary education of good quality;

(iii) ensuring that the learning needs of all young people and adults are met through equitable access to appropriate learning and life skills programmes;

(iv) achieving a 50 per cent improvement in levels of adult literacy by 2015, especially for women, and equitable access to basic and continuing education for all adults;

(v) eliminating gender disparities in primary and secondary education by 2005, and achieving gender equality in education by 2015, with a focus on ensuring girls' full and equal access to and achievement in basic education of good quality;

(vi) improving all aspects of the quality of education and ensuring excellence of all so that recognized and measurable learning outcomes are achieved by all, especially in literacy, numeracy and essential life skills.

Consultations are currently underway to determine the appropriate indicators to measure these goals, and to ensure their coherence with the Millennium Declaration Goals.[10] Note that while these are PIs at the country-level, they are also needed at the international level. In Dakar 2000 it was determined that no country should be thwarted in their achievement of these goals through a lack of funds, provided that they have a credible education plan. Two key questions to answer are:

1. How does one determine whether or not an education plan is 'credible'?, and
2. How does one determine whether a country has adequate funds?

These indicator/targets for improvement include: resource funding provisions; student access, participation and retention rates; teacher qualifications, training and certification; student achievement progress; and population literacy rates.

However, one of the main problems for the creation of indicators in many countries is the lack of correspondence between the population data and the education data. For example, in Saudi Arabia, the education data collected provide counts for Saudi children only, whereas the population data include resident non-Saudi children.

Sources from which PI data may be obtained are inherently multilevel and multi-faceted. That is, data can be gathered from multiple levels of a system, namely: student, class, school, district, region or province, state, national and international – such as the *Third International Maths and Science Study* (TIMSS).[11] The data may also be gathered from

administrative data, school surveys, household surveys and population censuses. The ways such data are gathered range from rudimentary manual methods to sophisticated computer-based management information systems implemented by governments and their supporting bureaucracies (see below). In the latter case, the rapid development of information and communication technology, increased pressures to 'measure' student, school and system performance, are major factors that have influenced the development of powerful education management information systems (EMIS). The purpose and role of EMIS are discussed in more detail later.

ESSENTIAL FEATURES OF USEFUL INDICATORS

A *useful* performance indicator (PI) is one that informs the processes of strategic decision-making and taking – resulting in measurable improvements to desired outcomes following implementation. Similarly, the *quality* of a PI is comprised of many components including:

- Validity;
- Reliability;
- Relevance to policy;
- Potential for disaggregation (e.g., by gender, socioeconomic and ethnic groupings, education administrations, etc.);
- Timeliness (i.e., currency and punctuality);
- Coherence across different sources;
- Clarity and transparency with respect to known limitations;
- Accessibility and affordability (i.e., cost effectiveness);
- Comparability through adherence to internationally agreed standards;
- Consistency over time and location; and
- Efficiency in the use of resources.

The optimum combination of these components is dependent upon the use to be made of the data. Data acceptable for one purpose might be inadequate for another and, since most data are used for many different purposes, the process of determining 'fitness for purpose' is extremely complex and requires wide consultation. The features of five of these characteristics of useful PIs, are outlined in more detail below.

Relevance: Judgments related to the *relevance* of a given PI depend on the *purposes* for which is it gathered and *how* it is used to inform policy, planning, practice and reform. Moreover, the *relevance* of any PI is location-specific and context-dependent in terms of prevailing policy priorities and demands for information. In general, however, a PI is deemed to be *relevant* if it provides *useful* information for strategic decision-making and decision-taking. For example, a key guiding principle of the *UNESCO Institute for Statistics* (UIS) in their work of supporting PI data-gathering in 189 member countries and states is that PI data **should not** be collected for their own sake, but rather, because they are needed for specific policy purposes. In this regard, a visit to the UIS web-site is helpful, at: http://www.unescostat.unesco.org/

Cost-effectiveness: Regardless of the perceived *usefulness* of particular indicators, cost-effectiveness and logistic feasibility are important considerations that need to be taken into account. In the case of indicators of students' achievement outcomes, for example, the cost and feasibility of obtaining estimates derived from full cohort or population data collections may be unjustifiably great compared with those obtained from appropriately designed samples. Decisions about the cost-effectiveness of PIs, however, must be balanced against considerations of their utility to inform policy, planning and reform.

Timeliness has two components: *currency* and *punctuality*. Indeed, an important characteristic of the usefulness of PIs is their availability at times when key policy and planning decisions need to be made. At such times, the absence of PI information often leads to misinformed enterprises that have a tendency to rely on opinion rather than evidence. Whereas the relevant information for some PIs require longer periods to collect and analyze (e.g., student achievement progress rates), findings at key stages of the data collection should be reported to inform policy makers and planners of possible trends and other PI factors affecting those trends.

Reliability: Determining the *reliability* of a PI involves evaluating how *accurately* it has been *measured*. This is a crucial technical issue for the formulation and interpretation of PI information that is frequently overlooked by gatherers, purveyors and consumers. Rather, obtaining and reporting evidence concerning the reliability and sources of measurement error for PIs are fundamental responsibilities of PI developers. The same applies to large-scale monitoring procedures employed in national or system-wide testing and public examination systems that involve the estimation of composite scores from multiple modes of assessment. At the very least, evidence about the uncertainty associated with observed scores is required to minimize the potential 'risks' of misinterpretation.

Validity: In the present context it should be noted that an estimate of the *reliability* of a PI is not necessarily commensurate with its *validity* – both *content validity* and *criterion-related validity*.[12] While it is possible to have a highly reliable PI that lacks validity (e.g., an assessment task), a *valid* PI that has low reliability is of little or no value. For example, conclusions about students' achievements are *valid* only when measured *reliably* and based on evidence about intended and achieved learning outcomes. Nonetheless, the c*ontent validity* of an indicator – including its *face validity* and *logical validity* (see Note 12) – may only be established via a rational analysis of its content and utility, based on subjective judgment, albeit by consensus.

In sum, *useful* PIs are those that are *relevant*, *cost-effective*, *timely*, *reliable* and *valid* – in terms of their capacity to inform the processes of strategic decision-making and decision-taking – resulting in measurable improvements to desired outcomes.

CONSTRUCTING AND USING PIs TO ADVANTAGE

Among the justifications for system-wide, student achievement monitoring programs is that the provision of accurately measured, responsibly analyzed and presented PI data has the potential for generating ***improvements*** in teaching, as well as in students' learning and achievement outcomes. If such improvements are to be realized, access to such data – particularly those measured on common scales over time – is essential. However, in the absence of explanatory, 'value-added' indicators (see below), any improvement strategies adopted are likely to be unproductive since mere *measurement* and *location* of students on performance scales does not, *ipso facto*, generate improvement, regardless of how well or often student progress is measured and monitored. Moreover, estimation of the effects of factors influencing students' educational outcomes and the identification of major sources of within- and between-school variation are not possible.

Nonetheless, one of the most promising features of the current international emphasis on 'outcomes-based' educational PIs is the emerging effort by several countries (and state/province systems within them) to focus on *gains* or *improvements* in student performance rather than simply attainment levels at a given point in time. In this context it is worth noting the increasingly popular approach of PI feedback services such as those offered by the Curriculum, Evaluation and Management Centre (CEM) at the University of Durham (UK). In recent promotional literature published by the CEM,[13] it is claimed

"…we have become the largest provider of **performance indicators** to schools and colleges in the world" (their emphasis). The promotional information continues thus:

> We develop, produce and provide tests and questionnaires to be completed by students under standardized conditions. We analyze these and provide clear graphical feedback, and comparisons with many hundreds of other schools and colleges. Data on pupil progress (value-added) is provided when outcome measures become available. At each stage we try to measure what matters, be it attitudes, safety, relationships, learning and teaching processes, etc. Having high quality, externally analyzed data, with fair comparisons, greatly assists schools in validated self-evaluation and management.

These PI services are provided to participating schools throughout the years of schooling, ranging from: Performance Indicators in Primary Schools (*PIPS Baseline Assessment* for 5 Year-old children); *PIPS* (for children aged 6-11 years); the *Middle Years Information System* (*MidYIS* for 12-14 Year-olds), *the Year 11 Information System* (*Yellis* for 15-16 Year-olds); and the *A-level Information System* (*Alis*+ for 17-18 Year-olds). These Information Systems are illustrated in Figure 1 below.

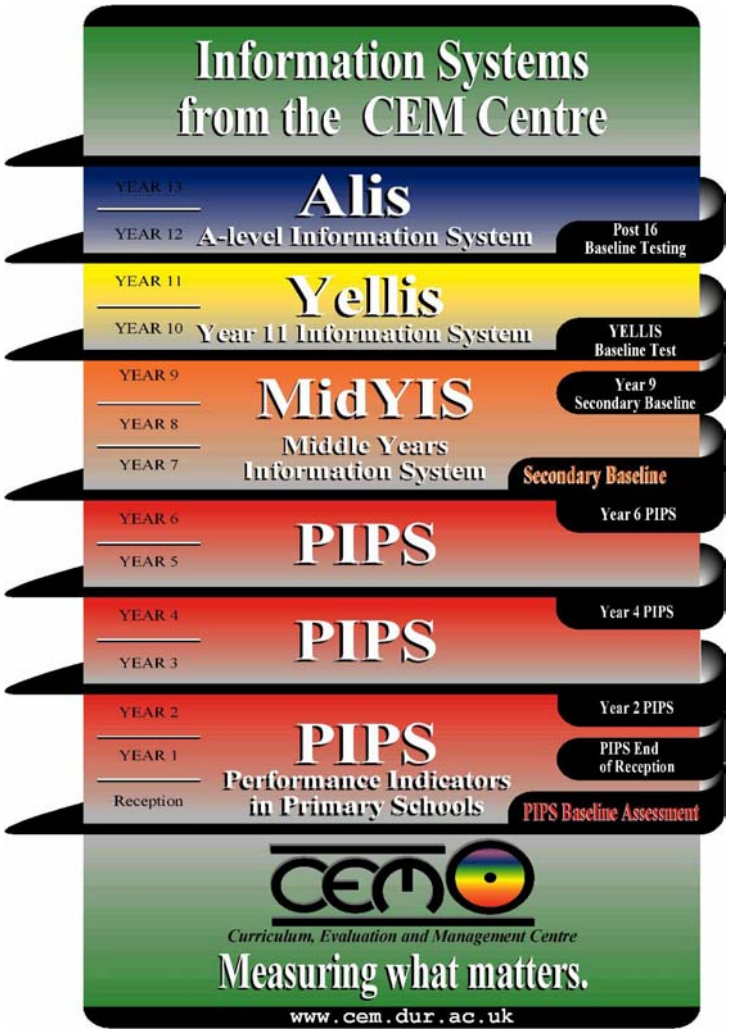

Figure 1 CEMs PI Information Systems

Although such PI services to schools have the potential to be powerful and effective modes of engendering strategic improvements and reform, their inherent weakness lies in the fact that even with the availability of repeated measures on the same students at different stages of their schooling, the outcome measures derive from assessments of

performance on **discrete** standardized tests. In other words, since the items from each of the tests across the Performance System levels have not been calibrated on a common measurement scale,[14] claims of 'value-added' ***progress*** cannot be made legitimately. Moreover, any explanatory modeling of the performance data in terms of factors affecting student 'growth' or 'rates of progress' for policy-related intervention purposes, is neither possible nor justifiable.

In contrast to this approach, a feature of ACER's *Longitudinal Literacy and Numeracy Study* (LLANS)[15] is that since students' performances in Literacy and Numeracy are measured on common, qualitatively-described developmental scales, it is possible to model 'growth' and determine the magnitude of factors affecting that growth. Figure 2.1 illustrates and describes the Literacy progress made by one student during the first two years of schooling (i.e., 2 assessment occasions per year), and Figure 2.2 presents the modeled growth curves for 10 students in one class over the same time period.

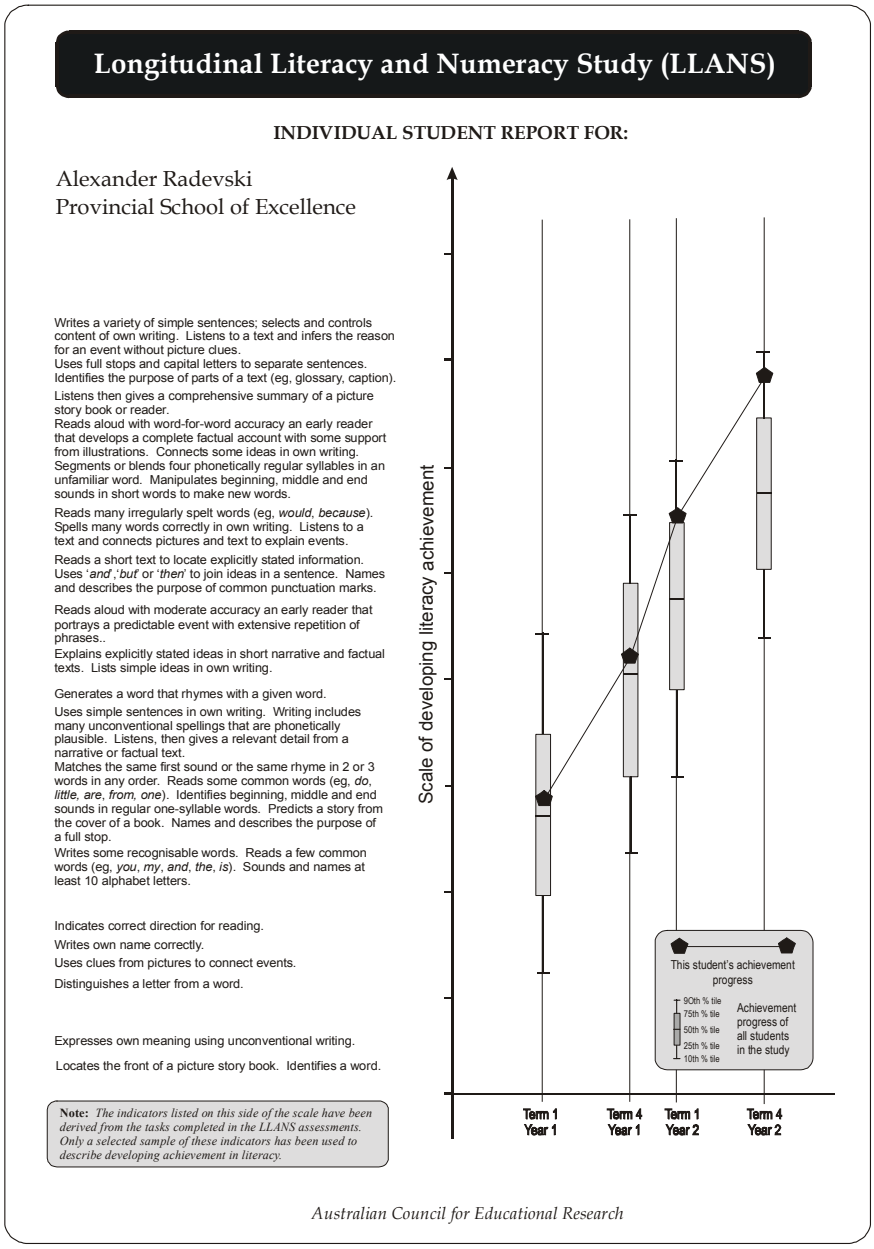

Figure 2.1 Report of individual student progress on LLANS Literacy Scale

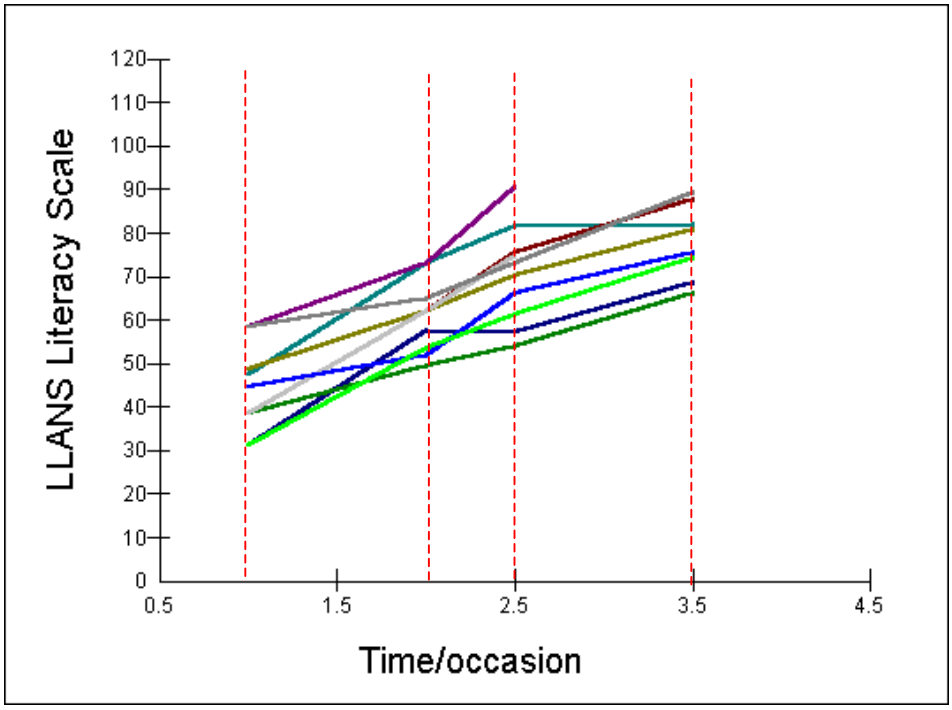

Figure 2.2 LLANS Literacy growth 'curves' for 10 students in one class, adjusted for 'intake' characteristics

'Value-added' educational PIs

Whereas the notion of 'value-added' is common to the parlance of *accountability*, *monitoring*, *performance indicators* and *school improvement*, it is not well understood, and until recently, little use of 'value-added' measures has occurred outside of research projects.[16] Nevertheless, with increasing recognition of the power of such information to motivate and shape improvement efforts, this situation is beginning to change rapidly, and warrants comment here – albeit briefly.

Based on fundamental substantive and methodological considerations, researchers have developed ways of statistically adjusting examination and/or test score data to take into account factors such as the prior attainment and 'ability' of students, and a range of student and school background characteristics. It is widely recognized that such adjustments provide a more balanced and equitable picture of the performance of individual students and the schools in which they are enrolled – at least in terms of achievements measured by scores on standardized tests and public examinations (see references cited in Note 21).

These measures have been described as 'value-added' because they provide an indication of the educational 'value' that schools 'add' to students' achievements – over and above that which could be predicted given the backgrounds, abilities and prior achievements of their student intakes. A variety of 'value-added' measures can be constructed to indicate different aspects of educational effectiveness. At least three kinds of 'value-added' measures can be identified, namely:

1. *Unpredicted Achievement* (achievement level adjusted for family background factors and student ability);
2. *Learning Gain* (student achievement level adjusted for initial achievement level);

3. *Net Progress* (student achievement level adjusted for family background factors, ability and initial achievement level).

Several researchers have advocated the use of 'value-added' measures of *net progress* in which the key control variables are prior attainment measures obtained as soon as possible after students commence at a school, with background and ability measures being used to further adjust schools' raw results.[17]

Example of PI feedback to schools leading to improvement

The following section describes a project conducted in the State of Victoria (Australia) to engender within-school improvements in teaching, learning and student achievement via the provision of performance feedback data from the Year 12 *Victorian Certificate of Education* (VCE) assessment program. Further examples of the use of assessment data for school improvement purposes are given by Coe and Visscher (2002), and by Yang, Goldstein, Rath and Hill (1999).[18]

The project described here is known as *The VCE Data Project: An information service about Year 12 student and school performance on the VCE, across studies and over time*. A more detailed account of the project is given in the chapter by Rowe, Turner and Lane (2002) – see Note 20. A key aim of the project is to provide schools with student performance feedback data presented in forms that are designed to assist school leaders to monitor and **improve** the effectiveness of their teaching and learning programs, as well as their students' achievements on the VCE. The PI data base consists of 'ability'-adjusted student achievement scores for up to 53 subjects of the VCE, obtained over 5 years (1995-1999) from approximately 65,000 students annually, located in more than 600 schools in three sectors: government, Catholic, and independent.

The basic premise underlying the notion of *improvement* in this project is that feedback to schools of responsibly analyzed and presented student performance data is **necessary** to identify 'strengths' and 'weaknesses' in teaching and learning programs, as well as in students' achievements. Whereas feedback of such PI data is a *necessary* condition for subsequent improvement, access by schools to the data *per se* do not constitute a *sufficient* condition to bring about change for the better. Rather, such change is crucially dependent on careful and responsible management of performance information by school administrators and leadership teams within the context of a shared commitment to strategic, continual improvement among all stakeholders within a given school community.

The feedback to each participating school is via a specially designed, interactive computer software package that contains the schools' **cognitive achievement data** for up to 53 subjects of the VCE – the normalized scores of which are each adjusted for measures of general 'ability' (as measured by a *General Achievement Test*) at the individual student level (level-1), and for the within-school average 'ability' of students undertaking a given subject (level-2). Data are provided for each VCE subject offered at the school over a five-year period. Brief statistical model specifications for these analyses are provided in Appendix 2 of the chapter by Rowe et al. (2002 – cited in Note 20).

Using the interactive software package referred to above, the data for each subject are presented graphically in the form of 'ability'-adjusted means of school-level **residuals** (see Note 22), bounded by 95% confidence (uncertainty) intervals – as illustrated for one cohort year in Figure 3.1.

Additional plots are also available for 3-year and 5-year trends for multiple subjects, as illustrated below in Figure 3.2. Such plots may be selected from an interactive menu, for: all students, females in all female classes and/or schools, males in all male classes and/or

schools, females in coeducational classes and/or schools, and males in coeducational classes and/or schools

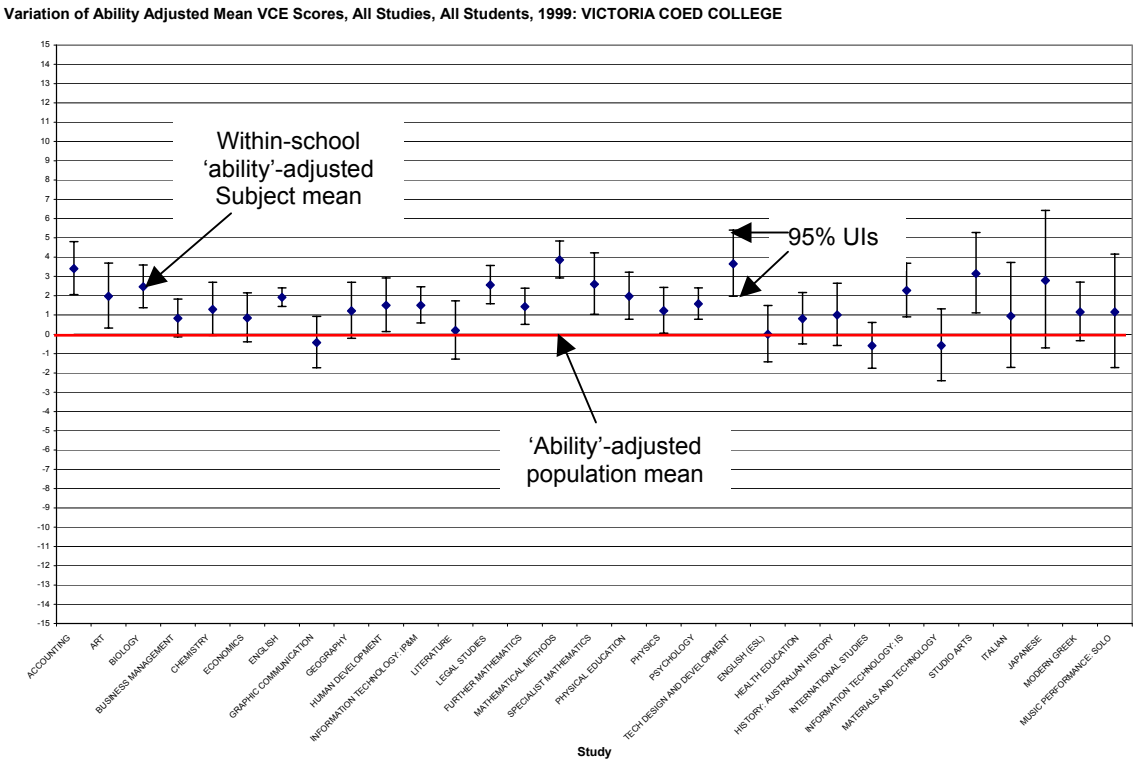

Figure 3.1 Residuals plots for 31 Year 12 subjects in one school
(Note: the bolded zero line is the 'ability'-adjusted population mean)

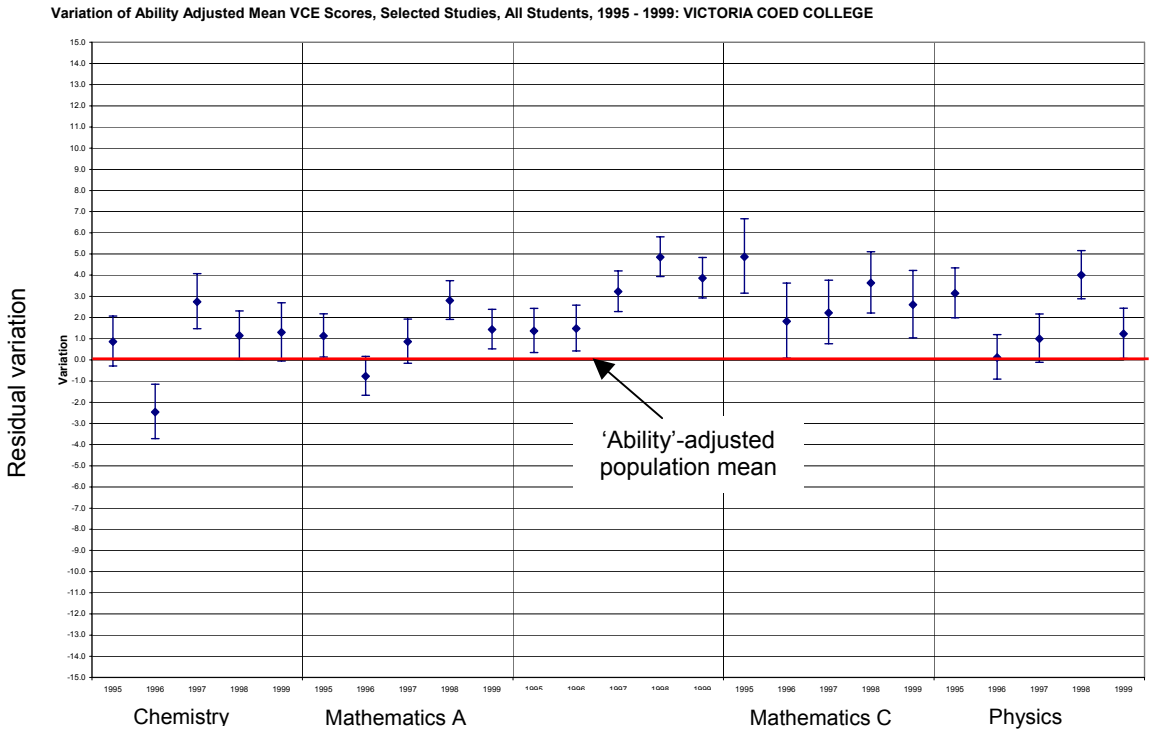

Figure 3.2 Residuals plots for five YEAR 12 subjects in one school, over five years (1995 – 1999)

A key feature of the data shown in Figures 3.1 and 3.2 is illustrative of the general finding that there is greater variability *within* schools than *between* schools. Whereas there is systematic and significant residual variation in students' 'ability-adjusted' scores due to differences between schools (ranging from 10%-21%), there is considerably greater variation in aggregated subject scores within-schools – both across subjects in any one year and over time. These estimates at the subject-level range from 54% to 59% of the residual variance, after also adjusting for the effects of gender and school sector.

By any criterion, the PI data arising from such performance feedback to schools are sensitive – particularly in respect of their potential to identify differential *teaching and learning 'effectiveness'* within a given school. Thus, it is important that such information be managed carefully and responsibly. Ultimately, both the interpretation and management of the information must reside with particular schools. Nevertheless, it is vital that the available data be neither over- nor under-interpreted by those having access to the information. To this end, professional development consultations have been provided to assist participating schools to interpret their data. Further, with the assistance of Principals and senior school administrators, suggested guidelines have been made available for what constitutes 'appropriate' responses to such data, with the aim of sharing 'best practice' in teaching and management strategies that might lead to improvement.

In developing the guidelines, extensive interviews were conducted with senior management team members of schools participating in the initial pilot study, as well as those participating in the more extensive trial. With few exceptions, schools' management teams chose to restrict the information to team members only, and embarked on specific personnel management strategies that included combinations of:

- Personal counseling and encouragement of particular staff by the school's senior management team (i.e., Principal, VCE Coordinator and Curriculum Coordinator);
- Team-teaching approaches designed to engender both *peer support* and *goal congruence* – as well as to maximize and minimize the effects of 'effective' and 'ineffective' teaching practices, respectively; and
- Curriculum-specific teacher *professional development* with the assistance of key personnel from relevant Subject Associations.

To date, feedback from the project has been overwhelmingly positive. Specific details are documented in the chapter by Rowe et al. (2002 – see Note 20).

Problematic effects of published PI data

For organizations of all types there is increasing pressure to produce publicly available PIs. However, making best use of PIs is problematic since there is always a risk of them being misused. Further, if PIs are themselves flawed, there is a risk that their public dissemination will be misleading. In education, where PIs are used for both internal and public monitoring/accountability purposes, particular attention must be given to their construction, use and publication. For a discussion of the intended and unintended effects of publishing PIs at the school-level, see Visscher et al., (2000).[19]

Example of problematic effects of PIs

In the United Kingdom, for example, where recent government policies have centered on *educational accountability* and *standards monitoring,* the use of PIs have had notable impacts on schools. Foremost among these stem from the implementation since 1990 of a national curriculum, national assessment, an external school inspection system administered by the

Office for Standards in Education (OfSTED), and the publication of schools' average, unadjusted, achievement scores on tests and public examinations.[20]

As part of a general policy initiative by the British government since 1987 to promote the use of indicators by which public service institutions can be compared and their performances evaluated, the *Parents' Charter* (DES, 1991), for example, requires that comparative 'league tables' of examination and national curriculum test results be published for every educational institution (schools) and local education authority. The 'league tables' consist of schools' rankings computed from students' average achievement scores (raw and unadjusted) on national curriculum test results at ages 7, 11 and 14 years, together with similar scores for the *General Certificate of School Education* (16 year-olds) and *A-levels* (18 year-olds). The stated intention of the *Parents' Charter* is that these PI tables be used by parents to assist in choosing schools for attendance by their children.

The British government's intention in pursuing these policies is to meet presumed public demands for *accountability* and the maintenance of *educational standards*. The UK case is an interesting one that warrants some comment here – the purpose of which is to highlight the potential misuse of PIs.

In the opinion of several commentators,[21] the prevailing market ideologies that underpin such policies have fostered a climate in which competition has begun to dominate cooperation. The focus on allowing market forces to predominate makes it possible for governments and educational regulatory bodies to locate blame for 'poor performance' or 'ineffectiveness' at the local and/or school level. Since markets operate through competition in which there are 'winners' and 'losers', the designation of schools as 'effective' or 'ineffective' is seen as an inevitable consequence.

However, the results for the recipients of a 'failing'/'ineffective' label can be catastrophic. They may simply go out of business; they may be taken under the direct control of a state education department or, as happened recently in the UK, a 'failing' school was investigated by a government appointed commission and subsequently closed. From recent UK experience, the impact of 'league tables' has been evident in:

- Political and media 'bashing' of schools and teachers.
- A test-dominated curriculum (particularly in literacy, mathematics and science) that has resulted in an over-emphasis (exclusive in some cases) on curriculum content that is to be tested or examined.
- Overt lobbying of the government by principals of non-selective schools to 'select' up to twenty per cent of their school enrollments in an attempt to improve their schools' rankings on the 'league tables'. This has resulted in a reluctance, and in some cases, direct refusals to enroll 'low-achievers'. Further, some schools have responded by concentrating their efforts on those students considered capable of improving their average examination and test scores, while giving less attention to those perceived less likely to improve.
- Parents have 'voted with their feet' by choosing to enroll their children in schools on the basis of 'league table' rankings. In some cases, this has meant changing their former residential locations to those in closer proximity to the chosen schools.

In any event, an inevitable result of comparisons among schools, whether by publication of crude 'league tables' as in the UK, France and in several Australian states, or more sophisticated 'value-added' ones like those published in the US State of Tennessee, is that there are 'winners' and 'losers'. Once the losers are deemed to be 'failing' or 'ineffective' it is difficult to find ways of helping them in prevailing political, economic and

social atmospheres of blame, recrimination and retribution. Moreover, such atmospheres are not conducive to the implementation of within-school improvement strategies.

For schools in the UK, the annual 'league table' rankings of schools on the basis of examination results have attracted such criticism that even the government which introduced them has conceded that they can be misleading (DfEE, 1995). Nonetheless, these 'league tables' continue to be published with wide political, community and media support. Whereas there is some consensus that appropriately contextualized, or 'value-added' comparisons are desirable, they are rare, and there are considerable practical difficulties in implementing them. Moreover, all rankings are fallible. In fact, several studies have now shown that there are serious and *inherent* limitations to the usefulness of such indicators for providing reliable judgments about educational institutions. The reasons for these limitations are:

- Given what is known about *differential school effectiveness*, it is not possible to provide simple summaries that capture all of the important features of schools.
- By the time information from a particular school has been analyzed, it refers to a 'cohort' of students who entered that school several years previously so that its usefulness for *future* students and the making of judgments about *school effectiveness* may well be dubious. Moreover, where information is analyzed on a yearly basis, it is necessary to make adjustments for prior contributing factors that extend over two or more years in time. In fact, it is increasingly recognized that schools, or teachers within those schools, should not be judged by a single 'cohort' of students, but rather on their performance over time. As noted by Goldstein (1997a), this makes the historical nature of school effectiveness judgments an acute problem.

Above all, even when suitable adjustments for students' intake characteristics and prior achievement have been taken into account, the resulting *value-added* estimates have too much *uncertainty* attached to them to provide reliable rankings. This point, illustrated graphically in Figure 4, is vital and one that is all too-frequently ignored by advocates of published 'league tables'.

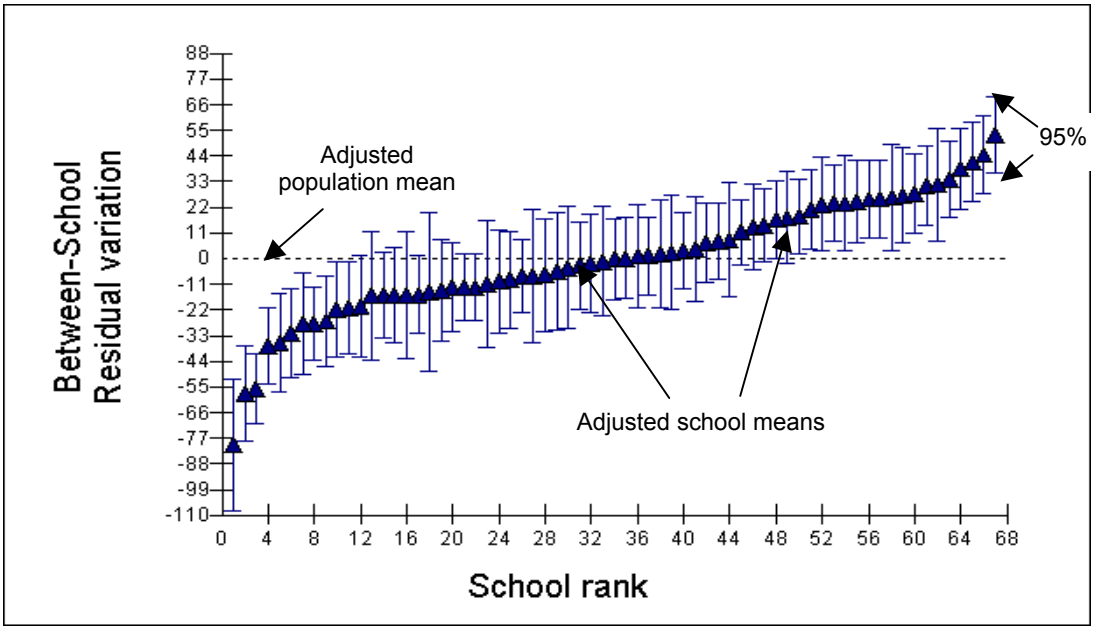

Figure 4 Typical pattern for ranked residuals[22] of schools' average test scores, within 95% 'uncertainty' intervals (Uis)[23]

KEY ELEMENTS OF EDUCATION MANAGEMENT INFORMATION SYSTEMS (EMIS)[24]

Up-to-date and accurate statistical information on the constituent elements of education systems is necessary to support educational reform efforts and to guide decision making at all levels. In turn, timely and accurate school census data are the basic building blocks for high-quality school, district, state and national information systems. Strengthening these education management information systems (EMIS) is a high priority for many countries worldwide.

Recent advances in computer-based information and communication technology is providing schools and school systems with the means of greater access to timely, relevant and detailed PI information on many functions and processes. More complex information can now be collected, analyzed and used at the school, district and system levels. Indeed, those working in and responsible for the management and administration of schools need to become informed users of EMIS systems so that the maximum benefit is gained from them.

Management information can be conceived in different ways, but its core purpose is well illustrated by Davies and Ellison (1990):[25]

> If managerial functions are to be carried out both efficiently and effectively, then it is a prerequisite that high quality information is available to inform decision-making at the various managerial levels...An information system...should be reliable and provide: the right information, to the right people, at the right time, in the right way, to achieve clear objectives.

The information systems developed are generally conceived as rational, and with the ability to supply a range of information to different parts of an organization efficiently for subsequent effective use. The emphasis is on the construction of appropriate information from the data collected and subsequently analyzed.

Davies and Ellison describe a framework that links the information systems to the decision making requirements of an organization:

- **Data-bank** - classifies and stores information (e.g., student records)
- **Predictive Information** – provides for some prediction and inference (e.g., student enrollment trends)
- **Decision Making** – based on agreed values, an information system provides the basis for recommendations to be made to management for approval or veto
- **Decision Taking** – combined information/decision-making system that utilizes available information to make, take and communicate decisions.

The decision-taking level requires the greatest level of preparation and consideration with respect to the algorithms that are used in making decisions. Laudon and Laudon (1996)[26] describe six major types of information systems that incorporate data collection with various levels of analysis and decision making:

1. **Transaction Processing Systems** (TPS) - Computerized systems that perform and record the daily routine transactions necessary to conduct the business; they serve the operational level of the organization.
2. **Office Automation Systems** (OAS) - Computer systems, such as word processing, electronic mail systems, and scheduling systems, that are designed to increase the productivity of data workers in the office.

3. **Knowledge Worker Systems** (KWS) - information systems that aid knowledge workers in the creation and integration of new knowledge in the organization.
4. **Management Information Systems** (MIS) - information systems at the management level of an organization that serve the functions of planning, controlling, and decision making by providing routine summary reports.
5. **Decision Support Systems** (DSS) - information systems at the management level of an organization that combine data and sophisticated analytical models to support semi-structured and unstructured decision making.
6. **Executive Support Systems** (ESS) - information systems at the strategic level of an organization designed to address unstructured decision making through advanced graphics and communication.

There are increasing levels of complexity in EMIS – from the level of data collection and simple reporting (TPS) to the sophistication of systems that provide a range of synthesized information suitable for a range of interpretations and needs (ESS). Interestingly, at the DSS and ESS levels the emphasis is not on the provision of information that will provide definitive answers as in the decision taking level of Davies and Ellison (this is more attuned to the MIS level), but to the provision of information that can be used in a variety of ways to address many routine and non-routine problems. At the DSS and ESS levels there is a departure from the rational assumptions of the previous levels. Decision-making is seen as involving both rational and non-rational processes, aided by the skill and experience of those interpreting the information.

It is important to note that management information systems (MISs) are not simply part of organizational decision-making processes (as outlined above), but have a complexity of both functions and purposes including:

- Housekeeping - those basic elements, such as payroll and fiscal accounting, that lead to increased efficiency of the day-to-day operations of the organization;
- Problem identification - systems designed to guide performance;
- Management support tool - including housekeeping, budgetary and fiscal control, and/or planning;
- Political and public relations - these may be peripheral to the core function of the organization and can create what appears to be irrelevant demands on organizational members;
- Provision of different types of information to different organizational levels; and
- Multiple purposes and use of information.

Example 1: Ohio's Education Management Information System

The Ohio Education Management Information System (EMIS) was established in 1989 by a Senate Bill. The Bill required the Ohio Department of Education (ODE) to develop a system that increased the amount of information available to state-level policy makers and the public – essentially for *accountability* purposes. The initial legislation reflected the following key purposes for the EMIS:

- Obtain uniform data for various input and output measures (e.g., number of teachers, students' test scores);
- compare schools and districts across the state;
- tie operating costs to output measures for efficiency ratings;
- provide numerical data by school, grade level, and subject area to be used in the identification of excellent and deficient schools and districts; and

- establish a flexible database for answering complex questions regarding schools and districts.

Within the context of a major emphasis on ***accountability standards***, the ODE has developed and continued to improve a comprehensive computer-based system of information about students, staff and finances that was previously collected on paper forms. A more recent account of this EMIS (June, 1998) entitled: *Improving Ohio's Education Management System (EMIS)* is available at: http://www.loeo.state.oh.us/pdf/emisfinal.pdf. At the outset of this document it is claimed that the EMIS is "…the most important source of information for evaluating public elementary and secondary schools." This information (about school and district performance, as well as fiscal accountability) is published – for public consumption – in the form of a "report card".

In brief, the Ohio EMIS requires the collection of 202 data elements (see Table 1) on students, staff and finances: 94 financial elements are required by state law; 98 elements related to students and staff are required by state and/or federal law; and 10 elements are needed for verification of other data elements, or for required calculations.

The system is managed via the *Ohio Education Computer Network* (OECN) and its 24 self-governing data acquisition sites that act as 'collection points' for schools and school districts submitting their EMIS data to the ODE – as illustrated in Figure 5.

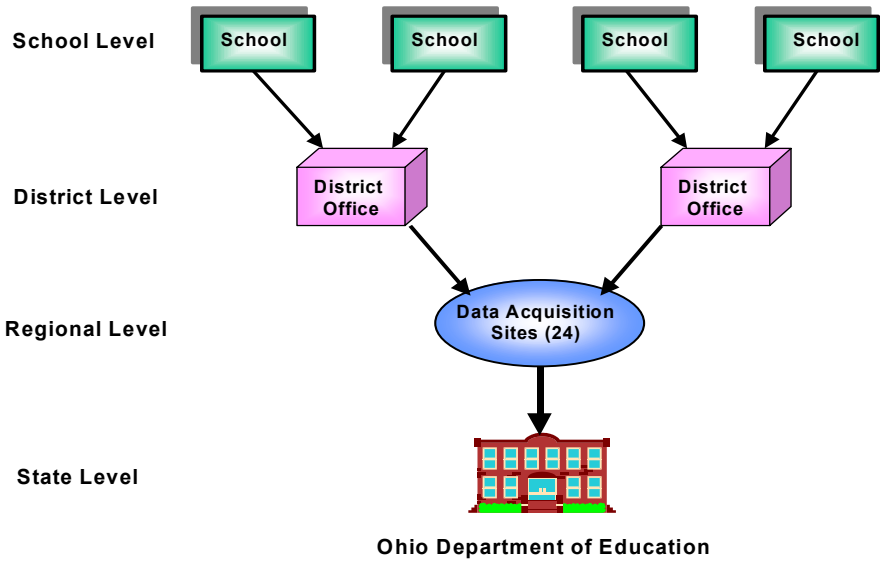

Figure 5. Ohio Education Computer Network (OECN)

Whereas the Ohio EMIS is viewed by other states and researchers throughout the US as a forerunner in collecting elementary and secondary education data, the increasing reliance on the EMIS to inform accountability standards places a premium on the provision of a quality system that is characterized by *timeliness*, *reliability* and *validity*. Such quality is heavily dependent on ensuring adequate processes at the school and district levels for collecting and entering data into the system, not the least of which involves the submission of accurate EMIS data on time. Of necessity, these processes and subsequent analyses of the data (and their presentation) require the expertise of trained personnel at the school, district and state levels.

Table 1. Illustrative Data Elements of Ohio's EMIS*

Student Data	Staff Data
Demographics Student ID code Student name Date of birth Gender District of residence Racial/ethnic category Student status Grade level (this year) Grade level (next year) Disadvantagement Limited English proficiency Disability condition	**Demographics** Staff ID code Employee name Date of birth Gender Racial/ethnic category Degree type Authorized experience (years) Total experience (years) Semester hours Attendance (days) Absence (days) Absence/long-term illness
Attendance Admission date Attendance (days) Authorized absence (days) Unauthorized absence (days) Non-attending reason Suspension Re-entry Expulsion Corporal punishment Graduation credit units Award of Merit Diploma date Diploma type	**Employment** Assignment area Certificate application Contracted pay amount/rate Local contract code Position code & start date Position status & type Position fund source Separation date & reason Course code, level & type Teacher's social security number
Programs Academic extra-curricular and intra-curricular Athletic extra-curricular programs Child day-care Drivers education Educational options Enterprise options Gifted educational programs Immigrant education program Open-enrollment Post-secondary enrollment options School-related service programs Special education placement option Special education related services Special programs Vocational programs	**District-wide Data** Building location code Building area dimensions (square feet) District code Instructional plan Number of days in session Transportation percentage Vocational education correlated classes
Academic trak and achievement Course code (for each course taken) Course grade Course status Local classroom code Competency-based education results Proficiency testing results Student summer school results	**Financial Data** Account description Amount Amount outstanding at end of period Capacity of Board Current fund balance Current payables Depository name Expense for proceeding for current fiscal year Federal contribution received Federal expenditure during current fiscal year Fiscal year Fiscal year actual expenditure Fiscal year actual receipts Fund Grant title Tax receipts: personal-general Tax receipts personal-public utilities

* The Ohio EMIS data *Definitions, Procedures and Guidelines* are available at the following Web site: http://www.oecn.k12.oh.us/www/emis/1998/emis_guide.html

However, regardless of the quality of any EMIS, unless it has the potential to better inform educational governance and practices that generate **improvements in teaching and learning – especially in students' learning outcomes**, it becomes little more than an expensive data-gathering exercise that is difficult to sustain and justify. That is, excessive emphasis on ***accountability*** *per se* leads to negative wash-back effects. For example, in the case of the Ohio EMIS, "…many school districts view the EMIS as a low priority and a burdensome state mandate" (p. ii). Moreover,

> "…very few district administrators see the value of the EMIS to better inform educational practice. In fact, most district administrators do not use the system for local purposes at all" (ibid.)

The Ohio *Legislative Office of Education Oversight* (LOEO) is well aware of these issues and has taken proactive steps towards addressing them.

Example 2: The ED*ASSIST approach to EMIS

A recent approach to the development of a generic, micro computer-based EMIS, is the *Education Automated StatiStical Information Toolkit* (ED*ASSIST), developed by the Systems Services Department of the Academy for Education Development (AED), Washington, USA. ED*ASSIST has been designed to facilitate the assessment of country-specific EMIS needs and priorities, leading to action plans for change. Specific details about ED*ASSIST, including downloadable demonstration software, are available at the AED web site: http://www.aed.org/edassist/

ED*ASSIST was developed in response to the need for more complete, accurate, and timely statistics about schools, which are understandable by a wide variety of users. Its key features include:

- An integrated set of tools (software, illustrative examples and technologies) for data gathering, processing, presenting (both tabular and graphical) and utilization;
- Built-in enforcement of data quality standards and practices to produce timely, efficient and reliable data;
- Supports demand-driven EMIS, including Fundamental Quality Level indicators, as illustrated in Figure 5;
- Provides for many educational data inputs required by most education systems for analysis, policy dialogue, projections and simulations.

The underlying rationale for the Toolkit is that in any given country, different information is required for different education activities. For example, state and national education planners need information structured differently than regional officers or communities. To this end, ED*ASSIST has been designed to organize and support information for three types of users:

- Educational policy planners and strategists;
- Management personnel; and
- Operational personnel.

In addition to its core set of outputs and indicators, ED*ASSIST supports a demand-driven information flow by facilitating the creation of additional outputs and indicators as required by the range of users and stakeholders seeking them (see Figure 5). Current information on students, teachers, schools, infrastructure, and finances can then inform decisions about day-to-day operations, management, and long-term strategy.

At this stage, Ministries of Education in at least six countries are using all or parts of the ED*ASSIST approach, including: Benin, Kenya, Lesotho, Nicaragua and Uganda.

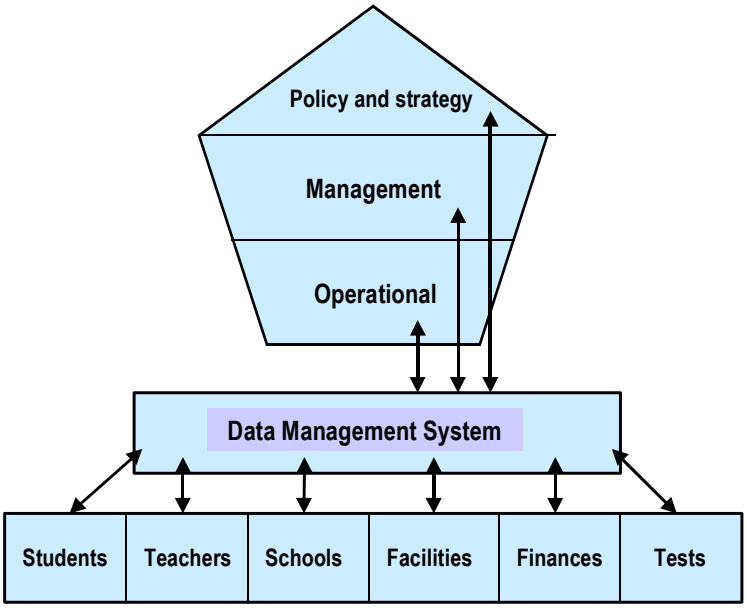

Figure 6. ED*ASSIST's demand-driven information flow

Summary

In brief, *useful* educational performance indicators (PIs) are those that are *relevant*, *cost-effective*, *timely*, *reliable* and *valid* – in terms of their capacity to inform the processes of strategic decision-making and decision-taking – resulting in measurable improvements to desired outcomes – especially in student achievement. Such is also the case for both rudimentary and sophisticated Education Management Information Systems (EMIS). An important guiding principle is that PI data **should not** be collected for their own sake, but rather, because they are needed for specific policy purposes.

It is also important to stress that **feedback** to schools and districts of responsibly analyzed and presented PI data is a **necessary** basis for identifying the 'strengths' and 'weaknesses' of resource provision, in teaching and learning programs, as well as in students' achievements. Whereas feedback of such PI data is a *necessary* condition for subsequent improvement, access by schools and/or districts to the data *per se* do not constitute a *sufficient* condition to engender reform and change for the better. Rather, such change is crucially dependent on careful and responsible management of performance information by school administrators and leadership teams within the context of a shared commitment to strategic, continual improvement at all levels of an education system and among all its stakeholders.

==========

Notes

[1] Lievesley, D. (2001). Making a difference: A role for the responsible international statistician? (Presidential address). *Journal of the Royal Statistical Society* (Series D), *50* (4), 367-406.

[2] Scheerens, J. (2001a). School effectiveness in developing countries. *School Effectiveness and School Improvement, 12* (4), 353-358.
Scheerens, J. (2001b). Monitoring school effectiveness in developing countries. *School Effectiveness and School Improvement, 12* (4), 359-384.

[3] See especially:

Dorn, S. (1998). The political legacy of school accountability systems. *Education Policy Analysis Archives*, 6 (1), 1-32.

Hill, P.W., & Crévola, C.A. (1999). The role of standards in educational reform for the 21st century. In D.D. Marsh (Ed.), *ASCD Year Book 1999: Preparing our schools for the 21st century* (pp. 117-142). Alexandria, VA: Association for Supervision and Curriculum Development.

Tucker, M.S., & Codding, J.B. (1998). *Standards for our schools: How to set them, measure them and reach them*. San Francisco, CA: Jossey-Bass.

Visscher, A., Karsten, S., de Jong, T., & Bosker, R. (2000). Evidence on the intended and unintended effects of publishing school performance indicators. *Evaluation and Research in Education, 14*, 254-267.

[4] *Summit of the America's – Line 2*: *Educational Assessment, Brasilia*, March 12-14, 2002. INEP.

[5] National Centre on Education and the Economy. (1997). *New standards for performance*. Washington, DC:Author.

Mortimore, P. (2001). Globalisation, effectiveness and improvement. *School Effectiveness and Improvement, 12* (2), 229-249.

[6] Mortimore, P. (1991). School effectiveness research: Which way at the crossroads? *School Effectiveness and School Improvement, 2* (3), 213-229; p. 214.

[7] OECD (1989). *Schools and quality: An international report*. Paris: Organization for Economic Cooperation and Development.

OECD, (1995). *Indicators of education systems: Measuring the quality of schools*. Paris: Organization for Economic Cooperation and Development.

[8] Source: http://www2.unesco.org/efa/wef/countryreports/nepal/

[9] Manno, V.B. (1994). *Outcomes-based education: Miracle, cure or plague?* Hudson Institute Briefing Paper No. 165, June 1994.

[10] *Millennium Declaration Goals*, The United Nations Millennium Summit, New York, September 2000.

[11] See: Beaton, A.E, Mullis, I.V.S., Martin, M.O., Beaton, A.E., Gonzalez, E.J., Kelly, D.L., & Smith, T.A. (1996). *Mathematics achievement in the middle school years: IEA's Third International Mathematics and Science Study*. Chestnut Hill, MA: Boston College.

Mullis, I.V.S., Martin, M.O., Beaton, A.E., Gonzalez, E.J., Kelly, D.L., & Smith, T.A. (1997). *Mathematics achievement in the primary school years: IEA's Third International Mathematics and Science Study*. Chestnut Hill, MA: Boston College.

[12] ***Content validity*** is established through a rational analysis of the content of an indicator or set of indicators – based on individual, subjective judgement. There are two major types of content validity: face validity and logical validity. ***Face validity*** is established when it is agreed (by consensus) that that an indicator (e.g., a math test score) is a valid measure of a relevant trait (i.e., math achievement). ***Logical*** or ***sampling validity*** involves a careful definition of the **domain** of elements to be measured and the logical design of indicators to cover all the relevant areas of this domain.

Criterion-related validity is established when indicator measures can be related to a predicted criterion. For example, in order to have criterion-related validity, measures of 'inputs' (e.g., per capita cost of education), must be related to (or positively correlated with) a relevant 'output' criterion (student achievement). Useful definitions and discussions of these constructs are given by Allen, M.J. and Yen, W.M. (1979). *Introduction to Measurement Theory*. Monterey, CA: Brooks/Cole Publishing Co. (Chap. 5, pp. 95-117).

[13] A well-presented outline of these services is available in:

Tymms, P. (1999). *Baseline assessment and monitoring in Primary Schools: Achievements, attitudes and value-added indicators*. London: David Fulton Publishers.

14 Masters, G.N. *Educational Measurement: Assessment Resource Kit*. Camberwell, Vic: Australian Council for Educational Research.

15 Meiers, M. (1999, September). *A national longitudinal literacy and numeracy study.* Paper presented at the 24th Annual Congress of the Applied Linguistics Association of Australia, Perth, September, 26-29.

Meiers, M., & Forster, M. (1999, October). *The Longitudinal Literacy and Numeracy Study* (LLANS). Paper presented at the ACER Research Conference 1999, Improving Literacy Learning, Adelaide, October 18-19.

Rowe, K. J. (2001). *Progress report for LLANS Literacy: Surveys 1-4*, August 2001. Camberwell, VIC: Australian Council for Educational Research.

16 Ladd, H.F., & Walsh, R.P. (2001). Implementing value-added measures of school effectiveness: Getting the incentives right. *European Educational Research Journal, 43* (1), 1-23.

17 See, for example:

Bosker, R.J., & Witziers, B. (1995). *School effects: Problems, solutions and a meta-analysis*. Paper presented at the eighth annual International Congress for School Effectiveness and Improvement, CHN, Leeuwarden, The Netherlands, January 3-6, 1995.

Mortimore, P., Sammons, P., & Thomas, S. (1994). School effectiveness and 'value-added' measures. *Assessment in Education*, 1, 315-332.

Saunders, L. (1999). A brief history of educational 'value added': How did we get to where we are? *School Effectiveness and School Improvement, 10* (2), 233-256.

Thomas, S., Sammons, P., Mortimore, P., & Smees, R. (1997). Stability and consistency in secondary schools'effects on students'GCSE outcomes over three years (Leading article). *School Effectiveness and School Improvement, 8*, 169-197.

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

Yang, M., Goldstein, H., Rath, T., & Hill, N. (1999). The use of assessment data for school improvement purposes. *Oxford Review of Education, 25* (4), 469-483.

19 See Visscher et al. (2002) reference cited in Note 18.

20 Relevant literature references to all that follows in this section are provided in:

Rowe, K.J., Turner, R., & Lane, K. (2002). Performance feedback to schools of students' Year 12 assessments: The *VCE Data Project*. In A.J. Visscher and R. Coe (Eds.), *School improvement through performance feedback* (pp. 163-190). Lisse, The Netherlands: Swetz & Zeitlinger.. A complete version of this chapter is available on ACER's web site, at: http://www.acer.edu.au

21 See in particular:

Goldstein, H. (1997a, March). Value added tables: The less-than-holy-grail. *Managing Schools Today*, pp.18-19.

Goldstein, H. (1997b, July 18). From raw to half-baked. *Times Educational Supplement*, p. 15.

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

Goldstein, H. (2001). Using pupil performance data for judging schools and teachers: Scope and limitations. *British Educational Research Journal, 27*, 433-442.

Goldstein, H. & Myers, K. (1996). Freedom of information: towards a code of ethics for performance indicators. *Research Intelligence*, 57, 12-16.

Goldstein, H., & Spiegelhalter, D. (1996). League tables and their limitations: Statistical issues in comparisons of institutional performance. With discussion. *Journal of the Royal Statistical Society, A, 159*, 385-443.

Goldstein, H., & Thomas, S. (1996). Using examination results as indicators of school and college performance. *Journal of the Royal Statistical Society* (Series A), *159*, 149-163.

Myers, K., & Goldstein, H. (1996). Failing schools in a failing system. *Association for Supervision and Curriculum Development (ASCD) Year Book*.

Rowe, K. J. (2000). Assessment, league tables and school effectiveness: Consider the issues and let's get real! *Journal of Educational Enquiry, 1* (1), 72-97.

[22] A useful way of interpreting a **residual** in this context is that it represents what *remains* after adjusting for factors that are known to *influence* or *explain* variation in students' achievement outcomes, including measures of 'ability' and prior achievement.

[23] Rather than referring to these intervals as *uncertainty intervals* (UIs), it is more common to refer to them as *confidence intervals* (CIs). Note that 95% *confidence intervals* for a statistic (a mean point-estimate for each school in this case – $\bar{x}_s$) are calculated from: $\bar{x}_s \pm 1.96 \times$ the school's standard error (i.e., the school's standard deviation divided by the square root of the school cohort size – $\sigma_s / \sqrt{n_s}$). These intervals imply that we can be 95% 'confident' that the estimate of a school's mean lies between these upper and lower limits. However, in the present context of making comparative judgements about the relative performance of schools, these limits are more properly referred to as *uncertainly intervals*. That is, when the intervals for two or more schools overlap, there is *no certainty* that their relative performance differs significantly. For a presentation and discussion of the relevant conceptual and technical issues, see: Goldstein, H., & Healy, M.J.R. (1995). The graphical presentation of a collection of means. *Journal of the Royal Statistical Society, A, 158*, 175-177.

[24] The relevant references cited in this section derive from two sources related to EMIS, namely:

Gurr, D. (1997). *The development of Management Information Systems in education*: Web site: http://staff.edfac.unimelb.edu.au/david_gurr/papers/

LOEO (1998). *Improving Ohio's Education Management Information System* (EMIS). Columbus, Ohio: Legislative Office of Oversight.

[25] Davies, B., & Ellison, L. (1990). Management information. In B. Davies, L. Ellison, A. Osborne, and J. West-Burnham (Eds.), *Education management for the 1990's*. London: Longman.

[26] Laudon, K.C., & Laudon, J.P. (1996). *Management information systems: Organization and technology*. NJ: Prentice-Hall.