A FRAMEWORK FOR EVALUATING e-LEARNING ENVIRONMENTS

Frank Fursenko, G. Stewart von Itzstein

ABSTRACT
The management of many universities is in a state of transformation from the traditional collegiate, to a more structured corporate style of management. These universities are increasingly modelling their educational administration structures on structures used by business enterprises. At the same time there has been a remarkable technology shift which has enabled effective computer based learning linked to sophisticated online delivery mechanisms. Academics and educational managers are embracing the new technologies for a variety of reasons not all of which focus on improving educational outcomes. The transformation of universities from traditional to corporate management which has developed in conjunction with the technology shift has produced a whole new set of problems and opportunities and educational managers have yet to capitalize on the opportunities. This paper identifies these problems and opportunities and provides a framework for taking more control of the e-learning process in order to provide a better outcome for both students and educational managers. This paper initially examines the problems inherent in online delivery of both traditional university education and competency-base vocational training. We also present our research into student evaluations and how they can lead to false metrics. The paper then identifies better evaluation mechanisms for quality control. These mechanisms are presented as a unified framework for the ongoing evaluation of e-learning outcomes.

KEYWORDS
e-learning, metrics, framework, evaluation.

INTRODUCTION
A decade after the dotcom hype predicted the empty lecture theatre student demand for interactive learning is driving online education into its next stage of development (Palloff, Pratt, 2001). Universities have been warned to adapt to the next generation of online learning or lose ground to private providers, who with the advantage of flexibility and have, overtaken many universities in developing expertise in e-learning and are even offering post-graduate diploma programs in e-learning¹. Most universities had failed to undertake the rigorous research and clear thinking (as exemplified by university policies and administrative processes) necessary to understand the effective use of e-learning. Apart from a handful of universities there is no set agenda, no set policy and no clear understanding as to how e-learning should be designed or delivered. In particular there is no clear understanding of proper evaluation techniques. Dr Stephen Quintin²

Compounding these problems is the fact that e-learning is not seen as a discipline in its own right. Often it is seen as an add-on in much the same way that distance education was seen in the 1990s as an add-on to normal face-to-face delivery (Fursenko et al, 2003). Universities have failed to fully recognise that

the way students get their education has changed. It can no longer be assumed that students will submit without question to a routine that requires that they attend all lectures and then undertake exercises to prove they have absorbed the knowledge imparted by the speaker. Students are fully aware of the technology shift that has radically changed the face of education and has given them many new options.

The main challenge facing many universities is to rethink what face-to-face delivery is good for and then work out how to get the right blend of face-to-face and online education (Brabazon, 2002). Universities lack a framework that will enable them to design and evaluate programs such that they can harness the benefits of one delivery mechanism without losing the strengths of another. Without a proper framework it is difficult for universities to identify the real educational needs of students and therefore to provide better outcomes to students and better evaluation mechanisms for quality control.

**A FRAMEWORK FOR PLANNING AND EVALUATING PROGRAMS**

This research is proposing a framework for the evaluation and planning for e-Learning environments. The framework consists of four models: Each of these models work in conjunction at different levels of abstraction to support an evaluative framework.

1. Stakeholder Model
2. Strategic Model
3. Operational Model
4. Abstract Student Model

As we will show in this paper the main benefit of this framework is in evaluation. The framework is designed to guide us in setting parameters for measuring the development of graduate qualities, developing subject-specific competence and diagnostic type measures. Done properly it will lead to better outcomes for students, academics and administrators by bringing together as far as possible student self-evaluation, teacher evaluations and administrator evaluations.

**MODELS**

The framework supports the evaluation planning that must encompass the differing needs of the three key stakeholders as illustrated in the Stakeholder Model. The four models are used to explain how we can plan and operate a common set of metrics that can be presented to these stakeholders.

**Stakeholder Model**

The Stakeholder Model (see Figure 1) illustrates the differing viewpoints or perceptions of metrics (measures of educational outcomes) by the three key classes of stakeholders in universities; students, academics and administrators. This model is designed to represent a top level view of our framework and can be considered as analogous to an entity relationship model. This level represents who are we measuring for?

For academics involved in teaching the most important metrics are those that can be used in the evaluation of students both individually and as cohorts (or classes). The realities of tertiary teaching - large classes, multiple campuses, online options, etc – often require academic teaching staff to adopt and use a conceptual (or abstract) student model for the purposes of evaluation. That is represented by the use of scaling of students to a curve, the curve representing an ideal average student undertaking the course.

Students need metrics for reflection on their own performance in both absolute and relative terms. For example “how am I going?”, “How am I going in comparison to my classmates?” etc. They are at the core of most evaluations, measures and surveys but there needs to be meaningful measures of performance. Currently the metrics are poorly served by academic teaching staff and administrators largely due to the lack of established frameworks for evaluation. In many universities the only official evaluation of performance a student receives is a transcript with minimal information – often one
indicator of performance (usually a percentage with a grade but often only a grade) per course and an overall grade point average (GPA).

Administrators in general (and senior decision-makers in particular) are rightly focussed on measurements of performance. They want metrics that measure how well the university is doing in absolute terms (using standard key performance indicators), in relative terms (using monthly, yearly or 5-yearly figures) and in comparative terms (against other national universities or categories of national universities or against international universities). Administrator-motivated metrics are essentially number-driven and therefore the strong emphasis on key performance indicators (KPIs).

![Stakeholder Model](image)

Figure 1. Stakeholder Model

**Strategic Model**

This model (see Figure 2) incorporates various strategies for planning the evaluation of undergraduate degree programs. This level of the framework represents what we are going to measure. Evaluation planning needs to be conducted concurrently on 3 different levels:

1. **Program (Long Range) Level.** The two main strategic goals of tertiary education are to develop a student to the highest possible educational (professional) standard and to produce a “work-ready” graduate who has the specified graduate qualities.

2. **Course (Medium Range) Level.** At this level the main concerns are tactical rather than strategic and they are of most importance to academics who teach the courses. Here the focus is on how well this course (or subject) is taught to a cohort (or class) but we it also must allow the development of selected aspects of graduate qualities.

3. **Task (Short Range) Level.** At this level we evaluate how well an individual task is taught how well it is absorbed. Diagnostics such as tests and quizzes whether online or otherwise are critical at this level. Cultural factors and assumptions of skill levels need to be examined carefully at this level.

The short range evaluation feed into the medium range evaluation strategies and in turn these evaluations feed into the long range evaluation. Each evaluation short range, medium range and long range seeks to baseline the student then measure their progress along the way. This leads to a life of student approach to evaluation where the focus of evaluation is how much difference have we made to that student and in turn to all students undertaking our programs. Or more simply what is the sum value benefit that we are adding to our students. The sum average of value that we add to all our students indicates the performance of our teaching. Measurement of performance is hence a more direct measure than the current approaches.

More specifically, from the perspective of the classes of metrics (how do we measure these performance indicators?):

174
1. The Program level requires a framework (or strategy) that incorporates graduate qualities and a well established taxonomy for education such as Bloom’s Cognitive Domain.

1. The Course Level requires a framework that can outline how both academic and professional skills can be developed. It also requires an appropriate course evaluations mechanism.

2. The Task Level requires a framework that can incorporate techniques for measuring how well students are absorbing content and developing basic skills on a week by week basis.

Figure 2. Strategic Model

Operational Model
The Operational Model (shown in Figure 3) deals with the vectors that are used to undertake the evaluation of the three levels listed in the Strategic Model. That is, how do we evaluate the students’ performance?
For different delivery modes (face-to-face, distance and e-learning) we will require a different mix of vectors. Generally the main vectors that are currently used in tertiary education can be classified and described in the following ways:

- Surveys generally ask users for their opinion on various issues and therefore can be very subjective. Surveys are a very useful vector for measuring perceptions and evaluating satisfaction levels but they are not suitable for measuring quality or any educational value adding that has occurred in a course or program. Furthermore there is the very real question as to whether the students have developed the emotional maturity as outlined in Bloom’s Affective Domain to respond appropriately to questions that ask them to evaluate their program, a specific course, a lecturer or aspects of course delivery.

- Hard metrics are derived from raw data. Examples are pass rates, completion rates and grade point average and in most cases hard metrics would be considered objective rather than subjective. Meaningful comparisons can be more easily made using hard metrics rather than soft (or scaled) metrics because the baselines for these metrics are easy to establish.

- Soft metrics are essentially hard metrics that have been scaled in some way. These metrics therefore represent qualitative measurements as opposed to the purely quantitative measurements we get from hard metrics. It should be noted that data (as represented by soft
metrics) cannot be extrapolated to the entire population hence are of little value when attempting to statistically analysis/project populations of students.

- Diagnostics are usually understood to be tests (whether formative or summative) but should also include practical classes and in some disciplines (such as Information Technology), tutorials as well.

- Situational awareness is the vector that attempts to determine how aware the subject is of the domain they are in. In a military situation, for example, we can apply situational awareness by stopping an exercise at some point and asking a barrage of questions to determine how aware the subject is of the domain they are in. Situational awareness can be measured in all major delivery modes using a variety of tests such as oral questioning during lectures in the case of face-to-face or pop-up tests in e-learning situations.

- Assignments which may include group or individual course-specific projects.

- Final year project which attempts to evaluate a substantial part of the work done in an entire undergraduate degree program.

- Work experience which allows the student to demonstrate his or her competency in many aspects of their undergraduate training.

**Abstract Student Model**

By changing the focus of our evaluation from the course to the student we provide a number of significant opportunities for better-targeted evaluative strategies and metrics. In this research we have invented the concept of the abstract student (S0). The abstract student is a hypothetical student undertaking a course within the school. The abstract student can be configured to represent a number of variants. For example S- is an abstract student representing the lower quartile group, conversely S+ represents the upper quartile student and finally the basic average student is represented by S0. An abstract student (S0) can be the average student from a subject, cohort or program group. By cohort we mean a group of students that are travelling through the degree together (taking the same subjects at the same time). We define these as, average abstract student from subject S0S, average abstract student from cohort S0C and abstract student from program (sometimes referred to as a degree) S0P. Finally we also define an abstract student as they were at the beginning of the subject or program and a student at the end of the subject or program. This is denoted by a B or E at the end of the definition to indicate beginning and end respectively. By defining these types we can make more realistic evaluations of our tasks/subjects and graduates. We use this symbolism for conciseness when generating analysis of our students’ progress.

To generate the abstract student for the group in question we use a statistical method to get the average raw marks for a group of students in that course. For example, the abstract student for the group of students undertaking the subject ethics301 would be the average raw mark of every student undertaking that subject S0SB<grade>. The average exam mark for the abstract syntax student would be the average raw exam marks for all students doing that subject from their previous subjects S0CB<exam mark>. Any statistic that can be collected for an individual student could then be represented in the abstract student. One point to note is that it is vital that we use the raw results of the student rather than the scaled result normally used for constructing their final grades. That is because the scaled results are designed to give a comparative mark for the student assuming the abstract student in a particular subject has a result that sits in the bell curve of standardized grades.

By collecting this information we are able to better evaluate the progress a student, a group of students, or all students in a program. For example by comparing the S0SE<exam> to the actual raw average for the course we can see if the course is performing well in comparison to the performance of that particular group of students. If we were to make a comparison of the scaled results we would end up with the same result for every course due to the scaling factor.

**Example Use of Abstract Student Model (ASM)**

This is an example of the need for the ASM; a model capable of delivering some quantitative measures or metrics.
Consider a course called User Interface which is compulsory courses in all undergraduate IT programs including the elite LHIS program. UI appears in all program schedules as a semester 1 course but is also delivered in the second semesters for pragmatic reasons. Most of the students in the first semester cohort are those who are progressing normally through their various degree programs and especially the selected students in the elite program LHIS. For argument sake, assume that the failure rate for this cohort is 20% which (according to the university KPI on pass-fail rates) is deemed to be acceptable. In the second semester cohort we will have students who failed in the first semester, other students who are not progressing through their various degree programs at the normal rate due to poor performance, part-time students, and transferees from other programs. Also there would not be any students from the elite program. For argument sake, assume that the failure rate for this cohort is 40% which according to the university KPI is far too high and clearly not acceptable.

A comparison of first and second semesters obviously suggests poor teaching performance in the case of the second semester delivery. But, if we could establish an ASM cohort for both semesters it may show that:

1. the pass rate of the first semester cohort should have been 90%.
2. the pass rate of the second semester cohort should have been 50%.

The proper conclusion is that the first semester cohort performed below expectation and the second semester cohort performed above expectation. Whether this can be put down to teaching performance is another matter but the ASM would probably support this conclusion.

**Architecture of the Framework**

The interaction between the 4 models of the framework is clearly illustrated in Figure 4. The Stakeholder Model represents the top level relationship between stakeholders and illustrates the “WHO” of the framework. The Strategic Model deals with evaluation planning and illustrates the “WHAT” of the framework. The Operational Model illustrates the “HOW’ of the framework. Finally the abstract student model (ASM) ties all these models together to allow analysis of collected data.

**THE CASE FOR A FRAMEWORK TO EVALUATE OUTCOMES**

This is a framework for evaluation. It does not provide the questions - instead it provides a mechanism to ensure that the entire spectrum of requirements are met. Indeed the framework cannot provide the questions simply because there is no panacea for an ultimate set of questions that will cover everything and every context of use.

The University of South Australia provides a Student Course Evaluation Survey (SCES) that all students are requested to complete at the end of each course that they take. This questionnaire (shown below) consists of a standard set of ten questions that is the same for every course within the university regardless of its content, student experience, student culture, structure or mode of delivery.

**Student Course Evaluation Survey**

1. I have a clear idea of what is expected of me in this course.
2. The ways in which I was taught provided me with opportunities to pursue my own learning.
3. The course enabled me to develop and/or strengthen a number of the qualities of a University of South Australia graduate.
4. I felt there was a genuine interest in my learning needs and progress.
5. The course developed my understanding of concepts and principles.
6. The workload for this course was reasonable given my other study commitments.
7. I have received feedback that is constructive and helpful.
8. The assessment tasks were related to the qualities of a University of South Australia graduate.
9. The staff teaching in this course showed a genuine interest in their teaching.
10. Overall I was satisfied with the quality of this course.

Four of the ten questions in this survey (shown underlined) relate to Graduate Qualities. The other questions judge the students’ perception of their educational experience and in this way asked students to evaluate the course (and the lecturer). This style of questioning assumes that students have developed emotional and professional maturity as described in the five major categories that form Bloom’s Affective Domain. This domain includes the manner in which we deal with things emotionally, such as feelings, values, appreciation, enthusiasms, motivations, and attitudes (Krathwohl, Bloom, Bertram, 1973).

![Diagram of Models Interaction]

Figure 4. Framework Showing Interaction of Models

Surveys such as the Student Course Evaluation Survey shown above are typical of the surveys used in many universities and often form the basis for evaluating the performance of academics and of the university. Furthermore, similar surveys are often used to produce metrics for key performance indicators which in turn are used by administrators to evaluate the success or otherwise of management objectives. The many assumptions that are made in constructing such surveys because of inadequate evaluation planning presents a strong argument for the type of framework that we present in this paper.

**Main Criteria for Evaluation**

The Strategic Model is the most critical model in our framework as it deals with evaluation planning at 3 different levels; the program level, the course level and the task level. The Strategic Model must be based on well-established criteria, especially at the program level.

The major criteria we have chosen and used in our framework are:

1. Bloom’s Taxonomy – the Cognitive Domain
2. Bloom’s Taxonomy – the Affective Domain
3. The University of South Australia’s Graduate Qualities.

Bloom’s Taxonomy is the best known and most widely used classification of cognitive learning objectives (Anderson, Sosniak, Rehang, 1994). Learning is organised as a series of levels or prerequisites, and suggests that higher learning levels cannot be addressed until those below have been
Learning becomes effectively serial in structure. The model includes six levels of thinking: knowledge, comprehension, application, analysis, synthesis and evaluation (Anderson, Krathwohl, 2001) and the realization of all sub-domains represents an optimal educational outcome. Bloom’s Taxonomy provides tools to assess level at which students are assessed by identifying the specific requirements of assessment tasks. Bloom’s Cognitive Domain is a vital criterion in our Strategic Model.

The “synthesis” and “evaluation” sub-domains (in Blooms Cognitive Domain) are the most abstract and their development will spans many courses and several year levels. The evaluation planning of these sub-domains is part of the program level of our Strategic Model and the most relevant vectors at this level are the final year project and work experience. The “application” and “analysis” sub-domains are the principal foci of evaluation planning at the course level of the Strategic Model. These two sub-domains are well researched and consequently there are many suitable vectors (such as assignments) to apply to the evaluation planning for these sub-domains (Anderson, Krathwohl, 2001). It must also be noted that an appropriate Abstract Student Model is essential to any evaluation planning at the course level. The lowest two sub-domains of Bloom’s Taxonomy, “knowledge” and “comprehension” are central to evaluation planning in the task level of the Strategic Model. They can be measured using a variety of vectors that deal with diagnostics but they must be designed carefully. A lot of work has been done in this area on different diagnostics that are appropriate in face-to-face situations, self-paced video presentations and in e-learning (Gillani, 2003).

The Strategic Model also requires that, in evaluation planning we do not assume that students possess sufficient maturity in all categories of Bloom’s Affective Domain. This domain deals with emotional things such as feelings, values, appreciation, enthusiasms, motivations, and attitudes. This taxonomy unfortunately is usually ignored by course developers and “tertiary-level” educators at all levels and yet it is crucial to the validity of surveys that form the basis for many surveys (Krathwohl, Bloom, Bertram, 1973).

A growing number of universities (including the University of South Australia) have defined a comprehensive set of graduate qualities which is summarized as follows:

On graduation, a student of the University:

1. Operates effectively with and upon a **body of knowledge** of sufficient depth to begin professional practice;
2. Is prepared for **lifelong learning** in pursuit of personal development and excellence in professional practice;
3. Is an **effective problem solver** capable of applying logical, critical and creative thinking to a range of problems;
4. Can **work autonomously and collaboratively** as a professional;
5. Is committed to **ethical action and social responsibility** as a professional and citizen;
6. **Communicates effectively** in professional practice and as a member of the community;
7. Demonstrates **international perspectives** as a professional and as a citizen.

And, each of the seven graduate qualities can be decomposed as shown below in the case of Graduate Quality #3 (Is an effective problem solver capable of applying logical, critical and creative thinking to a range of problems.) A graduate will:

- Gather, evaluate and deploy relevant information to assist problem solving – i.e. analysis and synthesis;
- Define researchable questions in the discipline or professional area;
- Initiate creative responses to problems and frame such responses as opportunities;
- Apply strategies to conceptualise problems and formulate a range of solutions;

In our previous research (Fursenko, Gelade, Dansie, Itzstein, Li, Wahlstrom, 2005) we showed how graduate qualities were designed to shape the teaching, learning and assessment that make up a program and, we focussed on the issue of embedding graduate qualities into course and program objectives. This
work resulted in the development of a data-based mapping tool (based in part on Bloom’s Taxonomy) that can be used to assess the extent to which graduate qualities have been achieved in a course and, by extrapolation, in a program. Finally this research also assessed the degree to which the embedding of graduate qualities into course and program objectives produced outcomes that aligned with the requirements of the modern knowledge economy in which most graduates work.

In another paper Fursenko, Itzstein (2003) we also showed that, from a management perspective, graduate qualities introduce a framework for defining an educational product and therefore give management a range of new key performance indicators that can be applied in performance management and give substance to educational management by objectives.

VISION OF FRAMEWORK IN USE

We provide an example of the use of the framework (with particular reference to Figure 2) by academic staff, university administrators and the students themselves. It is envisioned that when students begin their degree programs they are required to undertake a series of short tests (during the Orientation Week that is common to many universities) to determine baseline metrics for the framework and for the Abstract Student Model:

- A personality test to establish their initial attitude, motivation and values, all of which are critical aspects of Bloom’s Affective Domain and will vary considerably depending on their maturity and past experience. Students enter university through a variety of different pathways; directly from secondary school, after completion of post-secondary training programs or mature-age entry where most students have had some work experience in various industries. The results of these tests will give us significant insight into the responses students give to survey questions that often include evaluation of course materials and delivery, programs and lecturers.

- A domain test to establish a student’s knowledge of their particular program of study, in addition to general and related topics.

- A graduate qualities test to establish a student’s ability in written comprehension and communication, problem solving and understanding of the complexities of working in groups or teams.

Academic staff will be asked populate the software that supports the Abstract Student Model with all assignment and test marks as well as the examination marks and to record the graduate qualities that each assignment was designed to develop. This will be done for each course. The university already requires all academic staff to state, for each assignment, which graduate qualities will be developed or tested and provide relative weightings. As the student cohort progresses through the program, academics will be able to compare each student to their previous result history to determine whether a particular student is achieving below their normal level and require remedial work. The academic can also determine the average improvement of all students, compare that to their normalized performance and gauge if the course is too challenging or not challenging enough.

University administrators can use statistical measures to understand how a student performs in a course relative to their previous performance and can identify courses that add most value to the students based on actual performance. Anomalies such as yearly fluctuations for different intakes and teaching staff can be put into proper context. Administrators also can determine whether they need to provide workshops or seminars for students to further develop any aspects of graduate qualities in the second or third years of a students program.

Finally students will be able to access the system periodically and monitor their own progress in relation to their fellow cohort members, course mates and program participants. This allows them to understand their progress and whether they are improving relative to other students in their program.
CONCLUSION

In the past decade we have witnessed a remarkable technology shift that has delivered the capacity to transform education and create a new paradigm. At the same time we have seen the management of many universities transformed from the traditional collegiate style to a more structured corporate style characterized by key performance indicators and management by objectives. The opportunities presented by e-learning have, in many cases, not been realised because the objectives derived from key performance indicators were poorly evaluated and technology was used inappropriately to simulate old teaching methodologies rather than embracing e-learning as an opportunity to deliver a new and innovative teaching paradigm. The convergence of technology and educational management has not resulted in the best learning experience for students in many universities.

Our previous research (Fursenko, Itzstein, et al. 2003) clearly identified the advantages of e-learning as the multi-vector interaction modes provides the ability to give equality in learning experience irrespective of campus and, making programs accessible for students irrespective of where they live. The “more advanced aspects” of e-learning – where students have web-chats and where they have some form of pod-casting and where they can actually see the other person speaking – significantly mitigates the impact of not having face-to-face teaching experience. It is widely acknowledged that providing these features is a costly exercise but the start-up costs of e-learning would be offset by long-term savings as face-to-face teaching declines.

In this research we have addressed some of these issues by providing a framework for evaluation of student outcomes. Current evaluation techniques are far too subjective and often do not focus on what should be the primary goal of the evaluation that is “How much benefit did we provide for the student?” For example students are routinely asked to evaluate a course rather than to evaluate their own progress within that course. In properly designed evaluations a student’s progress should be the primary goal of the evaluation and the evaluation of the course a secondary consideration. Asking students to evaluate courses will always be subjective and non-expert (Ebel, Frisbie, 1991).

This framework acts as a guide for designing evaluative systems used for analysing the total benefit of the university education experience to the student. This enables students to reflect on their progress, academic staff to evaluate and understand their own performance as educators and administrators to measure the university’s performance. In the end we as facilitators of education are most interested with the sum total of value we add to our students’ lives. If we can measure this more directly we will be able to react quicker and more appropriately and hence be more effective.

REFERENCES


Gelade, S. and Fursenko, F. (2006) Can Intrinsic Qualities be Developed through Assessment? Mapping Assessment Practices in IT Degree Programs. To be published in a book titled “Enhancing Teaching and Learning through Assessment” which will be published on the Assessment Resource Centre (ARC) website, the official website of the First International Conference on Enhancing Teaching and Learning through Assessment, the Conference CD-ROM and in other forms of conference materials.


Note: The full PowerPoint conference presentation of this paper can be viewed at http://www.vonitzstein.com/research/teaching

Frank Fursenko and G Stewart von Itzstein
School of Computer and Information Science
Mawson Lakes, Adelaide
Email: frank.fursenko@unisa.edu.au

G Stewart von Itzstein
School of Computer and Information Science
Mawson Lakes, Adelaide
Email: stewart@vonitzstein.com