Professionalising academics: A review of issues in higher education assessment systems

Anil Hira and Darcie Cohen

Simon Fraser University

e-mail: ahira@sfu.ca, tel. 778 782-3286

revised: Mar 1, 2011

We argue that academic systems are poorly designed from the point of view of evaluation of results and accountability and transparency of systems. The basic management literature suggests that performance stems from the development of clear strategic targets and priorities; appropriate administrative and incentive structures to support, enhance, and evaluate performance; and the development of collective values towards continual improvement on an institutional, not just individual, level (Wang, 2010, p. 10). Through a review of North American graduate training and placement; tenure, promotion, and salary review; and the particularly problematic aspects of teaching and service assessment, we find a series of vexing challenges that need to be tackled. We suggest the direction of systemic changes that could lead to collective improvement of academic performance.

Keywords: evaluation, assessment, teaching, research, service

Introduction

Higher education plays a vital role in society in terms of creating both good citizens and a capable workforce. Academics are thus key actors in socialisation, adaptation, innovation, and self-reflection in society. It is no mistake that most academic enterprises are public enterprises, reflecting the collective goods nature of this social enterprise, including continual improvement. The triumvirate roles of teaching, research, and (public) service all dovetail in theory to serve together to accomplish this mission.

Given the attempts to move towards more accountable management practices in the public sectors, reflected in the "new public management," and "results-based management" approaches, one might expect similar evolution in higher education. Indeed, increasing global economic competition has led to more focus on improving educational outcomes, as reflected in the international PISA tests ranking country educational attainment at lower levels, a raft of new global university rankings, and increased experimentation in educational systems, such as school choice. World university rankings continue to show a solid performance for North American, and particularly U.S. universities, but these ranking systems rely primarily on citation counts. In other words, even in an age of global competition, the old adage of "publish or perish" still holds, too often to the neglect of teaching and service. We argue that the root of the problem lies in fundamental shortcomings in the various systems of academic assessment. We back up this argument with observations from a decade of personal experience at 5 different North American universities, a review of the literature, and primary data gathered from Canadian Political Science departments about teaching and service assessment. Though we are focused in our evidence on North America, the literature indicates that such issues are of universal interest.

In fact, despite recent efforts by professional societies to develop ethical guidelines for fair academic assessment, they remain more in the realm of statements of principles rather than recommendations for accountable procedures. Because of increasing job competition as well as a lack of guidelines for assessment, academic staff members report increasing stress and declining job satisfaction (Cooper & Nojima, 2002, p. 168).

There are many reasons to question whether citation counts really reflect improvement. As the pressure to publish continues to accelerate, there is a natural increase in the number of outlets, leading to even more pressure to specialise and more difficulty in keeping up with research. The proliferation of new publication outlets also leads to an emphasis on short-term quickly published research (Bruneau & Savage, 2002, p. 84). The natural result is that the overall quality of research and the sharing of results across the fields (as all evaluation is still done by disciplinary departments) declines, as long-term research becomes reserved for the chosen few who can garner large grants. Policy impact becomes less relevant, and new ideas (as opposed to adding more proof to existing literature) highly risky and poor investments of time. There is even less time for those who do consider teaching and service to be important obligations (though not proportionately rewarded) to act upon such convictions.

Such internal pressures continue in the face of accelerating countervailing winds of public calls for some form of accountability in higher education. As Bruneau & Savage (2002, p. 58) state,

> A public university or college is, we assert, truly accountable when it ensures members of the public can readily survey teaching, learning, inquiry, and governance. It is accountable when open to scrutiny, open to participation, open to criticism, and demonstrably committed to critical research….

These authors (p. 61) go on to note that the desire for information and accountability is what explains the proliferation of ratings at the national and global levels. Such ranking systems are statistically inaccurate, and fraught with "one size fits all" and other problems. They tend to reinforce already-resourced institutions, and somehow seem accurate with an occasional climber, often based on the definition of a particular criteria rather than a distinctive policy (Harvey Editor, 2008).

What criteria should be used instead to assess how academics reach their 3 missions? We assert the principles can be derived from those for good governance and research. By good governance, we mean commonly shared principles of management, such as: wise resource allocation, rule of law, equity, accountability, transparency, information flow and accessibility, the ability to plan for the long-term and measure progress, the ability for an organisation to set collective goals and learn collectively, stable and clear duties and steps for promotion, and clear and stable rules for the succession of leaders. By good research, we mean the principles taught in Methods 101, such as: clear conceptualisation, identification of variables for measurement, validity, reliability, and accuracy of measurement, and coherent logical and empirical testing as part of an ongoing evaluation process. Using these criteria as a guide, we find present academic assessment systems in North America require major reforms.

Problems with training and placing new academics

There is widespread concern that we are training too many PhD students for academic posts. The breakdown of the system is reflected in the well-documented augmenting exploitation of sessional, part-time, and graduate student teachers, developing into an academic underclass with limited prospects for competing with recent graduates from the high reputation universities, and limited long-run ability to improve their skills. The reason for a continued and perhaps increasing glut of PhDs is a simple reflection of incentive structures. Departments without PhDs are not taken as seriously in research competition and thus have a much harder time securing grants. Most departments are rewarded based on the number of graduate students they have, and graduate students provide a very cheap labour pool for teaching and research assistance. At the graduate level, this means taking on students who may or may not be suitable and/or may not be a good fit for the Department. However, graduate training is rarely rewarded.

There is no doubt that the number of part-time academic staff members has ballooned in proportion to the number of full-time members in North America, reflecting the continually increasing number of graduates in all fields. Some estimates put the increase at 5X as fast as full-time members from 1970-2000. However, while supply increases, demand is shrinking. The fact is that the average age of full-time academic staff has increased dramatically. Positive trends in terms of the rising proportion of women and minorities are one exception to this downward cycle (Schuster & Finkelstein, 2006, pp. 40, 47, 58, 71 & 163).

"Casualisation" of the academic labour market reflects not only over-supply, but the trend of using non-tenurable contracts as well as strictly lecture and adjunct positions (Schuster & Finkelstein, 2006, pp. 175 & 232 and Louis, 2009). In fact, there are now 3 possible "career" paths: tenure-track, renewable contracts, and non-renewable contracts (Gappa, Austin, & Trice, 2007, p. 15). Yet, most academic departments continue to allow the hiring process to be vulnerable to ideological and personal politics. For example, one does not find in a general description for comparative politics that a significant minority want a game theorist or another faction a Marxist. Many think they should only hire candidates from departments with higher reputations, but can not state such discrimination openly. Therefore part of the hiring process is discerning and navigating the winning "fit" for a coalition that will vote for the candidate. Sometimes this happens by consensus, where there clearly is a superior candidate who also fits. But in today's job market, it is not uncommon to see candidates with less publications, teaching experience, or real life work experience hired for political reasons (i.e. "least objectionable"). There is therefore no learning curve from failure in many aspects of academics, including failure to find a position. Thus, while "fit" is certainly an important criterion, the lack of clear hiring procedures or results, and the fear of lawsuits preventing the adoption of such a system, open hiring up to abuse and continual factionalisation of departments. It is no wonder therefore that, particularly in large research departments, there is a tendency to move towards a strong homogeneity of perspectives, just the opposite of what a lively intellectual life and strong teaching call for.

The perverse market incentives therefore call for greater information about labour markets and higher level intervention. Occupational outcomes for graduate students need to be monitored by discipline over time and passed on to potential graduate students so that they can make informed decisions. Provincial and federal governments might need to limit the

number of departments creating graduate students (particularly PhDs) and/or limit the total number in order to restrict supply to more realistic levels. As Mann and Nelson Hochenedel (2006, p. 299), state,

> "This may seem heavy handed, but it's no worse than churning out 100 graduates in discipline X each year, *knowing* 75 of them have no real chance of ever gaining full-time employment in the field to which they will be dedicating up to ten years of their lives studying" (emphasis is theirs).

Obfuscation in tenure, promotion, and salary review systems

Though tenure is supposed to guarantee freedom of expression, the reality is far different (Williams & Ceci, 2007). In the current system, junior professors live in fear and are often forced to choose between factions or take their chances when their departments are divided. The notion of a decentralised university with independent departments whose members alone are solely qualified to evaluate their work also means that the entities of review and accountability, namely upper levels of administration, are usually reluctant to weigh in, unless there is an egregious case of clear sexual or racial discrimination.

The overall lack of accountability in the academic process has led to rising public opinion against tenure. Polls in 2006-7 indicated that 82 per cent believe it needs to be changed. The perception of lack of accountability after tenure is great and perhaps justified, given the general lack of systems for post-tenure review, or, lack of post-tenure review entirely in many cases. Indeed a recent survey of Political Science chairs revealed that a substantial proportion believe that tenure has "shielded incompetent academic staff members from dismissal" (Neal, 2008).

Assessment procedures for academic performance do not lend themselves to the basic principles of good research. The tenure process is a particular point of stress, related in good part to the "lack of comprehensive, clear, and rational guidelines and procedures (Gappa, Austin, & Trice, 2007, p. 75)." Such pressures are naturally higher for women given additional duties of childcare, and for minorities who may also feel excluded from the collegial club.

In research-oriented universities, an "international reputation" in research is often evoked as the threshold for promotion to full professor. It is a well-known fact that most academic staff believe that research counts for much more. Analysts note that the proportion of academic staff members who publish regularly rose from ½ to 2/3 from 1969 to 1998. Though there are still stark differences by institution type, liberal arts (4 year) universities increasingly emphasise publications and grants as well as excellent teaching throughout careers, as necessary for success (Schuster & Finkelstein, 2006, pp. 99-103 & 171). The general tendency at most comprehensive institutions is simply to count the number of publications, thus eschewing quality as a factor (Youn & Price, 2009, p. 218).

In terms of teaching assessments for tenure, there is a tendency to rely on broad quantitative surveys of students at the end of term, which is a useful tool but does not really get at what students learned and what they do with that knowledge. Evaluation of teaching tends to be based on being "not cataclysmically bad," or just acceptable.

Service is an even more ambiguous category, as it entails service at a variety of levels from the department to the university to the discipline and finally the general public. As the rewards for internal and general public service tend to be unrecognised in comparison to

those in the discipline, where club benefits prevail (and international reputations ensue), there is a tendency to over-burden junior academic staff members with the internal chores (Macfarlane, 2007, p. 267) and lower level teaching assignments (where teaching arguably matters the most), thus reducing the sense of collective buy-in or mentoring by senior colleagues as well as the possibilities of using their experience to improve department performance.

Under such circumstances, external reviewers are one possible objective check. However, in a world of dense specialisation, the reality is that most reviewers will be have some connections with the candidate or his/her mentor, reflecting the Gramscian world of academics, based on reputational stratification as well as merit, thus biasing against new approaches or unconnected peers. In fact, in academics the rich tend to get richer, in the sense that a scholar with the right credentials and connections has better opportunities at grants, more familiarity with how to publish, and better funding to conduct research and to simply buy the time to develop a career.

These problems are multiplied when we consider interdisciplinary scholarship (Hurtado & Sharkness, 2008), given the iron chains of disciplinary organisation of almost all North American universities, a system that militates against multi- or true inter-disciplinarity regardless of the amount of lip service paid to it. In Political Science surveys, 67 per cent of PhD-granting departments in North America require articles in prestigious (disciplinary) journals for tenure, while just 10 per cent of BA and 19 per cent of MA institutions do (Rothgeb & Burger, 2009), indicating that top notch interdisciplinary scholars are likely to end up in less prestigious universities, with less resources for scholarship. In short, while most universities now proclaim interdisciplinarity on the one hand, they give all power and decision-making to individual departments on the other, revealing the severe blockages to cooperation across silos.

In fact, the lack of an open system and the extreme path dependency of academics are well-documented in studies (Youn, 1988, p. 17). The system separates candidates into 2 discrete markets, one for teaching-intensive careers, and, for the elite, research careers (Youn & Zelterman, 1988, p. 69). Reputation of institution, and networking with like-minded colleagues are major factors that determine who gets hired, published, and funded, and militate against new paradigms. Lest we be accused of hyperbole, we note the findings of a recent study of political science PhD hiring and placement, which concludes, “Comparing Figures 1 and 3 further reveals how remarkably little change has occurred in the centrality of the very top departments in the network over time, although some departments have become (marginally) more central and others (marginally) more peripheral, with only a few departments exhibiting substantial shift in relative location,” (Fowler, Grofman & Masuoka, 2007, p. 736), with location referring to the originating departments of hirees.

Peer review is also used for promotion and salary review. The peer review system for publication and promotion is also fraught with well-known problems. Absent clear guidelines, reviews of work can vary quite wildly; an author can feel that editors can cherry pick or over-determine end results based on their choice of reviewers. This leads to a natural clustering of scholars within camps for mutual protection, further entrenching divisions. A would be author frequently does not get clear feedback on what is wrong with the way he/she has approached the problem or examined it, preventing learning from taking place.

Moreover, the degree of specialisation means that comments from one journal do not readily help an author in revamping an article. The overwhelming nature of being an editor with the proliferation of production based on the job market leaves little time for helping a junior scholar to fix the problems with their research.

Moving forward on tenure, promotion and salary review assessment systems

How do we move towards a more transparent and accountable system for tenure? There appear to be no quick fixes. In the UK and Australia, a national academic evaluation system based primarily on number and place of publications has evoked considerable controversy. For example, highly ranked departments were closed down for lack of enrolments, bringing into question whether the system of quality review had any consequences (Bruneau & Savage, 2002, pp. 106 & 117).

General guides to review recommend a set of clear guidelines be given to junior academic staff, and that performance reviews result in a career development plan that includes clear teaching, research, and service expectations and the steps that will be taken to achieve them for the next few (3-5) years. These should include the classes that will be taught and elements such as teaching workshops that are designed to strengthen capacity in each of the key areas. Academic staff who are performing at a higher level should be rewarded with larger salary increases as well as tenure and promotion. Informal rewards should also be used, including greater access to teaching and research assistants, differential assignments to areas of research/teaching strength, and travel and lab funds. However, success should also bring responsibilities for chairing important committees and mentoring junior academic staff and graduate students (Curry, 2006). More explicit guidelines could help to reduce political games, such as statements like "we expect 4 peer-reviewed articles or 1 book in well-recognised journals or presses for tenure or promotion."

Teaching and service: the neglected cornerstones

As academics, our most profound influence on the world will be in training and shaping the next generation of students. Yet, Linda Nilson (2009) likens the challenge of teaching well under current circumstances as being asked to perform "magic." With increasing classroom size, ever more heterogeneous student populations, including large numbers that are under-prepared, rising research demands, and higher level training needs and competition needed for employment on the one hand, but shrinking resources on the other, teaching excellence is an increasingly difficult summit to climb. The most precarious aspect of that climb is the fact that we do not have any reliable system for assessing and rewarding teaching.

Therefore, a particularly worrisome outcome of the above trends and structures is that as research expectations increase, there will be less time and attention for teaching well. There is often a general perception that those who emphasise teaching are unable to conduct research on the same level. The fact that research is visible and brings in outside money further reinforces the notion that competence in teaching is all that is required (Serow, 2000). Yet, there is no evidence that excellent researchers can not become excellent teachers (Hoffmann & Oreopoulos, 2009, p. 92). In fact, the quality of the undergraduate education and the capacity of graduate students to teach should directly affect the reputation of the

institution, the availability of researchers, and the intellectual vitality of the instructors, and thereby the ability to succeed in research.

Most academic units give teaching equal weight in their formal systems of review. Indeed, most Political Science departments in North America formally rely upon a combination of student evaluations, a teaching portfolio, a review of syllabi and peer reviews (Rothgeb & Burger, 2009). However, the reality is different. For example, in tenure and promotion decisions, teaching often does not figure into the external reviews. Peer review is not always carried out. There is almost no reward for attending teaching workshops (Cooper & Nojima, 2002, p. 167).

Particularly problematic is the over-reliance on quantitative student surveys. Most universities simply look at the average for overall satisfaction, often ignoring other pertinent questions, such as amount of learning, level of workload, difficulty of course, and expected grade. More importantly, studies show distributions of grades have weak if any correlation with excellent teaching, and there is no chance to examine actual feedback given, or improvement in papers (Cooper & Nojima, 2002, p. 167). In fact, researchers have found a wide variety of variables directly affect evaluations, such as time of day, size of class, and level of class, that are never taken into account. As Baldwin & Blattner state (2003, p. 29), "Most of us would not dream of giving our students a final exam as the only measure of their performance in the class." As Sadler (1999, p. 115) notes, moreover, students are "uncalibrated instruments." Such evaluations are likely to lead to selection bias in that the most unhappy people are the most likely to fill out surveys. It is no wonder that student surveys show little faith among university staff (Douglas & Douglas, 2006).

Quantitative student evaluations do not really get at the range of skills taught in a course. These might include "complex and transformative learning" that goes beyond facts towards developing thoughtful analysis, interpretation, and use (Wright, no date). Course-based evaluations also do not get at a department's overall vision and goals for teaching (if it exists at all!). As Van Note Chism (2007, p. 17) notes,

> Specifically, experts indicate that while students are the most appropriate judges of day-to-day teacher behaviour and attitudes in the classroom, they are not the most appropriate judges of the accuracy of course content, use of acceptable teaching strategies in the discipline and the like. For these kinds of judgments, peers are the most appropriate source of information… (including giving feedback on): subject matter expertise, course goals, instructional materials and methods, assessment and grading practices, student achievement, professional and ethical behaviour, and thesis supervision.

In short, we have not taken seriously the basic methodological principles or long-term outcomes of research design for evaluating teaching performance. A recent in depth focus group study of teaching evaluations revealed that most student evaluations responded most directly to how the student "felt" about the course in terms of enjoyment level; whether the instructor was "likeable" and "not boring;" whether the professor agreed with the student's perspective on the subject matter; and whether a grade was given commensurate to their own perception of their effort. The author notes the fact that no follow-up studies of the methodological effectiveness of teaching surveys are apparent, that survey scores are in fact used as the sole measure of teaching effectiveness, and that they are treated as ordinal (sliding scale) rather than the categorical variables that they are (as averages are compared) (Titus,

2008). Basic statistical rules of validity, reliability, and tests of range and variance are largely ignored in such systems, though outcomes are considered unimpeachable in review systems.

The end result is often the creation of a student as client/consumer culture in academics. While certainly students and parents need to have strong input into teaching, there are 2 obvious problems unless there are counter-balances. The first is the obvious tendency to grade inflation, as popular professors are rewarded for their ability to entertain, rather than to teach. The second is that students, many younger than 21, by definition are inexperienced and not experts in education, life, or the subject that they study, and thus under-qualified to be the sole judges of how to teach.

Peter Kennedy, a renowned international professor of Economics and winner of teaching awards at SFU, summarised the core principles of good teaching at university as integrity, fairness, perseverance and courage. By these he means: the intellectual integrity and quality of the course; the integrity of the professor in fair evaluation of students and conveyance of the core material of the course; perseverance in continual improvement in knowledge and skills in both the area of studying and teaching itself; and courage, in risking disapproval by pushing students hard, taking an active public interest in teaching, and being willing to innovate in teaching. As Kennedy pointed out, education may likely take several years of lag time before its full effects becomes appreciated. Many of us have noted the value of the "hard" professors who pushed us, but only well after the fact.

Graduate education further perpetuates the cycle by centering around the production of top quality research, but with little to no attention in most cases paid to preparing academic staff members for teaching or service (Mathews-Gardner et al., 2008). Comprehensive studies of graduate education reveal a dismaying lack of preparation for academic life, with little systemisation to graduate training in terms of expectations or guidance (Wulff, Austin, Nyquist, & Sprague, 2004).

Research on teaching rewards in Canadian political science

To find out if the clues from our personal experience and the literature were accurate about assessment, we conducted a survey of teaching assessment systems in Political Science Departments across Canada. We note that our focus is on regular, full-time traditional departments in Canada, rather than distance and other education providers. We initially began our research with an on-line survey of practices with Chairs, but found the need to supplement this with in person and telephone interviews due to poor response. We found that the issues we discuss in regard to teaching incentives are ones that most Canadian Political Science departments are struggling with.

During 2009, we sent out 55 surveys, receiving back 13 responses, for a response rate of 24 per cent. We then conducted seven telephone and in person interviews (using the same survey) to bring the total to 21, for a final response rate of 38 per cent.

Table 1: Profile of survey respondents.

	Average	Range
Number of FT Faculty	14.6	3-35
Course Load	4.6	4 to 8
% buy-outs/year	27%	0-80%
Number FT Lecturers	1.4	0-5
Number Sessionals	10.3	0-49
Has a Grad Program	71% yes	
Has a PhD program	38% yes	

The responses reflect the diverse population of Canadian higher education, including responses from small, medium-sized (comprehensive) and large research universities. We note also that responses came from all over Canada, including institutions in British Columbia, Alberta, Saskatchewan, Manitoba, Ontario, Québec, New Brunswick, Newfoundland and Nova Scotia.

Several interesting observations can be made. The first is that while formally salary review procedures put teaching on an even keel with research, *almost no one thinks teaching is given equal weight in reality.*

Table 2: How is teaching rewarded in salary review?

Salary Review- formal weights	Average	Range	18 responses in total, 3 n/a
% research	46%	0-70	
% teaching	37.6%	15-100	
% service	21.4%	0-33	
Does formal weighting correspond to real?	72% say no		

We then asked a series of questions related to how teaching was evaluated. The responses are not that surprising. There is an almost exclusive reliance on student questionnaires. There is emphasis on adequate teaching at the tenure stage, but not on a periodic basis. There are more long-term studies of teaching according to our surveys than we would expect, so this response deserves follow-up study. Scrutiny of teaching appears to happen just once in many Departments, at the time of tenure.

Table 3: How is teaching generally evaluated?

How is Teaching Evaluated?			21 responses	
Student Questionnaires	21			
Peer Review	6			
Professional Staff Eval.	2			
Teaching dossier	2			
Is Teaching Evaluated Differently in Tenure/promotion vs. Salary review?				
	Yes	No		
	48%	19%	21 responses in total; 7 n/a, 10 yes, 4 no*	
Is there long-term evaluation of teaching?, e.g. alumni surveys				
	Yes	No		
	29%	71%	21 responses	

* Note: where peer assessment is used, it is only for tenure

One particularly sticky question is how to handle workload equity. We found that most respondents use a regular rotation to take care of this problem, and about 40 per cent consider workload during tenure and promotion.

Table 4: Teaching Equity

Teaching Equity- Large Course Rewards		20 responses
Rotation (taking turns)	9	
considered in tenure/promotion	6	
point system	3	
course release	1	
Other	7	

What efforts are political science departments in Canada making to improve teaching? Table 5 demonstrates that the main (and generally sole) avenue is through university awards. There is quite limited payback for attending workshops, despite universities putting considerable resources into such efforts. There is some effort to push PhDs to attend teaching workshops, but the results suggest that these are university-wide efforts, rather than political-science specific ones. New hires get course releases often, but no real training program for teaching political science.

Table 5: Teaching Improvement

Rewards for Attending Teaching Workshops?			
	Yes	No	
	10%	90%	21 responses in total
Teaching Awards?			
-University-level	19	2	
-Faculty-level	10	11	
-Department-level	1	20	
Aid to Junior Scholars' Teaching			
	Yes	No	21 responses in total
Course Release in 1st Year?	17	4	
Research Stipend?	18	2	20 responses in total, 1 n/a
	Yes	No	
PhDs Required to attend Teaching Workshops?	66%	33%	6 responses in total, 15 n/a
PhDs matched w/faculty mentor?	37%	63%	8 responses in total, 13 n/a

New hires arrive at universities subject to a maelstrom of possible political agendas. They often suffer from a lack of clear guidelines about what constitutes merit and the grounds for advancement in our profession. One interviewee suggests the adoption of a system of points given in ranges for different types of publications. Such a system does not deal with differences in quality of outlets (other than peer-reviewed), but it does at least give a clear rubric about how publications will be counted (e.g. articles vs. book reviews). The message new scholars get about teaching is clear- one interviewee put it bluntly, "you can't substitute teaching for scholarly activity and remain a credible university professor." Even traditional liberal arts programs have transformed themselves, partly through top down pressure, into research pressure cookers. One interviewee at such a place observed,

> "…this is an institution that is going through a metamorphosis. I think that traditionally this was an institution that really valued teaching and that was a really important criteria in terms of people's career advancement. There was a real emphasis on maintaining small class sizes, teaching was important in this institution but it has become rapidly less important because all of the criteria for promotion have to relate to research. Problems in this academic staff include everyone trying to avoid teaching because they need to perform in other areas like career development. The problem with this is, of course, someone has to teach the students! There seems to be very little incentive to take teaching seriously at the moment."

Moving towards a more reliable teaching and service assessment

There has to be some way to give appropriate signals (i.e. incentives) for assessing and improving performance. Obviously, the current system does not, leading to continual internal

political battles over the “burden” of teaching and service. As one interviewee stated, “we used to have "merit." Lousy teaching and poor service may have prevented the odd person from getting merit, but very few people ever got merit for teaching or service. It has now been abolished in favour of a series of awards.”

The evaluation of teaching also needs considerable renovation. Evaluation should be based on course materials, delivery, and outcomes (Fink, 2002, p. 50). Departments should state the core topics, skill sets, opportunities for real work experience, and a sequential set of courses, including possibly a capstone course, that allows students to build up and reflect upon the learning process (McClellan, 2009, p. 48).

Generally, there are 2 approaches to achieving such goals on the departmental level. One is to set up a clear mission, evaluative criteria, assessment procedures and course planning from the top (university and then chair) down. The second one is for more decentralised situations or where there is some academic staff disagreement, and allows the academic staff themselves to get together and decide upon such matters. In either case, objective assessment, peer review and incentive systems are vital to success (Deardorff & Folger, 2009). Once teaching matters, the process should be seen by academic staff as a method of collectively improving their performance, recognising the diverse talents of colleagues, and improving student outcomes. Assessment at the departmental or university level could include studies of alumni, current students, academic staff, and employers. The wide variety of methods we use for our own research could be used here: from interviews to focus groups to surveys. For core materials, comparing pre- and post- exams and placements could help to reveal collective success. Student evaluations are best done longitudinally (including after graduation), allowing time for the student to better put the class material into perspective and use (Kane, Williams, & Cappucccini-Ansfield, 2008). These could include exit surveys of graduating students and alumni asking about the performance of professors, techniques, and knowledge and skill bases.

It is extremely odd that we accept peer review for our research but do not embrace it as a tool for teaching. Hesitation to embrace it comes from our points about the politics of salary review. Supposedly it can only be done by someone with subdisciplinary knowledge (a point with which we disagree), therefore it supposedly opens up a potential lack of specialisation, conflict of interest and personal politics if a fellow member of the department carries one out. Even if it’s voluntary and not counted, there are likely to be strong suspicions by some members of the department that such information could be used against them in salary review or promotion. The obvious solution would be to have an outside person, perhaps someone from a neighbouring university, do the review. This brings into play a resources and coordination problem, thus, such a system could work only occasionally, perhaps every few years. For young scholars who need intensive guidance, pairing with a mentor would make the most sense. Rather than being latent resources that are underutilised such as poorly attended teaching workshops, teaching, learning skills, and library support staff at the university level could be an integral part of improving instruction and spreading knowledge across universities, if they were integrated into the assessment of attempts by faculty members to improve teaching outcomes.

On a class level, a pre- and post-test would reveal improvement in student outcomes, rather than simply how students feel at the moment of angst (generally right before a final exam) about the experience. A recent article (Stark-Wroblewski, Ahlering & Brill, 2007, p. 410), conducted such an experiment in a psychology course. They found, reinforcing the conclusions of other literature, that there was quite a low correlation between improvement in test scores and student evaluation scores – just 0.18. Such tests could be a valuable instrument to help move the speculative nature of teaching results forward. Yet, even such arrangements require supplemental peer review observation, to pick up the more subtle aspects such as cohort and individual differences, learning to analyse, interact with peers, and make decisions in complex environments. Others note that such tests are inadequate for evaluating real learning as opposed to reproduction. They suggest other instruments, such as a student journal to be kept during the class with observations (Voparil, 2009, p. 27), as well as the peer and expert review mechanisms discussed above.

Macfarlane (2007, p. 270) offers some suggestions about service. He suggests that a service portfolio can be created, with attention to criteria, such as engaging audiences; mentoring; organising and representing; reviewing; and advising. More fundamentally, we need to come out of our ivory tower once in a while to reveal the contribution we are making and interact with the public so that our work is more recognised, valued, and responsive.

Conclusion

We have seen that while research, particularly in the liberal arts, has continued to improve in terms of number of publications and accumulation of knowledge within specialisations, teaching and service have been neglected. We provided evidence that the root of the problem lies in faulty assessment systems in North America. Teaching (and service), by contrast, is more inherently collective than often acknowledged (Fairweather, 1996, p. 205), thus poor or stagnant performance inherently more difficult to address.

To begin to respond, we would need to re-orient academic staff members towards a scholarship of teaching and learning as well as the particular niche of their research. Absent greater attention to outcomes and balance of attention among the 3 missions, the scenario for future academics is bleak. Greater separation between researchers and teachers, and tenured and non-tenured academic staff will result. Researchers will be concentrated among those who can pull in external grants, with strong club effects. However, if they are tenure track they will also be over-burdened with internal service, administrative, and mentorship duties as their proportional number shrinks due to waning support for tenured positions. Those who are excellent teachers but poorer (but competent) researchers will find decreasing recognition, if they are able to get onto the tenure track at all.

Strong research efforts certainly have benefits, but it is hard to see how students or society will be better served by neglect of teaching and service, or for that matter, of explaining the value of the taxpayer-funded research to the public. While there should be some exclusive places for top notch researchers and teachers to specialise in their area of rare proficiency, the majority of the academy needs to be able to do both, particularly at the level of graduate training where exposure to research is a key ingredient. The way forward begins with an embrace of good governance and research principles as outlined throughout this essay.

Acknowledgements

We would like to thank Simon Fraser University's Institute for the Study of Teaching and Learning in the Disciplines for their generous support of this research, and Adrienne Burke, Cheryl Amundsen, and the late great Peter Kennedy of SFU for their very helpful comments.

References

Baldwin, T., & Blattner, N. (2003). Guarding against potential bias in student evaluations: what every faculty member needs to know. *College Teaching. 51*,1(Winter), 27-32.

Bruneau, W., & Savage, D.C. (2002). *Counting out the scholars: how performance indicators undermine universities and colleges.* Toronto: James Lorimer & Co.

Cooper, J. E., & Nojima, S.E. (2002). Teaching: academic whitewater rafting. In J. Cooper & D. Stevens (Eds.), *Tenure in the sacred grove: issues and strategies for women and minority faculty* (pp. 163-78). Albany: State University of New York Press.

Curry, T. H. (2006). Faculty performance reviews. *Effective Practices for Academic Leaders. 1*,2(Feb), 1-16.

Deardorff, M. D. & Folger, P.J. (2009). Making assessment matter: structuring assessment, transforming departments. In M. D. Deardorff, K. Hamann, & J. Ishiyama, (Eds.), *Assessment in Political Science* (pp.77-96). Washington: American Political Science Association.

Douglas, J. & Douglas, A. (2006). Evaluating teaching quality. *Quality in Higher Education. 12*(1), 3-13.

Fairweather, J. S. (1996). *Faculty work and trust: restoring the value of teaching and public service in American academic life.* Toronto: Allyn and Bacon.

Fink, L. D. (2002). Improving the evaluation of college teaching. In K. H. Gillespie, L. R. Hilsen & E. C. Wadsworth, (Eds.), *A guide to faculty development: practical advice, examples, and resources* (pp.46-58). San Francisco: Jossey-Bass.

Fowler, J. H., Grofman, B., & Masuoka, N. (2007). Social networks in political science: hiring and placement of Ph.D.s, 1960-2002. *PS: Political Science and Politics. 40*(4), 729-39.

Harvey Editor, L. (2008). Rankings of Higher education institutions: a critical review. *Quality in Higher Education. 14*(3), 187-207.

Hoffmann, F. & Oreopoulos, P. (2009). Professor qualities and student achievement. *The Review of Economics and Statistics. 91*(1), 83-92.

Hurtado, S. & Sharkness, J. (2008). Scholarship is changing, and so must tenure review. *Academe 94*(5), (no pages).

Kane, D., Williams, J. & Cappuccini-Ansfield, G. (2008). Student satisfaction surveys: the value in taking an historical perspective. *Quality in Higher Education. 14*(2), 135-55.

Louis, D. (2009). Adjuncts: solutions for a mistreated majority. *Chronicle of Higher Education.* 55(June 12), no pages.

Macfarlane, B. (2007). Defining and rewarding academic citizenship: the implications for university promotions policy. *Journal of Higher Education Policy and Management. 29*,3(Nov), 261-73.

Mann, D., & Hochenedel, H.N. (2006). A manifesto of the twenty-first century academic proletariat in North America. In R. Barrow & P. Keeney (Eds.), *Academic ethics* (pp.291-306). Burlington, VT: Ashgate.

Mathews-Gardner, L., Deardorff, M.D., Reeher, G., Hudson,W., Borrelli,M., & Halva-Neubauer,G. (2008). Getting a job at a teaching institution- and then succeeding: a Q&A with experienced teacher-scholars. *PS: Political Science and Politics. 41*(3), 575-8.

McClellan, E. F. (2009). An overview of the assessment movement. In M.D. Deardorff, K. Hamann, & J. Ishiyama, (Eds). *Assessment in Political Science* (pp.39-60). Washington: American Political Science Association.

Neal, A. D. (2008). Reviewing the post-tenure review. *Academe. 94*,5 (Sept/Oct): (no pages).

Nilson, L. B. (2009). Editor's Introduction: the educational developer as magician. In L.B. Nilson & J. E. Miller (Eds.), *To improve the academy: resources for faculty, instructional, and organisational development* (pp.3-13), vol. 27. Somerset, NJ: Jossey-Bass.

Rothgeb, J. M. & Burger,B. (2009). Tenure standards in political science departments: results from a survey of department chairs. *PS: Political Science and Politics*. *42*, 513-19.

Sadler, D. R. (1999). *Managing your academic career: strategies for success*. St. Leonards, Australia: Allen & Unwin.

Schuster, J. H. & Finkelstein,M.J. (2006). *The American faculty: the restructuring of academic work and careers.* Baltimore: The Johns Hopkins University Press.

Serow, R.C. (2000). Research and teaching at a research university. *Higher Education. 40*, 449-463.

Stark-Wroblewski, K., Ahlering, R.F., & Brill, F.M. (2007). Toward a more comprehensive approach to evaluating teaching effectiveness: supplementing student evaluations of teaching with pre-post learning measures. *Assessment & Evaluation in Higher Education. 32*,4 (Aug), 403-15.

Titus, J. J. (2008). Student ratings in a consumerist academy: leveraging pedagogical control and authority. *Sociological Perspectives*. *51*(2), 397-422.

Van Note Chism, N. (2007). *Peer Review of Teaching: A Sourcebook.* 2nd ed. Bolton, MA: Anker Publishing.

Voparil, C. J. (2009). Assessing for understanding: toward a theory of assessment as learning. In M.D. Deardorff, K. Hamann, & J. Ishiyama (Eds.), *Assessment in Political Science* (pp. 17-38). Washington: American Political Science Association.

Wang, X. (2010). *Performance analysis for public and nonprofit organisations.* Toronto: Jones and Bartlett.

Williams, W. M. & Ceci, S.J. (2007). Does tenure really work? *Chronicle of Higher Education*. *53*(27), B16.

Wright, B. D. (no date). Guidelines for good assessment of student learning at the department or program level. Working paper, from http://apsanet.org/content_13026.cfm.

Wulff, D.H., Austin, A.E., Nyquist, J.D., & Sprague, J. (2004). The development of graduate students as teaching scholars: a four-year longitudinal study. In A.Wulff & Associates (Eds.), *Paths to the professoriate: strategies for enriching the preparation of future faculty* (pp. 46-73). San Francisco: Jossey-Bass.

Youn, T.I.K. (1988). Studies of academic markets and careers: an historical review. In D.W. Breneman & T.I.K. Youn (Eds.), *Academic labor markets and careers* (pp. 204-37). Philadelphia: The Falmer Press.

Youn, T. I. K. & Price, T.M. (2009). Learning from the experience of others: the evolution of faculty tenure and promotion rules in comprehensive institutions. *The Journal of Higher Education. 80*,2 (March/April), 204-37.

Youn, T. I. K. & Zeleterman, D. (1988). Institutional career mobility in academia. In D.W. Breneman & T. I.K. Youn (Eds.), *Academic labor markets and careers* (pp. 52-73). Philadelphia: The Falmer Press.