Theme
QUALITY ASSURANCE FOR ENHANCEMENT

An Exploration of Key Performance Indicators for Academic Quality

Dr Jan Cameron, Assistant Vice-Chancellor (Academic)
University of Canterbury
P.Bag 4800, Christchurch, New Zealand.

Abstract

Over the last five years various attempts have been made in New Zealand to develop indicators of academic quality which derive from aggregate institutional data. However the indicators which have been considered have often been criticised for a variety of reasons, including a weakness in providing guidance for enhancement. Against this, many of the real attributes of academic quality of interest to university staff are not amenable to numerical representation. This paper suggests alternative measures which derive more directly from key academic quality assurance activities and which might also be valuable for benchmarking and benchlearning. In particular, the paper argues that performance indicators for academic quality are most likely to be meaningful for assessing and influencing quality if they are derived from the quality assurance processes themselves rather than from best available aggregate data.

Introduction

This paper explores the relationship between key performance indicators and academic quality in universities. In particular, it seeks to find ways in which the development and use of performance indicators can also fit the academic quality enhancement agenda.

The first serious proposal for institution-wide academic quality measures to be introduced to New Zealand universities emerged in 2003. The intention of Government, via the newly-established Tertiary Education Commission (TEC), was to include a performance element related to teaching in funding calculations for all tertiary education institutions[1]. This was intended to parallel the recently-introduced Performance-Based Research funding component (PBRF) whereby approximately 17% of government funding to the institution is determined on the basis of a collation of research measures.[2] The Executive Summary of the 2003 Technical Working Group's Report (Ministry of Education 2004) stated the intention: "The report recommends a package of performance indicators and measures, which are intended to be used to allocate a portion of funding for tuition according to provider performance and with

[1] Approx 230 organisations, of which 8 were universities. New Zealand universities are autonomous but they gain close to half their funding from government sources. They must therefore be accountable to government through its crown agencies, in particular the Tertiary Education Commission (TEC) with respect to quality, relevance and funding and the New Zealand Qualifications Authority (NZQA) with respect to academic programmes.

[2] The teaching version of the PBRF was colloquially referred to as the Performance-Based Teaching Fund.

a view to *enhancing the provision of quality in terms of educational gains by learners*" (p8, emph. added). To qualify for performance funding an institution would have to "report satisfactorily against all indicators" (p9). The indicators being proposed were benchmark measures for retention and course completion "averaged over all courses offered by the [institution]", plus measures derived from a Graduate Experience Survey, and applied both in the aggregate and also to target equity groups.

Most universities agreed that a student opinion survey in itself was not objectionable but noted they already delivered "in-house" surveys which were fit for purpose, ie, for their own purpose. Any common New Zealand-wide survey would have to be relevant to all tertiary education institutions (not just universities), and would need to be administered nationally if it was to gain acceptance. The proposal to include retention and completion rates was met with a range of objections, most of which are now well-rehearsed. In her university's submission the author of the current paper stated categorically, "The proposed measures of retention and completion bear no relation to the evaluation criteria of teaching performance".

Notwithstanding these arguments, a search for "quality indicators" has persisted. A joint working group between TEC, the NZ Vice-Chancellors' Committee (NZVCC) and the Universities' Academic Audit Unit in 2007-2008 failed to develop an agreed set of indicators for the universities.[3] In June 2008 the New Zealand Auditor General promulgated his "observation on the quality of performance reporting" of non-financial performance of public entities. The objective is transparency of operations and accountability to "Parliament and the public", and good management practice. While his observations "did not include performance reporting by tertiary education institutions" (para 1.17), the fact that the Auditor General has shown interest in such matters should alert universities that a search for performance indicators related to academic quality has not disappeared.

In part the difficulty in moving forward on the development of quality indicators derives from a perceived lack of compatibility between indicators used to satisfy accountability objectives and indicators which might be useful for guiding enhancement objectives. For the institutions, the prime focus is on understanding their own processes towards enhancement. For sector agencies the prime focus is on ensuring value in the use of public funds.

In practical terms, the concept of quality has been collapsed into the concept of performance indicators – quality is one of the factors on which "performance" is accountable. At the same time, though, the validity of some performance measures as indicators of academic quality, let alone sign posts for ways of improving quality, must be questioned. In this presentation it is argued that if indicators are to be useful for enhancing quality it is preferable to start from the quality assurance process itself and build indicators from it, rather than attempt to interrogate indicators calculated at the aggregate level in order to deduce and influence academic contribution effects. The paper seeks to assess usefulness, reliability and stability, and comparability of a selection of indicators in order to evaluate which might be more or less

[3] NZQA's most recent work for the remainder of the sector focuses on "evaluation questions" and "evidence-based self-assessment" rather than numerical indicators *per se* (NZQA 2008). This focus thus resembles the universities' academic audit approach more closely than the performance indicator approach used in earlier proposals and in university statements of service performance.

valuable in meeting an objective of being both useful indicators of academic quality and useful indicators of directions for quality enhancement. Some alternative ways of using available data are explored.

The paper is exploratory in that it traverses a range of existing measures as well as a variety of academic quality assurance processes. The paper invites input from participants particularly with respect to the validity of the indicators discussed and their benchmarking potential.

The New Zealand Context

The institutional quality framework for New Zealand universities derives from the responsibilities encumbent on the New Zealand Vice Chancellors' Committee (NZVCC) which is comprised of, and oversees, the eight universities. In particular, the NZVCC's Committee on University Academic Programmes (CUAP) has authority to approve academic programmes and accredit universities to offer academic programmes. In addition, the NZVCC has established an Academic Audit Unit which maintains its quasi-independence by way of an independent Board. The Audit Unit conducts regular, five-yearly academic audits of all universities. Its panels are appointed independently and include international auditors; its reports are public documents. Universities are thereby publicly accountable. In addition, universities report to the Crown on performance via Statements of Service Performance and Annual Reports.

During the various discussions outlined in the Introduction, the universities collectively have resisted incursions into what they see as their institutional autonomy. In addition to the rigorous process of programme approval, accreditation and audit, the universities repeatedly rehearse the catalogue of quality assurance activities which are routine business for them: peer review of new and existing programmes; moderation and accreditation by professional organisations; external examination of theses; quality audit and so on (see Scott and Cameron 2008). Such activities do not easily lend themselves to development of meaningful numerical indicators, especially those which might be used for assessing institutional quality or success, or for comparing sets of institutions.

Within our own university some resistance to formalised academic quality assurance persists. This is not unusual. As noted by Chalmers, the concerns of academic staff are well founded, "as the information gathered for institutional and national compliance and quality purposes is largely seen as unrelated to, and removed from, what is important to teachers and their students – engaging together in the process of learning and teaching" (2007, p12). This "disenchantment' is important, since quality enhancement has to be an academic enterprise. Institutional plans and compliance reports discharge their accountability responsibilities by reference to performance indicators. At the same time, effectiveness of these institutional plans and strategic documents depends, at least in part, on academic staff. It is thus essential that performance measures used within the institution are understandable, meaningful and useful for academic staff if they are to see these as anything other than a bureaucratic imposition. However the summary measure is calculated, it must be underpinned by activities

and processes which reflect the actual academic enterprise, and which can be influenced directly by academic action.

Critique of Conventional Performance Measures

Several writers (eg Coates 2007) observe that one of the problems with many existing performance indicators is that they have been developed from the data which are available, rather than the data which might be most reliable or valid. The tension is between "counting what is measured or measuring what counts" (HEFCE 2008). Chalmers claimed :

> "Much of the literature is critical of many of the current indicators in use, particularly input and output indicators. *It is likely that these indicators came into common use as a result of availability, rather than through any analysis of their appropriateness.*" (2007, p79, emph. added.)

The retention (or attrition) and completion rates proposed by the 2003 Working Party in New Zealand, and referred to in other jurisdictions, are seemingly attractive as performance indicators precisely because they are derived from data which are already reported to Crown agencies, albeit for other purposes. Problems with these particular measures which have been identified include

- lack of any clear theoretical link between attrition/retention and academic quality;
- potential perverse relationships in that high completion rates might reflect low standards;
- variation among institutions in what non-completion actually means, and how it is measured;
- failure to address the complexity of the multiple factors which affect attrition;
- lack of differentiation between attrition due to institutional factors and attrition due to personal factors;
- lack of consistency in defining the unit(s) of analysis;
- changes in learning patterns, including dipping in and out of education (course sampling and life-long learning);
- variation in early departure patterns by different sub-groups, within and across institutions;
- the impact of user-pays on ability to stay in study;
- impact of articulation agreements on encouragement of transfer between institutions. (Coates p90; Chalmers p77-9; Cooper 2002; Yorke and Longden 2004)

More specifically, in the New Zealand context of open entry to university for any citizen over the age of 20, the objective of high retention and course completion sits in tension with a simultaneous objective of improved access, especially by disadvantaged groups and adult learners – groups which are known to have higher attrition rates. Crude measurements of qualification completion also sit in tension with a stated objective to increase opportunities for stair-casing, and portability of achieved units of study across the sector. A benefit of New Zealand universities' open entry to adult students is precisely because they can opt in and opt

out as lifestyle and other commitments permit. “Retention” for such students might well be a meaningless concept.

Retention and completion rates thus have the attraction of being [apparently] easily measurable and quantifiable. Their disadvantage is that they bear little necessary relationship to the quality of education being provided – and they risk prompting serious perverse effects which would undermine any notion of quality. They might have proved useful to facilitate funding decisions but even then the irony would be to under-fund where potentially the need for improvement, and resource injection, was greatest. At a practical level such measures might indicate apparent efficiency of the sector but they have no obvious enhancement consequence. New Zealand commentary reflected Chalmers’ observation that “there are growing concerns about the appropriateness of using retention and attrition measures to make conclusions about the educational quality of an institution” (p 76).

Unease within the university sector also derives from a lack of clear differentiation between measures or indicators of quality outcome, quality output and quality input. Retention and completion measures, for instance, might be seen as indicators of a quality output which bear only limited relationship to quality outcome, ie the quality of an education which students or graduates receive. For example, students who do not complete a programme of study might, despite this, have gained considerably in knowledge, understanding and skill (ie, education) from the experience they had at university even if their time there was truncated. This is a quality outcome *for those students.*

Academic processes, on the other hand, need to be evaluated both for their own intrinsic quality (quality of the process), and for the likelihood of producing a quality outcome or output (the quality effect). Measures of quality effect are most useful for quality enhancement. Extending this argument, the linkage between processes and outcomes or outputs needs to be unpacked: universities might demonstrate quality outcomes for graduates because of good processes, or despite poor processes.

The interactions outlined here, between processes, outcomes and outputs, have implications for how indicators are constructed and used. Output indicators might be useful to those focussing on accountability but processes, and indicators which reflect the interaction between processes and outcomes (quality of the effect), are likely to be more useful for those interested in quality enhancement. There is therefore a need to relate indicators directly to the *intended outcome* or *intended effect* for the organisations themselves. In the ongoing discussion about evaluation indicators the key questions around who the users are, and for what purpose they need these indicators, are often either assumed or unanswered. Yet these fundamental questions should be directing development of evaluation indicators in ways which are transparent, useful and understandable. Indicators, like the quality processes themselves, must be fit for purpose.

One further critical issue which must be managed is data quality. As already noted much of the data used to develop key performance indicators (KPIs) have been collected for other purposes. Their robustness and relevance as academic quality measures are therefore questionable. For example, within a New Zealand university the term “part-time” has at least two formal definitions and probably several informal descriptions. For student loan purposes, “full-time” is

defined, nationally, as a minimum of 0.8 efts[4]: a part-time student would therefore be someone not studying full-time, that is studying less than 0.8efts. Yet for some students who are studying less than 0.8efts, this is their sole occupation, ie their university study is their "full-time" activity, while other students who *define themselves* as "part-time" are studying while engaged in other commitments and might take only one or two courses at a time, say 0.3efts per year. At the same time, a part-time thesis student is defined for fees purposes as 0.65efts. Yet some postgraduate degree regulations allow up to four years for completion of a 2efts degree, implying that for completion purposes part-time might equal half-time. One of the variables in TEC assessments of retention and completion is a binary variable, full-time/part-time status (Gray, p19). Study status might be expected to have meaningful impacts on student engagement and, consequently, on student achievement and perhaps on retention. But it is clear from the above examples that analysis of data by full-time/part-time status is highly problematic. This example is a salient reminder of Coates' directive, that "excellent measures [should] precede measures of excellence" (2007).

The challenge now being considered is how to develop a new set of performance indicators which:

- are relevant, valid and useful, to both the institution and to staff;
- can be influenced directly;
- can be used for benchmarking within the institution, both through time and across academic units, as well as across institutions;
- will spur enhancement initiatives;
- will reflect achievements resulting from these; and
- apply across the institution such that at the next level these might inform the university's Statement of Service Performance and the Annual Report as meaningful indicators to Crown agencies (and other stakeholders) of the institution's academic quality.

Development of Performance Indicators Which are Useful for Academic Quality: Defining and Measuring Learning and Teaching Excellence

For the purposes of this paper four questions are asked which relate to a specific strategic planning target related to teaching and learning excellence: "To value, encourage, reward and celebrate excellence in teaching and learning". In particular the following questions related to "encouraging excellence" are pertinent to planning and reporting objectives.

1. How do we determine "excellent" teaching and learning ?
2. What are the processes in which we engage to encourage excellence in teaching ?
3. Apart from excellence in teaching, what are the processes in which we engage to encourage excellence in learning ?
4. How do we know that these work ?

[4] 0.8 of a full-time study year of 1200 hours. In New Zealand courses and programmes are defined in terms of efts (the study load expected of an equivalent full-time student). A three year degree is thus 3.0efts, or expected to take notionally 3 years (3600 hours) of study.

The first of these questions asks how we define quality, how we know it when we see it. The second and third questions ask how we encourage or facilitate improved quality. The fourth question implies that the encouragement actually works, ie that quality might be assessed. The answers to question four will presumably lie in shifts in the measures used to answer question one, ie the quality effect.

Questions one and four deliver the performance indicator, ie measures of change in teaching and learning "quality" (the outcome); questions two and three specify the processes and enable the locus of effective intervention to be identified.

There is real risk of terminology confusion in these relationships:

1. Performance indicators are numerical representations of *quality* – of what?
2. The processes used to achieve the quality represented by the indicators are frequently *quality assurance processes*. These are termed "interventions", including diagnostic as well as remedial or enhancement intervention.
3. The quality of the quality assurance processes might be evaluated via the outcome, which in practical terms is the *quality effect.*
4. The extent and direction of the quality effect is also therefore an indicator of the *quality of the processes.*

The following examples include an attempt to differentiate these different dimensions of quality.

Excellent Learning

Intuitively, the most obvious indicator of excellent learning is the academic achievement of the student – his or her grade, or GPA. This might sometimes be broken down by ethnic group or age of student or admission category, for example. Other indicators which have been used include number graduating from equity groups; number of academic prizes awarded and number of prestigious scholarships gained. Both of the latter two indicators are problematic since across the whole university they might represent very small proportions of all students and, furthermore, are constrained by numbers of scholarships or prizes available. Eligibility and availability therefore determine the outcome. Certainly they are reflections of learning excellence *for those students* but they might say little about the institution's overall standards.

Use of GPA or pass rate metrics prompts reflection on whether these measures of excellence are necessarily reflections of achievement, and of "excellent learning" – does a high grade mean, by definition, that the student has achieved excellence? Poor academic quality (lowering standards) could result in higher grades and higher GPAs. Secondly, are grades the only, or the best, indicator of what a student has achieved? Can a student who has high grades but is unemployable because his degree programme was incoherent, be said to have achieved highly?

Such apparently trivial questions prompt closer reflection on what we might be considering when we seek indicators of academic quality (in this case "excellent learning"), in particular to relate quality to purpose. In the above example, the key question might well be, "how well have students met the graduate profile of the course/programme they are studying?" ie their

achievement is related to course or programme objectives. It remains possible, for instance, for a student to achieve highly on individual course objectives but fail on programme objectives. That might be considered a "high quality" education as an *outcome* for the individual, but for the institution such a student fails to satisfy programme intentions, which would be a "low quality" *output*.

Individual student achievement, therefore, is not necessarily commensurate with what the institution deems to be achievement. When access is an institutional objective there will be some students for whom gaining entry is an achievement; for others just passing is a major achievement. At the same time, some "high achievers" might not have learned a great deal – at least not at the university – while some "low achievers" might have learned a lot. To "expose the net effect of education contexts and processes" some assessment of value added is needed (Coates p90). Methodologies for doing this, however, do not carry favourable ratings (see Chalmers p60-1). Until such time as valid, reliable and affordable value-added analyses are developed, institutions might well need to rely on such measures as "proportion of students who claim they have met the graduate attributes" (whether or not their teachers would agree) and "proportion of graduates in related employment or further study" (whether or not this was their objective in undertaking initial study). Use of these measures as performance indicators must therefore be accompanied by caveats on relevance and reliability.

Measuring achievement itself might be simplistic, and susceptible to corruption effects, but we do have evidence of institutional practices – the quality assurance processes - which facilitate good learning and, hence, should encourage high achievement. One factor which has a direct relationship with achievement is, not surprisingly, admission standard, demonstrated through different versions of school-to-university admission tests[5]. More than this, some New Zealand data show a relationship with more general admission *category*[6]. In this case the admission profile is an indicator of *likely* achievement (a performance indicator, albeit one over which the institution might have little direct control); identification of the relationship between admission category and first year achievement is a quality assurance process which enables the institution to identify at-risk students and apply appropriate interventions. For this example the schematic representation might be:

[5] Since 2005 the national school-based University Entrance qualification in New Zealand has been on the basis of the National Certificate of Educational Achievement, NCEA; prior to 2005 it was on the basis of the University Bursaries Examination. The Entrance Standard, as a school-based examination, is not set by the universities, but by the NZ Qualifications Authority.

[6] Significant admission categories for New Zealand universities are University Entrance (either via Bursaries examination pre-2005 or National Certificate of Educational Attainment 2005 onwards); Discretionary Entrance usually for high-achieving school students without UE; Cambridge International; International Baccelaureate; other overseas qualifications; university foundation certificates; adult (over 20) admission.

Input	Quality Assurance process		Outcome	Output
Students in specific admission category	Identify students at risk of poor achievement	Intervention e.g., provide learning skills support	Improvement in group achievement category over time	Mean GPA
Potential PIs	Proportion of first years in "at risk" admission categories (indicates risk of low achievement, ie not achieving excellence – but might not be in university's control) GPAs or Pass rates for "at risk" categories of students over time (indicates likely impact of identification and intervention – the quality assurance process).			

Admission qualification would appear to have a direct impact on learning outcome, since it is a representation for (indicator of) academic preparedness. Once a student has begun university study a raft of other effects might come into play. One area in which institutions are currently most interested is those activites which foster students' engagement, with their academic community and with their learning. Coates (p91), for instance, argues for robust measures of student engagement, since on the one hand we know that student learning is enhanced by active engagement and on the other hand we know of the broad range of activities which contribute to engagement (see, for example, Zepke *et al* 2005) But while these activities and experiences are known to facilitate learning, they do not of themselves represent excellence in achievement. Use of these measures might be more useful as academic quality indicators in revealing gaps or weaknesses in an institution's provision. They are blood pressure checks in that they signal risk areas and indicate where further exploration is required and interventions need development.

Excellent Teaching

The second component of question one above is "excellent teaching.

Most of our university's performance indicators related to teaching are indicators of "supporting" and "rewarding" [excellent] teaching. They include, for example, number of teaching awards; number of staff enrolled in tertiary teaching qualifications; number and value of teaching development grants. While it is important that the institution monitors these attributes, and the measures are true measures in that they measure what they say they do which is "value", "support" and "reward" teaching, they have some problems as institutional performance measures. Firstly, they relate to very small numbers of people who gain teaching development grants, win teaching awards or are enrolled in teaching qualifications. These institutional KPIs refer to less than 3.5% of academic staff. More critically, it is impossible for some of the figures themselves to increase markedly since – as with achievement of prestigious scholarships by students - achievement of grants and awards is constrained by availability, if not by eligibility. Even the number of staff enrolled in teaching qualifications is at risk of truncation effects, in that those who have already gained such qualifications will be excluded from the calculation. Just as the number of scholarships gained does not reflect the standard of learning across the whole institution, so too these teaching KPIs do not provide a robust indicator of excellent teaching across the institution.

Excellence in teaching is desirable for its own sake, but it is *necessary* in the quest for excellent learning. It is important that reliable measures of teaching quality are developed. Conventionally, teaching excellence has been measured via student surveys. At the very simplest, a performance indicator for "good teaching" might be a score in the range 0-5 on the measure "Overall this was a very good course". One institutional KPI which reflects this is the "the percentage of students surveyed who thought their course was of 'good quality'". While there might be some problems with use of such a general aggregate score as a performance indicator (see below), the more pressing weakness for quality assurance purposes is that it tells nothing about activities or groups which are amenable to intervention in order to improve the "good quality" evaluation.

One component of good teaching, and a facilitator of good learning, is assessment. This university does have a KPI related to assessment, the "number of appeals related to assessment". These numbers are low, in single figures, and do not reflect whether or not the appeal was upheld. Furthermore, as a performance indicator the "measure" collapses a measure of availability and access to an appeal process with the subject of the appeal – a low number could as easily indicate inaccessible or difficult processes, or good informal resolution processes, as indicate few complaints about assessment. The performance indicator serves neither performance measurement nor quality enhancement well. A different way of using assessment as a focus for both quality assurance, enhancement and performance indicator is to use one of the questions in the standard course survey, "feedback on assessment was good/very good". This statement addresses one of the parameters of a "good quality" course. For quality assurance purposes this score is much more useful: not only does it direct action towards improvement but it is also a score which is known to be directly related to good learning. These activities can be shown schematically thus:

Input	Quality Assurance process		Outcome	Output
Courses surveyed on assessment feedback	Mean score on assessment feedback	Intervention e.g., workshops on assessment feedback	Improvement in mean score on assessment feedback	Mean score
Potential PIs	Proportion of courses scoring lower than 4.0 on course survey assessment items Mean score on assessment items in course surveys.			

The quality of academic programmes offered is a different parameter of academic quality. Most institutions use programme reviews to assure themselves of programme quality and to benchmark against national, international and professional standards. How might this practice be translated into a performance indicator? In the New Zealand context a university might use the number of programme reviews carried out (as the process whereby quality is assured), the number of proposals approved by NZVCC, or the number of programmes gaining professional accreditation. Of these, the first implies only that the institution is concerned about maintaining quality, but the latter two do indicate programmes meeting national and professional standards. As with performance indicators such as numbers of awards won, the

problem is once again the low possible base. A more meaningful indicator of quality might be the proportion of proposals which receive approval/accreditation. But again, these measures are based on a very small proportion of the institution's total programmes.

The challenge is to find a measure which reflects the overall quality of programmes, which must relate to their objectives – proportion of degree graduates gaining employment or further study in fields related to the programme might be a valid indicator, but even this is limited in its quality enhancement potential unless disaggregated by actual graduate attributes.

This overview suggests that it might be difficult to establish meaningful, reliable and robust performance indicators for teaching quality (including programme quality) which are also useful for quality assurance purposes.

This discussion so far reveals that there are different purposes for different performance indicators and their associated processes. For some, such as admission qualification (an input), there is a direct link to an outcome which is measurable. This correlation also lends itself to more detailed analysis in order to develop institutional (or subgroup) performance measures which are robust (see below).

For other topics the performance indicators will refer to activities which reflect the processes applied to enhance the outcome. It is essential that these are not assumed to measure quality itself. Rather, they might be seen as diagnostic tools, to highlight points of potential intervention or to trigger analysis for further explanation. The second and third questions posed above are about the interventions intended to result in quality, ie the quality processes:

- what are the processes in which we engage to encourage excellence in teaching ?
- apart from excellence in teaching, what are the processes in which we engage to encourage excellence in learning ?

These indicators about intervention or encouragement are thus indicators of intent, not of achievement itself. For both quality assurance and quality enhancement the processes involved are necessary, but might not be amenable to calculation as performance indicators. Yet if strategic planning targets are to be met then the strategic plan needs to be able to link the performance indicators to the quality processes and their effects. The quality of the processes might be indicated by the outcome, ie whether the processes worked or not (the "quality effect"), but more probably will require some qualitative evaluation which is not amenable to transformation into a numerical indicator.

Use of Quality Assurance Data to Develop Institutional Performance Indicators

This paper's argument is that performance indicators will be most useful for quality enhancement if they are deduced out of the quality assurance process, not vice versa.

For example, we review the progress of students, which enables us to identify where excellence in learning is being achieved for individuals, and where students are at risk of not

achieving even passing standards. One performance indicator related to this process is "number of academic progress advice letters sent to students". Students identified as being "at risk" are advised in different ways, for instance they might have their workload restricted, or might be directed to other areas of study or other activities (eg study elsewhere) or might be advised to seek personal support. This is a *quality assurance process*, which focuses on identification and intervention. How might this process be reflected in a performance indicator?

The process is intended to "encourage excellence in learning" (eg by limiting a student's workload if they are struggling) and to encourage "excellence in achievement" (eg by directing them to study which is more appropriate or achievable). Performance indicators associated with these interventions need to be couched in terms of outcomes which relate to the processes, ie th*e quality effects* – in this case, the proportion of students reviewed who improve their learning (indicated by improved pass rates or improved GPAs) or the proportion of students reviewed who improve their study experience/achievement (which for some might mean exiting and going elsewhere).

This example illustrates how improvement of the outcome (*viz* more relevant or improved study) should reflect interventions in the processes (a review of progress). If the review of progress does not appear to improve study experiences then it, on its own, might be a necessary but not sufficient process. Its value might be in identifying risk, which some other process must then address (eg academic advice; remedial classes; peer support; pastoral support; English language classes).

Schematically, the above example might be presented :

Input	Quality Assurance process		Outcome	Output
Students with poor achievement, defined in terms of GPAs or pass rates	Academic Progress review identifies students at risk	Academic progress interventions e.g, limit workload	Change in achievement	Actual GPAs or pass rates
Potential PIs	Proportion of students identified in Year 1 who take advice or seek support to improve their study (demonstrates the process delivers intended intervention – quality of the process) Proportion of students identified in Year 1 who continue study, who demonstrate positive changes in Year 2 (demonstrates interventions work - quality of the effect) This performance indicator could be further disaggregated to differentiate the groups involved and the nature of the intervention			

While these performance indicators do not demonstrate excellence itself, they are relevant to a university's mission to "encourage excellence" since they highlight a direction towards improvement. At the same time, this particular analysis serves quality assurance purposes by highlighting evidence-based intervention. This simple example also demonstrates that single measures are unlikely to be sufficient indicators of the complexity of quality processes or outcomes.

The above discussion has been oriented primarily around how an institution might use its quest for KPIs to also serve the quality assurance process. This quest might also be reversed, namely how might quality assurance activities inform KPIs, or how might KPIs be developed out of quality assurance processes ?

The first significant quality assurance process which influences student achievement is admission to the university. In New Zealand any school student who meets the [University] Entrance criteria may go to university; so might any citizen over the age of 20. Other entrants include international students with various school qualifications and a small proportion of domestic students with alternative qualifications, or with approved "fast track" entry. Entrants thus demonstrate varied levels of achievement prior to coming to University. Given that first year achievement varies by admission category, any performance indicator which uses aggregate first year achievement as an indicator of academic quality should control for admission variability. That is, one must use *both* the admission profile and the achievement against each category.

Taken a step further, it might be possible to calculate a single measure provided these measures are standardized against either a benchmark cohort, or a model profile. Internal analysis of the constituent data would still be necessary to drive enhancement initiatives but, importantly, the two could be satisfied from the same data analysis.

A simple illustration follows of the way in which first year achievement as a performance indicator might control for admission category composition effects and at the same time allow points for potential intervention to be identified.

If first year achievement in a particular [hypothetical] university's four major degrees is disaggregated it is apparent that those who enter with superior school qualifications (A and B Bursary) perform better in first year than do those who enter with the minimal school qualification or who enter via open entry as adult students. In this example the overall mean GPA for these four degrees, here called "whole university", is an actual 3.722. Mean GPAs for each of the four degrees vary over a range of 0.641 points with BSc students having the highest GPA of 4.042 and BCom students the lowest at 3.401. These might be termed the "crude" means, since they are derived from the uncontrolled distribution of the GPA variable.

However, it is apparent from analysis of admission profiles at degree level that in this example the BSc attracts the lowest proportion of the low-performing adult admission students and the BA attracts the lowest proportion of high performing A and B Bursary students. (Table 1).

If the performance for each degree is standardized to the "whole university" admission profile then differences in admission profile for each degree are controlled. Direct standardization involves applying the outcome variable (in this case disaggregated GPA) in the group being evaluated to a "standard" composition on the variable which is known to affect the outcome (in this case admission qualification). In this example, direct standardization against "whole university" asks, if this particular degree had the admission category composition of the "whole university" and the mean GPAs for each admission group of the individual degree,

what would the mean GPA of the whole degree cohort be?[7] The results shown in Table 2 present a different picture of performance for each degree than appeared to be the case from Table 1.

Table 1. Proportion of admission groups by degree programme[8]

	Whole Univ	BA	BCom	BSc	LLB
A Bursary	0.23	0.20	0.21	0.35	0.37
B Bursary	0.26	0.22	0.27	0.31	0.28
Minimum Entrance	0.25	0.25	0.3	0.22	0.19
Adult Admission	0.23	0.33	0.22	0.12	0.15
Total #	2306	755	506	785	260
Mean GPA	3.722	3.731	3.401	4.042	3.835

Table 2. Degree First Year GPAs Standardized to Whole University

	# Univ	Univ GPA	BA	BCom	BSc	LLB
A Bursary	598	5.7	5.6	5.8	5.8	5.5
B Bursary	594	3.9	4.4	3.9	3.6	3.9
Minimum Entrance	583	2.9	3.1	2.6	2.9	2.4
Adult Admission	531	2.2	2.6	1.7	2.2	1.3
Total	2306					
crude mean GPA		3.722	3.731	3.401	4.042	3.835
rank on crude GPA			3	4	1	2
standardized GPA			3.968	3.557	3.671	3.337
rank on standardized GPA			1	3	2	4

When standardized in this way, the "elite" qualifications according to their admission profiles shown in Table 1 (BSc and LLB) slip in their ranking. When admission qualification profile is taken into account, the mean GPA is highest for the BA and lowest for the LLB. The BSc slips to rank number 2 and the LLB slips to rank number 4 out of the four. Reading across Table 2, one can see of course that BA students do perform better overall in all except the very best admission category, but the point is that this advantage disappears in the aggregate crude, or actual, figure in Table 1, precisely because the BA has a relatively high proportion (32%) of low-achieving adult students – even though adult students in the BA perform considerably better than do adult students in the other three degrees. Or put another way, students in the BSc

[7] This simple calculation for direct standardization of degree GPA = the sum of (GPA of test degree * # in whole University subgroup)/# in whole university.

[8] Figures derived from real data: "other" categories of admission omitted.

and LLB perform better according to their crude GPAs because they are more likely to be high achieving students at entry.

So while one might find GPAs by admission category useful for identifying internal differences, if one requires a summary measure as a performance indicator then the standardized measure is a truer reflection of "quality", taking composition effects into account.

A similar analysis might by applied to GPAs for courses where the proportion who pass (for example) might be standardized to admission category profile. It is "known" that some courses are more likely to attract high achieving entrants, sometimes by reputation, sometimes because of discipline pre-requisites, sometimes because conversely they are courses more attractive to people with life experience, i.e. adult entrants. This customary knowledge is used to explain grade distribution disparities. However it is a subjective assessment. Standardization in this case might provide an objective assessment, since it will effectively control for these "explanatory" effects.

One might ask, "does composition matter?" This paper has already drawn attention to the problems with using retention/attrition and completion data as summary performance indicators. Several of these problems derive from composition impacts. Part-time students, for instance, might be more likely to drop in and out of study than are full-time students. The proportion of part-time students does indeed vary between programmes within an institution and between institutions[9]. Standardizing retention figures for full-time/part-time composition could lessen the likelihood of an institution being identified as having "poor retention rates" or "high attrition rates" if this is at least in part due to high proportions of part-time students. Gray (2008) has used direct standardization and more sophisticated analyses to begin an exploration of controls for composition effects on retention and completion[10]. Similar analyses could apply within institutions when comparing different degree or discipline programmes.

Standardization procedures might be useful for other indicators where composition effects might distort summary measures. The structure and progression rules of many New Zealand degrees are such that it is possible for advancing students to take entry-level courses. Grade distribution in these first year courses might be shown to reflect admission qualification or year of study. In such cases the achievement statistics *for the course* will reflect the student profile. Any institutional KPI for courses must therefore take account of profile differences. While the difference in profile will probably not fully explain differences in outcome (for instance, a final student might well be aiming for a mere pass in a course of marginal importance to her degree), standardization at least *reduces* any distortion created by profile effects.

Other factors known to correlate with achievement, whether or not the relationship is causal, such as ethnicity and gender, might also be controlled in this way where there are significant variations between comparative units of analysis. The process of disaggregating and

[9] Note earlier comments about data quality with respect to definitions of "part-time".

[10] Since Gray's work is based on a single data source, namely the "Single Data Return" institutions make for funding purposes, his analysis does not account for other factors which affect completion and retention as outlined earlier in this paper.

recalculating summary measures opens up the possibility of using the same data set as a quality assurance measure (ie identifying points of variation for possible intervention) and as a more robust performance indicator (controlling for some variation which has a direct influence on performance – for instance, see Scott and Smart 2005 for analysis of variations in completion rates).

A similar approach might be taken when using an aggregate or mean teaching score derived from student survey data as an institutional performance indicator. It has also been observed that some disciplines habitually score more highly than others on teaching scores. At our university this tendency has been addressed by comparing individual scores to a faculty mean, rather than to the whole university mean. Courses identified for commendation or intervention are those with scores which lie more than two standard deviations from the mean *for that faculty*. This allows the score to be used as a quality assurance factor and thus supports enhancement in a focused way. As a summary measure one might aim to improve the faculty mean. But if this score is to be used as an institutional KPI then it should also take account of the variation across disciplines or degree programmes: changes in the institutional performance indicator might be brought about by improvements in courses or might be due to shifts in numbers of courses in certain areas.

For instance, experience might show that Education courses, in total, always achieve higher mean scores than do courses in other degree programmes and Law courses always score lower than courses in other degree programmes. Since scores are calculated per course, a decrease in the number of high-scoring Education course will result in a decrease in the actual (crude) overall score. However if the changed course distribution profile is standardized no change is evident (Table 3). Equally, if the entire low-scoring Law category is removed, then the aggregate score improves – even though there has been no actual improvement in any of the other categories. Conversely, if the number of courses remains the same but scores in two areas improve, the overall score improves (Table 4).

Table 3 Impact of profile composition changes on overall teaching scores.

	Year 1, Base Year Data		Year 2, # of high-scoring Education courses halved	
	# courses	mean score	# courses	mean score
BA	900	4.2	900	4.2
BSc	300	4.1	300	4.1
BE	350	3.5	350	3.5
LLB	120	3.8	120	3.8
BCom	250	4.0	250	4.0
B Ed	1200	4.3	600	4.3
BFA	30	3.2	30	3.2
Crude mean score		4.11		4.06
Standardized mean score				4.11

The decrease in the crude mean score is spurious. The overall standard actually stayed the same.

Year 1, Base Year Data,			Year 2, remove all Law courses	
	# courses	mean score	# courses	mean score
BA	900	4.2	900	4.2
BSc	300	4.1	300	4.1
BE	350	3.5	350	3.5
LLB	120	3.8		
BCom	250	4.0	250	4.0
B Ed	1200	4.3	1200	4.3
BFA	30	3.2	30	3.2
Crude mean score		4.11		4.12

The apparent improvement in course quality is spurious and due entirely to the removal of Law – the standard for the remaining courses is unchanged.

Table 4. Impact of changes in scores, with and without composition changes, on overall teaching scores

Year 1, Base Year Data,			Year 2, no change in composition; Improve scores for BA and BE	
	# courses	mean score	# courses	mean score
BA	900	4.2	900	4.3
BSc	300	4.1	300	4.1
BE	350	3.5	350	3.8
LLB	120	3.8	120	3.8
BCom	250	4.0	250	4.0
B Ed	1200	4.3	1200	4.3
BFA	30	3.2	30	3.2
Crude mean score		4.11		4.17

This is a real change, ie not due to composition effects.

Year 1, Base Year Data,			Year 2, # of Education courses halved Scores for BSc and BE improve	
	# courses	mean score	# courses	mean score
BA	900	4.2	900	4.2
BSc	300	4.1	300	4.2
BE	350	3.5	350	3.6
LLB	120	3.8	120	3.8
BCom	250	4.0	250	4.0
B Ed	1200	4.3	600	4.3
BFA	30	3.2	30	3.2
Crude mean score		4.11		4.09
Standardized mean score				4.13

The standardized mean shows that decrease in crude mean score is spurious: even though there has been a reduction in the number of high scoring courses the overall standard has actually improved.

While these results might all be self-evident from the Tables, the point is that the differing actual (or crude) figures reflect different forces at work in producing changes in overall score. Comparison of aggregate measures across institutions or across time is robust only so long as composition remains stable. If composition varies either through time or across institutions then this must be controlled for. Direct standardization is a simple way of exerting this control.

Summary

This paper has reviewed a range of performance measures currently used at one New Zealand university, and a range of related quality assurance processes. The paper has indicated that a number of aggregate measures imply little about institutional academic quality either because they relate to very small subsets of university experience, or because they are constrained by factors beyond quality assurance influence, or because they mask the actual activities which lie behind the measures.

The quality agenda is challenged to demonstrate relevance and the utility of strategic performance measures for guiding and possibly facilitating enhancement. The paper has suggested that performance measures might be more valid and more useful if they start from the academic quality activities – rather than the available institutional data – and work towards the aggregate in ways which (a) enable points of intervention to remain visible and (b) control for composition effects. It is consistent with the quality assurance approach that KPIs might also include factors which are known to correlate with excellent learning and teaching. Few of these, however appear to lend themselves easily to the kind of numerical representation which compliance authorities seek. But they might be employed to formulate aggregate KPIs by defining the composition of the cohorts being assessed or compared.

In outline, a protocol for the development of KPIs in this way is:

1. identify correlates of variation in the "quality" to be measured ;
2. identify how the quality variable might be measured (check validity and reliability);
3. check for composition differences on the correlate variable;
4. if differences emerge, standardize for these;
5. if there are composition differences calculate aggregate KPI on basis of standardized measure;
6. use variation in measures on the correlates to identify potential points of intervention for enhancement.

In summary, the protocol develops the KPIs from the quality assurance processes and because potential for disaggregation remains, these same measures may be used as directions for quality enhancement, thereby closing the loop. The process rests on foregrounding quality assurance, to serve the demand for performance indicators, rather than vice versa.

In a similar way it is suggested that institutional KPIs might usefully include measures of processes and intended outcomes which are directly related to quality assurance initiatives. Making visible the impact of such activities should serve the strategic agenda of institutions more effectively than will crude aggregate measures which have remote connections with academic quality.

Acknowledgments

I appreciate the support and advice of Ann Dupuis (School of Social and Cultural Studies, Massey University), and Heather Dickie (Manager, Academic Quality Assurance Unit), Bob Hall (Director, Planning, Information and Reporting Unit) and Kaylene Sampson (University Centre for Teaching and Learning) of the University of Canterbury.

References

Chalmers, Denise 2007 *A review of Australian and international quality systems and indicators of learning and teaching.* Carrick Institute of Learning and Teaching in Higher Education, NSW, Australia. August 2007.

Coates, Hamish 2007 Excellent measures precede measures of excellence. *Journal of Higher Education Policy and Management* Vol 29 (1), March 2007, pp87-94.

Controller and Auditor General 2008 *The Auditor-General's Observations on the Quality of Performance Reporting.* Discussion paper produced under section 21 of the Public Audit Act 2001, Office of the Auditor-General, Wellington, NZ, June 2008. 60pp.

Cooper, Trudi 2002 Why student retention fails to assure quality *Research and Development in Higher Education, Vol 25 pp152-158.*

Gray, Alistair 2008 *Statistical methodology for taking account of diverse learner characteristics and courses.* Unpubl paper prepared for the Tertiary Education Commission. Statistics Research Associates, Wellington, NZ. 13 November 2008.

Centre for Higher Education Research and Information 2008 *Counting what is measured or measuring what counts? League tables and their impact on higher education institutions in England.* Report to HEFCE, April 2008/14.

New Zealand Qualifications Authority 2008 Consultation Document "Policy Requirements for an Evaluative Approach to Quality Assurance in Tertiary Education Organisations". NZQA website http://www.nzqa.govt.nz/for-providers/tertiary/consultation/index.html downloaded 20 February 2009.

Scott, David and Smart, Warren 2005 *What Factors Make a Difference to Getting a Degree in New Zealand?* Ministry of Education, Wellington, NZ, October 2005 36pp.

Scott, Dugald and Cameron, Jan 2008 *Quality Assurance Processes in New Zealand Universities – An Overview* Paper to Assuring the Quality of National and Transnational Higher Education Programmes: An International Perspective, Chennai, India, 6-7 March 2008. 12pp.

Technical Working Group 2004 *Introduction of a Performance Element to Tertiary Education Funding* Ministry of Education, Wellington, NZ December 2003 (publ January 2004).

Yorke, Mantz and Longden, Bernard, 2004 *Retention and Student Success in Higher Education* Society for Research into Higher Education and Open University, Maidenhead.

Zepke, Nick; Leach, Linda; Prebble, Tom et al 2005 *Improving Tertiary Student Outcomes in the First Year of Study* Unpublished Report, NZ Council for Educational Research, Wellington, NZ, 26pp.