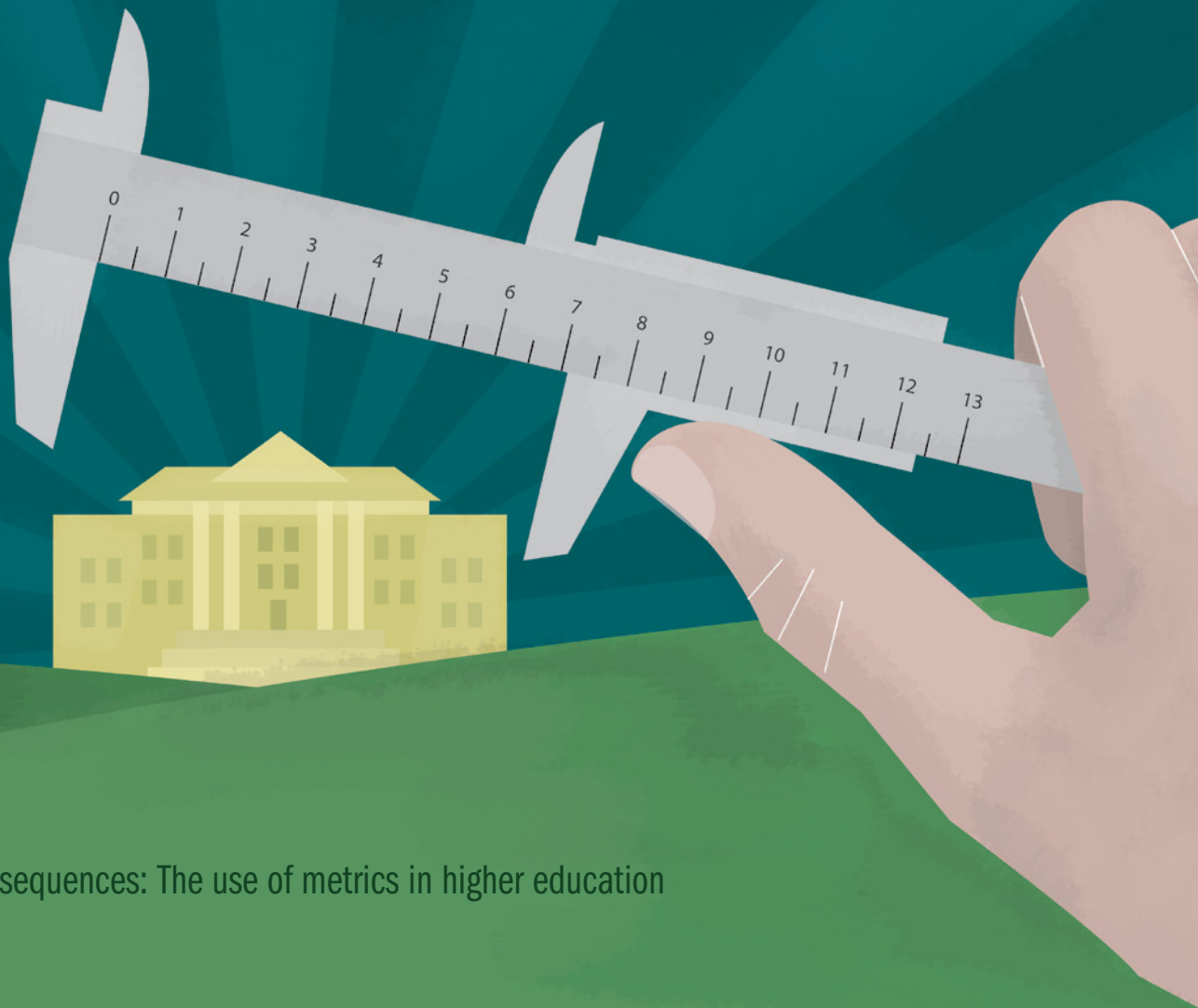


Academic Matters

OCUFA'S JOURNAL OF HIGHER EDUCATION
LA REVUE D'ENSEIGNEMENT SUPÉRIEUR D'OCUFA

Measuring Up?

Fall 2017



Gavin Moodie

Unintended consequences: The use of metrics in higher education

Yves Gingras

Dérives et effets pervers de l'évaluation quantitative de la recherche

Claire Polster and Sarah Amsler

Waking up to the reality of Canadian higher education



Athabasca University
FACULTY OF BUSINESS

For more information on our DBA:
1-800-561-4650
business.athabascau.ca/dba

“The DBA has continued to open doors for me, both in the advanced tools and learning materials as well as the implementations I can bring to my organization.”

Derek Prue, DBA Student
President
Windermere Properties

2 Editorial Matters

3 Unintended consequences: The use of metrics in higher education

Gavin Moodie

Metrics are used throughout Ontario's postsecondary education system—for determining university funding, judging institutional performance, and gauging student perceptions. But metrics are not always the best tool for evaluation, and often have unintended consequences.

9 Is there a metric to evaluate tenure?

Tim Sibbald and Victoria Handford

How much can data meaningfully inform decisions about tenure? If data only tell part of the story, perhaps faculty should be evaluated so that their different lived experiences are also taken into consideration.

14 The abuses and perverse effects of quantitative evaluation in the academy

Yves Gingras

The world of academic research is scored according to so-called "objective" measures, with an emphasis on publications and citations. But the very foundations of this approach are flawed. Is it time to abandon these simplistic ranking schemes?

20 Dérives et effets pervers de l'évaluation quantitative de la recherche

Yves Gingras

Les professeurs et les chercheurs universitaires sont de plus en plus évalués à l'aide de mesures dites « objectives », qui mettent l'accent sur les publications et les citations. Mais le fondement même de cette approche est problématique. Le temps est-il venu d'abandonner ces méthodes de notation simplistes?

25 Collecting data from students with students

Ruth Childs

Gathering data on university students can provide important information about how they interact with the postsecondary education system, but it is also important to consult students to determine what data are collected and how.

28 Waking up to the reality of Canadian higher education

Claire Polster and Sarah Amsler

Higher-education systems in Canada and the United Kingdom share much in common, but there are important differences that faculty on both sides of the Atlantic should appreciate. The UK experience can wake Canadian academics up to the urgency of resisting university corporatization and to the opportunities for resistance that remain.

33 Understanding the United Kingdom's Teaching Excellence Framework and its implications

Rob Copeland

The UK's new metrics-based teaching evaluation framework is methodologically and politically flawed. What will this mean for the country's universities and faculty?

MORE ON ACADEMICMATTERS.CA

OCUFA

Ontario Confederation of University Faculty Associations
Union des Associations des Professeurs des Universités de l'Ontario

Academic Matters

OCUFA'S JOURNAL OF HIGHER EDUCATION
LA REVUE D'ENSEIGNEMENT SUPÉRIEUR D'OCUFA

Academic Matters is published two times a year by OCUFA, and is received by 17,000 professors, academic librarians and others interested in higher education issues across Canada. The journal explores issues of relevance to higher education in Ontario, other provinces in Canada, and globally. It is intended to be a forum for thoughtful and thought-provoking, original and engaging discussion of current trends in postsecondary education and consideration of academe's future direction.

Readers are encouraged to contribute their views, ideas and talents. Letters to the editor (maximum 250 words) are welcome and may be edited for length. To provide an article or artwork for Academic Matters, please send your query to Editor-in-Chief Ben Lewis at editor@academicmatters.ca.

Reproduction of articles requires the permission of the author and publisher. Copyright in material accepted for publication remains with the author. The publisher may choose, however, to translate and/or reproduce material in print and electronic form. All published material expresses the view of the author(s) and not the publisher.

Academic Matters accepte volontiers des articles écrits en anglais ou en français.

Publisher:

OCUFA; Mark Rosenfeld, Executive Director

Editor-in-Chief:

Ben Lewis

Associate Editor:

Carol Anderson

Assistant Editor:

Brynne Sinclair-Waters

Assistant Editor:

Gabrielle Ross-Marquette

Art Direction:

Eva Kiss, Neglia Design Inc., www.NegliaDesign.com

Editorial Advisory Board:

Glen Jones, April Lindgren, Daniel Munro, Stephanie Ross

National Advertising Sales:

DOVETAIL COMMUNICATIONS
30 East Beaver Creek Road, Ste. 202
Richmond Hill, ON L4B 1J2

Sales Manager:

Kim Cristini, kcristini@dvetail.com
Phone (905) 707-3510

ISSN 1719-010X

For subscription information, please contact:

OCUFA@ocufa.on.ca

Return undeliverable Canadian addresses to:

Ontario Confederation of University Faculty Associations
17 Isabella Street, Toronto, ON M4Y 1M7

Phone (416) 979-2117 Fax (416) 593-5607
www.ocufa.on.ca www.academicmatters.ca

union
logo
here

Cover illustration: Marie-Claude Carignan



Alluring figures

There is an appealing simplicity in numbers. A number's value is never ambiguous, even if its meaning can be. Numbers are specific and easily compared to one another. They allow us to measure the dimensions of an object or to describe the outcomes of a decision.

There has long been a desire to use numbers to measure the impacts of specific changes on complex social systems. In recent decades, the dramatic increase in the power of computers has transformed the amount of time and energy these calculations require. Computers love numbers. Every decision they make is the result of a lot of very complex math. As a result, it has become possible to collect and run calculations on vast amount of data.

By defining a set of goals and metrics (standards by which to measure and evaluate data), entire industries are attempting to optimize delivery of services. From package delivery, to advertising, to electioneering, the assumption is that gathering and evaluating enough data will reveal ways to improve outcomes.

In Canada, and around the world, data analysis and metrics are increasingly being used in postsecondary education—to evaluate the teaching performance and research output of professors, to rank the performance of postsecondary institutions, to measure the diversity of student bodies, and to track the types of employment students attain after graduation. Performance on these metrics impact faculty tenure

and promotion, research pursuits, institutional funding, and enrolment.

However, making decisions based only on metrics overlooks those important aspects of postsecondary education that cannot be quantified. The very act of choosing the metrics used to evaluate success has serious consequences. Metrics reflect the values and priorities of those choosing them, diminishing the importance of the data not being measured.

Does a focus on metrics in higher education serve to optimize postsecondary education systems and make them more accountable? Do metrics distract from other important considerations, undermining the integrity of the system? Do metrics compound systemic biases within institutions or help reveal them so they can be addressed?

In this issue of *Academic Matters*, our contributors contemplate these questions and critically examine the impact that metrics have on the quality and integrity of teaching and research at universities in Canada, and around the world.

Gavin Moodie provides an overview of the different ways that metrics are used in postsecondary education and explores their unintended consequences, providing readers with a helpful primer on the issues.

Tim Sibbald and Victoria Handford examine how data inform decisions about tenure and argue for more holistic approaches to evaluation that take into account different lived experiences.

Yves Gingras dives into the world of academic research and questions the very foundations upon which published work is ranked and rewarded. Originally written in French, we have translated the article for our readers and published it in both languages.

Ruth Childs describes her experience designing student surveys, and the importance of consulting students about what data is being collected and how. She reminds us that it is vital to understand the perspectives of those that we are collecting data from if we are to ensure the data is useful.

Claire Polster and Sarah Amsler share their observations of higher education systems in Canada and the United Kingdom, and contrast the ways in which results-driven corporatization has impacted faculty.

Finally, Rob Copeland provides a background on the UK's new metrics-based Teaching Education Framework and contemplates what its flaws will mean for the future of the country's universities.

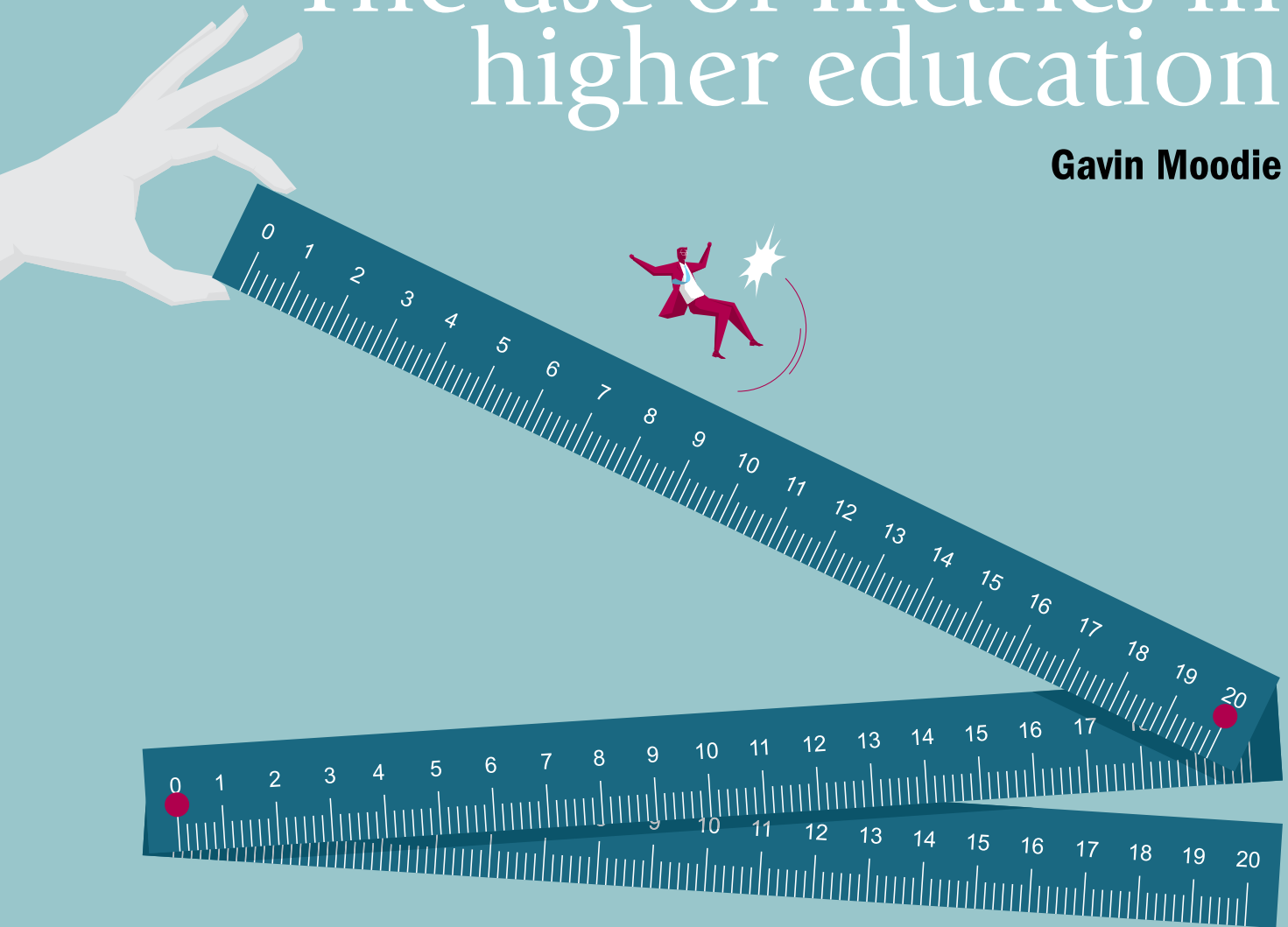
There is value in measuring educational inputs and outcomes in higher education, but it is crucial to be aware of their limitations. Metrics and quantitative data analysis can be valuable tools, but an over-dependence on numbers can have unintended consequences.

By defining results numerically, we invite the ranking of institutions, programs, and individuals; we encourage competition instead of collaboration; and, perhaps most importantly, an over-reliance on metrics devalues those vital aspects of the educational experience that numbers just cannot define. **AM**

Ben Lewis is the Editor-in-Chief of Academic Matters and Communications Lead for OCUFA.

UNINTENDED CONSEQUENCES: The use of metrics in higher education

Gavin Moodie



Metrics are used throughout Ontario's postsecondary education system—for determining university funding, judging institutional performance, and gauging student perceptions. But metrics are not always the best tool for evaluation, and often have unintended consequences.

Des mesures sont utilisées dans tout le système d'éducation postsecondaire de l'Ontario—afin de déterminer le financement des universités, de juger la performance des établissements, et de considérer les perceptions de la population étudiante. Mais les mesures ne sont pas toujours le meilleur outil pour l'évaluation, et elles ont souvent des conséquences imprévues.

MEASURED SCEPTICISM

Statistical measures, or “metrics” as we are now expected to call them, have become as extensive in higher education as they are deplored. The growth in the use of metrics has been neither recent nor restricted to Ontario. Faculty are therefore unlikely to be able to reverse metrics’ rise. But faculty could displace metrics from their core role of teaching and learning by promoting peer review of teaching, which is a far more valid indicator of teaching quality, may support teaching and learning as a community endeavour, and would remain very much the responsibility of individual faculty, rather than the domain of central data collectors and analysts.

AMBIVALENCE ABOUT METRICS

In an article published in 2000, English academic Malcolm Tight amusingly but informatively compared the ranks of English soccer clubs and universities. His work confirmed that there was a close relation between the distribution of universities and soccer clubs and the population of English cities and larger towns. Tight also found that, in many cities and towns, local universities shared similar ranks to local soccer clubs (if a university was ranked in the top ten, so was the soccer club). However, universities in the South of England were more likely to rank much higher than local soccer clubs, while universities in the North and Midlands were more likely to rank much lower.

Both soccer clubs and universities gain a considerable advantage from being old and well-established, and gain a further advantage when they have a higher income than their competitors (whether through endowments, tuition fees, ticket prices, or merchandise), something which is also strongly related to how long the club or university has been operating. University ranks are also similar to English soccer team ranks in that they are dominated by a stable elite that changes little over time.

Tight’s comparison of ranks illustrates an ambivalence with the usefulness of ranks and, more generally, with metrics, statistical measures, and performance indicators. On the one hand, these ranks seem to democratize judgments and decision-making about specialized activities. Those who know little about English soccer can readily determine the most successful clubs by scanning the league ranks. On the other hand, some highly ranked clubs may play too defensively and thus may not be considered by aficionados to play the “best” soccer. Ranking soccer clubs only by their winning ratio ignores more sophisticated judgements about the quality of the football they play.

Metrics are tools for transferring evaluation and monitoring from experts to people who are distant in location and seniority.

GOVERNMENT FUNDING AND METRICS

The Ontario Ministry of Advanced Education and Skills Development (MAESD) and its predecessors have long allocated funds to colleges and universities predominantly according to their level of enrolment. However, over the last decade MAESD has relied increasingly on performance indicators to monitor postsecondary institutions and influence their internal decisions. MAESD has been reporting each college’s rates for student satisfaction, graduation, graduate satisfaction, graduate employment, and employer satisfaction. For each university, the Council of Ontario Universities reports data on applications, student financial assistance, enrolments, funding, faculty, degrees awarded, and graduates’ employment outcomes.

Ontario universities get most of their operating revenue from tuition fees (38%), MAESD (27%), the federal government (11%), other Ontario ministries (4%), and other sources (20%).¹ Only four per cent of MAESD’s operating funding is allocated according to performance indicators, meaning that just over one per cent of Ontario university revenue is allocated in this way.² Yet performance funding and its indicators have been debated extensively.

Even more contentious is MAESD’s differentiation policy, which is informed by the Higher Education Quality Council of Ontario’s (HEQCO’s) analysis of metrics. The policy is primarily implemented through metrics-heavy strategic mandate agreements negotiated between the province and each university. Further, in a recent article for *Academic Matters*, the executive lead of Ontario’s University Funding Model Review, Sue Herbert, expressed a need for more “information, data, and metrics that are transparent, accessible, and validated.”³

It is therefore easy to conclude that MAESD’s direction for colleges and universities is driven by metrics that allow government officials and ministers to make judgements about institutions without a detailed familiarity with, or expertise in, postsecondary education. This is similar to arrangements in other Canadian provinces, a number of US states, the United Kingdom, and other countries, where governments and ministries have greatly increased their reliance on metrics.



There are three obvious alternatives to this scenario. The overwhelming preference of college and university management and staff is for governments to leave more decisions to the institutions alone. Funding would be provided to universities with few strings attached, tuition fees would be unregulated, and universities would be able to pursue their own visions for education, free of government interference. However, such a scenario undermines the democratic power of Ontario citizens, which is exercised through the provincial government and its delegates.

The second alternative would be for ministers and ministries to return to making decisions about postsecondary education by relying on their own judgement, attitudes, impressions, and others' anecdotes, as well as the advice of experts. This is opaque and relies on a high level of trust that decisions aren't affected by partisan interests or personal prejudices.

A third alternative would be for the government to delegate decisions to an intermediate or buffer body of experts in postsecondary education who would make decisions according to a combination of their own judgements, expertise, experience, and metrics. This was investigated by David Trick for HEQCO, who concluded that:

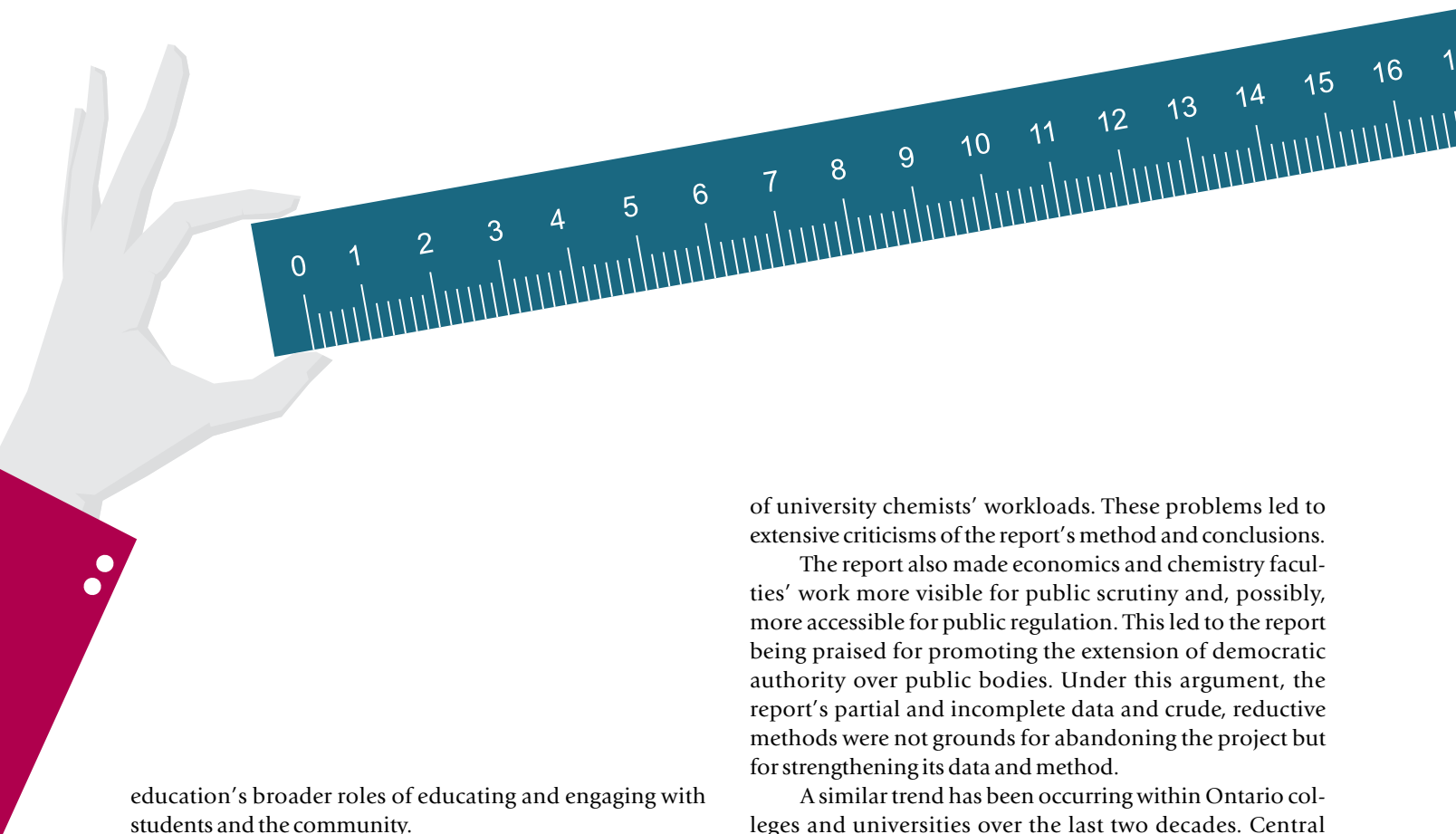
An intermediary body could be helpful as the Ontario government seeks to pursue quality and sustainability through its differentiation policy framework. Specifically, such a body could be useful for pursuing and eventually renewing the province's Strategic Mandate Agreements; for strategic allocation of funding (particularly research funds); making fair and evidence-based decisions on controversial allocation issues; and identifying/incentivizing opportunities for cooperation between institutions to maintain access and quality while reducing unnecessary duplication.⁴

However, governments and ministries are concerned that buffer bodies restrict their discretion and reflect the interests of the institutions they oversee more than the governments and public they are established to serve. In fact, the UK recently dismantled its higher-education buffer body, the Higher Education Funding Council for England.

INSTITUTIONAL ACTORS

Metrics are also tools for transferring evaluation and monitoring from experts, who are usually the people conducting the activity, to people and bodies who are distant in location and seniority, often senior management located centrally. No organization in Ontario or Canada has replicated the detail of the University of Texas' task force on productivity and excellence, which compiled data on each professor's pay, teaching load, enrolments, mean grade awarded, mean student evaluation score, amount of grants won, and time spent on teaching and research. The data on 13,000 faculty in nine institutions was published in a spreadsheet of 821 pages in response to open-records requests.

HEQCO's preliminary report on the productivity of the Ontario public postsecondary education system compared data for Ontario's college and university sector with those for all other provinces, examining enrolments, faculty/student ratios, funding per student, graduates, graduates per faculty, funding per graduate, tri-council funding per faculty, citations per faculty, and faculty workload. OCUFA criticized that report for being preoccupied with outputs at the expense of inputs such as public funding and processes such as student engagement, as well as for its narrow focus on labour market outcomes, which excluded postsecondary



education's broader roles of educating and engaging with students and the community.

In a subsequent report for HEQCO, Jonker and Hicks went further, analyzing data on individual faculty that were publicly posted on university websites and elsewhere. HEQCO wrote that the report:

conservatively estimates that approximately 19% of tenure and tenure-track economics and chemistry faculty members at 10 Ontario universities sampled demonstrated no obvious recent contribution of scholarly or research output, although universities generally adhere to a faculty workload distribution of 40% teaching, 40% research and 20% service.

Extrapolating from that sample, the authors say that Ontario's university system would be more productive and efficient if research non-active faculty members compensated for their lack of scholarly output by increasing their teaching load to double that of their research-active colleagues—for an 80% teaching and 20% service workload distribution.⁵

This report illuminates several issues with using metrics to measure productivity. Neither of the authors is a chemist, yet they felt competent, based on their use of metrics, to judge chemists' scholarly "output" and workload. Neither author works at a university with chemists, yet they believed it was appropriate for them to propose major reallocations

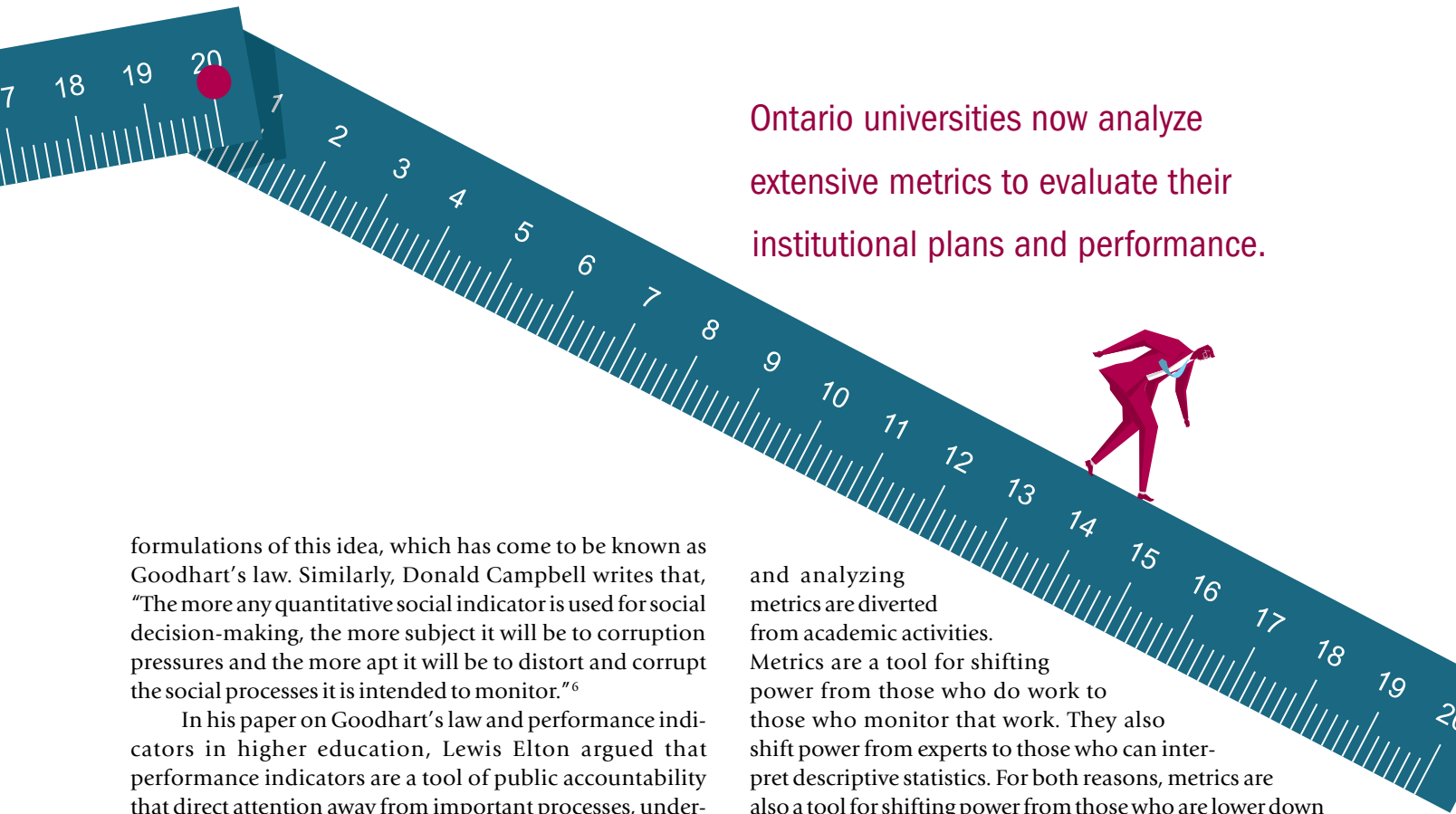
of university chemists' workloads. These problems led to extensive criticisms of the report's method and conclusions.

The report also made economics and chemistry faculties' work more visible for public scrutiny and, possibly, more accessible for public regulation. This led to the report being praised for promoting the extension of democratic authority over public bodies. Under this argument, the report's partial and incomplete data and crude, reductive methods were not grounds for abandoning the project but for strengthening its data and method.

A similar trend has been occurring within Ontario colleges and universities over the last two decades. Central administrations in Ontario's postsecondary institutions have long collected data to allocate funds internally and have increasingly collected and analyzed data to assess and monitor their institution's performance. Ontario universities now analyze extensive metrics to evaluate their institutional plans and performance. By a process of mimetic isomorphism—the tendency of an organization to imitate another organization's structure—institutions tend to allocate funds and evaluate performance internally according to the criteria on which their own funds are received and their performance evaluated. These measures are replicated, to varying extents, by faculties. While immediate supervisors and heads of departments still seem to share enough expertise and interests with faculty to trust in their own judgment and that of their faculty members, they still need to take account of the metrics used by senior administrators in their institution.

UNINTENDED CONSEQUENCES

A common criticism of the use of metrics is that they can have unintended and undesirable consequences by distorting the behaviour of those being measured. This idea was expressed rigorously by British economist Charles Goodhart, who wrote that an observed statistical regularity tends to collapse once it is used as a target. There are various



Ontario universities now analyze extensive metrics to evaluate their institutional plans and performance.

formulations of this idea, which has come to be known as Goodhart's law. Similarly, Donald Campbell writes that, "The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor."⁶

In his paper on Goodhart's law and performance indicators in higher education, Lewis Elton argued that performance indicators are a tool of public accountability that direct attention away from important processes, undermine academic professionalism, and reflect an erosion of trust in individual academics. However, he was not uncritically protective of academics, and argued that most traditional assessments of students use proxies that are similar to performance indicators (PIs). He argues that most grading is unreliable, suffering the methodological flaws of accountability through metrics:

Much of this traditional assessment is largely through the equivalent of PIs, with all the faults that stem from Goodhart's Law...

Also, it may be noted that what are normally called summative and formative assessment correspond to PIs used as a control judgmentally and PIs used as a management tool for improvement.

As long as academics use traditional examinations to assess students, they really have no right to complain if the Department of Education and Skills assesses them through quantitative PIs and targets (emphasis in original).⁷

IMPLICATIONS FOR FACULTY

Much of the data upon which metrics are based are collected from faculty, adding unproductive work to their many other duties. The resources invested in collecting, reporting,

and analyzing metrics are diverted from academic activities.

Metrics are a tool for shifting power from those who do work to those who monitor that work. They also shift power from experts to those who can interpret descriptive statistics. For both reasons, metrics are also a tool for shifting power from those who are lower down in an organization to those who are higher up. Metrics may change faculty priorities and increase the pressure to improve their performance on the measures monitored, as Jeanette Taylor found for some of the 152 academics she surveyed at four Australian universities. Metrics are likely to reduce faculty's discretion over the work they do and how it is evaluated. Metrics are also likely to intensify faculty work.

Metrics are limited and many have methodological flaws. Yet, rather than pausing the use of metrics, pointing out their problems leads to increased investment in attempts to make them more extensive and rigorous. This in turn increases demands on faculty to provide more and better data. Metrics are widespread in postsecondary education in many jurisdictions other than Ontario, and are pervasive in elementary school education. This suggests that faculty can do little more than moderate and perhaps redirect the metrics that flood over the sector. However, there is a major action that faculty can and should take that would redress much of the current distortion of metrics: promote widespread peer review of teaching.

There is currently no direct measure of the quality of teaching. This does not, of course, prevent believers in metrics from seeking to evaluate teaching by proxies such as student satisfaction and graduation rates. Compilers of ranks also incorporate faculty/student ratios and faculty reputation surveys. In contrast, all the measures of research performance are aggregations of peer evaluations: Manuscripts are published on the recommendations of peer

Peer review of teaching should become an important counterweight to the excessive reliance on research for evaluating the performance of institutions and faculty.

reviewers moderated by editors who are experts in the field, citations are by authors published in the field, and grants are awarded on the recommendations of experts moderated by chairs who are experts in the field.

Teams of scholars have developed comprehensive criteria, protocols, and processes that provide frameworks for the peer review of teaching. Typically, reviews are the responsibility of faculties, with the support of an expert in teaching and learning; reviewers are chosen by the faculty member from a team of accredited reviewers; the review is of the whole course, not just the observation of teaching events; and the faculty member meets their reviewers at least once before and after the review. In Canada, reviews are required for promotion and tenure at some Canadian universities, such as the University of British Columbia, as they are at several universities in the United States, the UK, and Australia.

Peer review of teaching should become an important counterweight to the excessive reliance on research for evaluating the per-

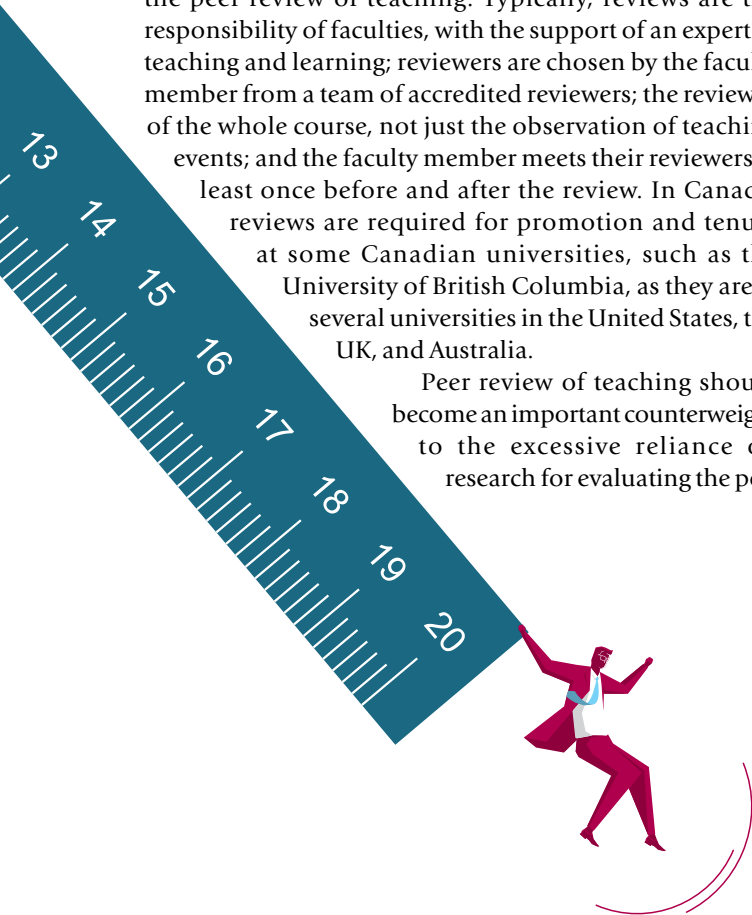
formance of institutions and faculty, as well as the excessive reliance on student satisfaction to evaluate faculty and institutions, and on graduation rates to evaluate institutions. Peer review of teaching enables teaching to become a community endeavour and, of course, remains very much the responsibility of individual faculty, rather than central data collectors and analysts.

MEASURED PROGRESS

Metrics have had a long and extensive history in higher education, despite the extensive critiques they have attracted and notwithstanding the clear dangers they pose. They are pervasive in Ontario, and probably more so in other jurisdictions in Canada, the US, the UK, and elsewhere. While faculty may curb the worst excesses of metrics, it seems unlikely that they will reverse metrics' advances. But there is a prospect of diverting the application of metrics from one of faculty's core activities and responsibilities, teaching and learning. Faculty can do this by promoting the peer review of teaching, which is a far more valid indicator of teaching quality than the proxy metrics that are currently used. ■■

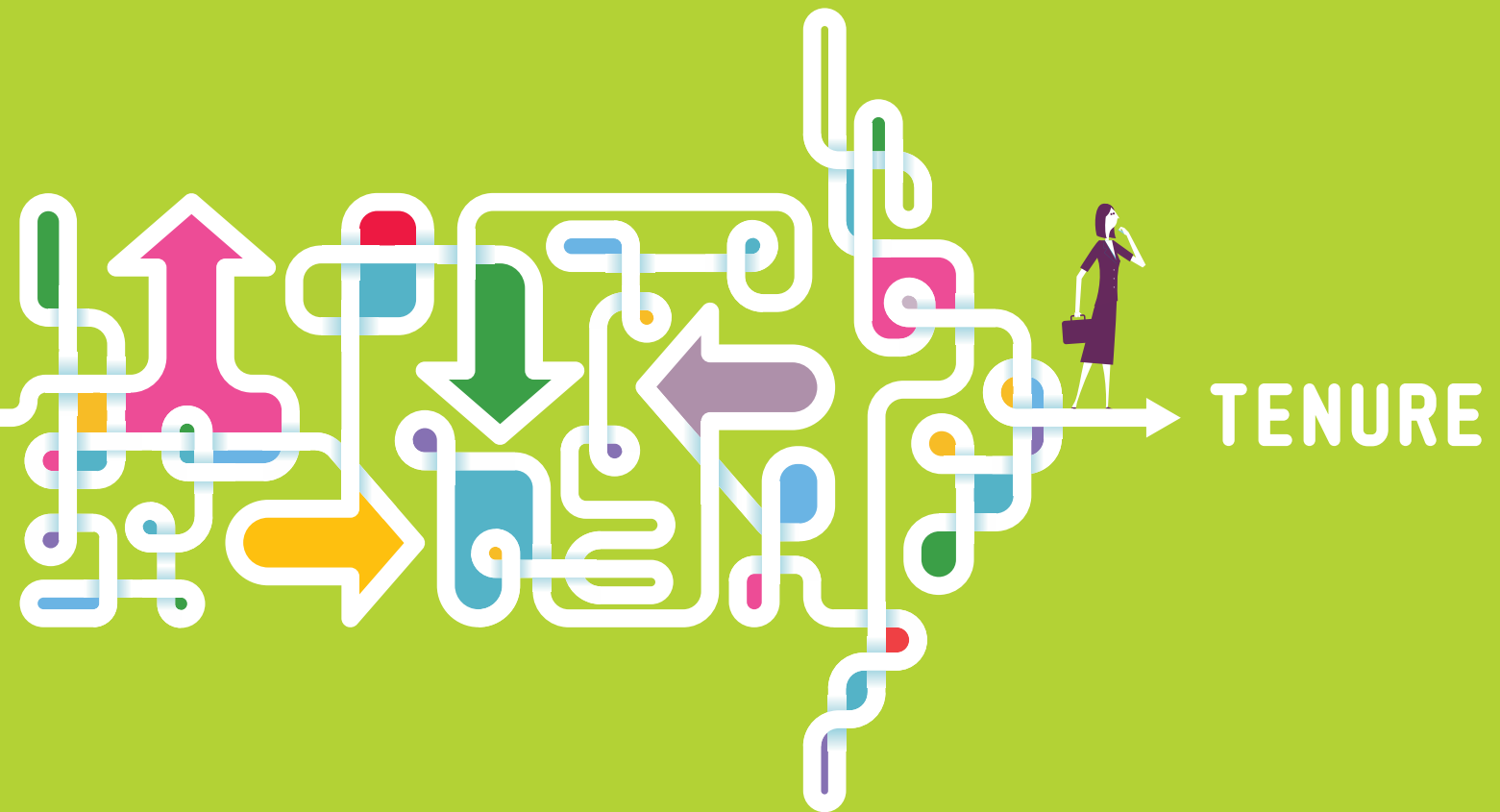
Gavin Moodie is Adjunct Professor in the Department of Leadership, Higher, and Adult Education at the Ontario Institute for Studies in Education at the University of Toronto.

1. HEQCO, "The Ontario university funding model in context," 2015, <http://www.heqco.ca/SiteCollectionDocuments/Contextual%20Background%20to%20the%20Ontario%20University%20Funding%20Formula-English.pdf>.
2. Ministry of Training, Colleges and Universities, "University funding model reform consultation paper," 2015, <http://news.ontario.ca/tcu/en/2015/03/ontario-launches-consultations-on-university-funding-reform.html>.
3. Sue Herbert, "Reviewing Ontario's university funding," *Academic Matters*, 2016 Jan., <https://academicmatters.ca/2016/01/reviewing-ontarios-university-funding>.
4. David Trick, "The role of intermediary funding bodies in enhancing quality and sustainability in higher education," HEQCO, 2015, <http://www.heqco.ca/en-ca/Research/Research%20Publications/Pages/Summary.aspx?link=196>.
5. Linda Jonker & Martin Hicks, "Teaching loads and research outputs of Ontario university faculty: implications for productivity and differentiation," HEQCO, 2014.
6. Donald T. Campbell, "Assessing the impact of planned social change," *Evaluation and Program Planning*, 2:1 (1979): 67-90.
7. Lewis Elton, "Goodhart's law and performance indicators in higher education," *Evaluation and Research in Education*, 18:1,2 (2004): 120-128.



Is there a metric to evaluate tenure?

Tim Sibbald and Victoria Handford



How much can data meaningfully inform decisions about tenure? If data only tell part of the story, perhaps faculty should be evaluated so that their different lived experiences are also taken into consideration.

À quel point des données peuvent-elles guider de manière importante les décisions concernant la titularisation? Si les données ne reflètent qu'un aspect des faits, peut-être que l'on devrait aussi évaluer les professeurs en tenant compte de leurs expériences différentes.

Tenure is one of the foundational concepts upon which Canada’s modern university system has been built. It is not without its issues, of course: Faculty often struggle in their early years as they work toward earning tenure.

How much can data-gathering inform tenure decisions? Or reflect the efforts and experiences of individuals who go through years of tenure-track in the hopes of meeting the expectations for tenure?

DATA ABOUT TENURE

Let’s begin by considering that there are data available about tenure—important and interesting data. One example is data from the *CAUT Almanac of Post-Secondary Education in Canada* (see Table 1), which provides quantitative details about the number of faculty with tenure, in tenure-track positions, and in other types of positions. The data are constrained by changes in government policy regarding the collection of the data and, to the best of our knowledge, more recent figures are not readily available. (While it is tangential to this article, when using data for evaluative purposes it is important to consider the possibility that data collection can be altered over time, causing the evaluative process to lose clarity.)

TABLE 1: TENURE BY THE NUMBERS

Year	Number in position			
	Tenured	Tenure-track	Other	Total for year
2010/2011*	21,870	7,158	5,049	43,455*
2008/2009	19,671	7,767	3,828	40,482**
2007/2008	19,137	7,803	12,915	39,855
2006/2007	27,633	7,590	3,840	39,063
2005/2006	26,916	7,530	3,822	38,268
2004/2005	15,543	7,104	11,361	34,008
2003/2004	15,111	6,666	10,995	32,772
2001/2002	18,099	5,402	3,369	26,870
2000/2001	18,266	4,758	3,313	26,337

* Figures in this row do not add up to the total because the statistics for Quebec were not broken down into these categories.

** Figures in this row do not add up to the total due to reporting problems in Quebec footnoted in the *CAUT Almanac*.

What Table 1 reveals is how rapidly faculty tenure numbers can change. The decline from 27,633 to 19,137 tenured faculty seen between 2006/2007 and 2007/2008 seems to defy the often suggested permanence of tenure.

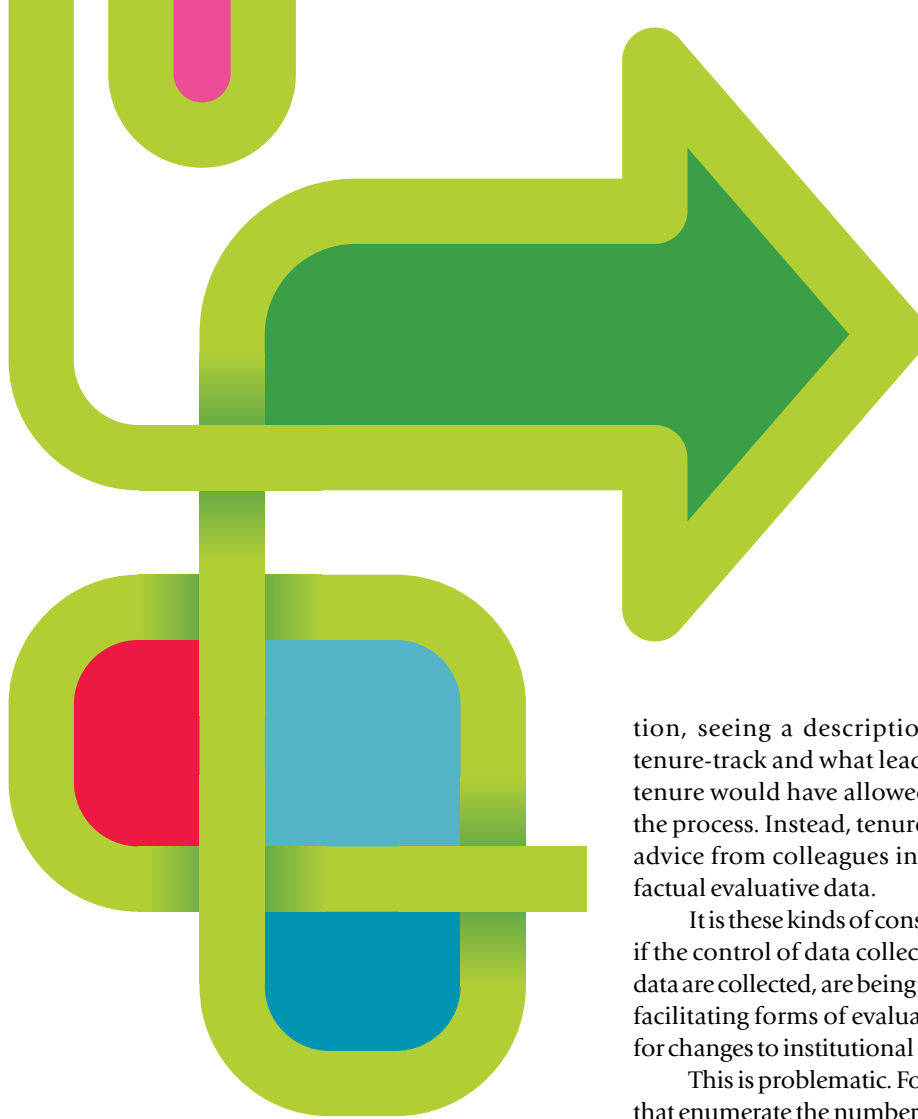
Similarly, the “Other” column values decline dramatically, from 11,361 to 3,822 between 2004/2005 and 2005/2006. These data allow us to evaluate the stability of faculty and confirm the “precarious” moniker used to describe university faculty. In that sense, data-gathering does facilitate some understanding of tenure.

Another collection of data, also from the *CAUT Almanac*, outlines the number of faculty within different ranks (see Table 2). It is often assumed that promotions hinge on tenure; however, the data clearly show that policies do allow promotion and tenure to be treated somewhat independently. Consider, for example, that in 2004/2005 there were 15,543 tenured faculty but 21,945 full and associate professors. This certainly allows one to conclude that university policies are not as simple as might be assumed. Over the decade covered by the table, the number of full professors increased by 24 per cent, associate professors and assistant professors both increased by 47 per cent, and lecturers by 151 per cent—demonstrating that growth has occurred and that it has favoured non-research faculty. While this category may include permanent faculty with no research requirement, it also includes sessional instructors, who hold the least permanent positions. Both undermine one of the core missions of the university: research. The data also reveal that new research faculty are embedded in a situation with proportionately fewer research mentoring options.

TABLE 2: RANK BY THE NUMBERS

Year	Number in rank				
	Full professor	Associate professor	Assistant professor	Lecturer	Total for year
2010/2011	14,022	15,036	10,077	4,293	43,428
2008/2009	13,479	13,752	10,788	2,517	40,536
2007/2008	13,293	13,161	10,914	2,442	39,810
2006/2007	13,143	12,747	10,794	2,340	39,024
2005/2006	13,149	12,279	10,614	2,256	38,298
2004/2005	11,325	10,620	9,981	2,082	34,008
2003/2004	11,412	10,401	9,045	1,914	32,772
2001/2002	11,282	10,295	7,661	1,609	30,847
2000/2001	11,304	10,248	6,878	1,710	30,140





Tenure-track faculty are given specious advice from colleagues in building corridors, rather than factual evaluative data.

THE STORIES THE DATA DO NOT TELL

Although the data presented show various trends and phenomena, they do not, in and of themselves, provide information about tenure decision-making or the experiences of faculty members entering academia. Even if the *CAUT Almanac* analysis was done at the institutional level and by subject discipline, which is feasible, the information would not suffice. There should be skepticism about its value for informing individual faculty members or tenure and promotion committees about the process of achieving tenure.

Previously, we shared our concerns that evaluative data did not reflect personal decision-making or lived experiences. This led us to look for narratives about the meaning of tenure-track. We were both well established in public education, Tim as an experienced teacher and Tory in educational leadership, before moving to positions in higher education, giving up employment security to pursue tenure. In making that decision, it would have been helpful to have data showing the success rate of achieving tenure. This would have allowed us to make more informed decisions about moving to higher education and, perhaps, such data would illustrate just how precarious it is to pursue tenure. In addi-

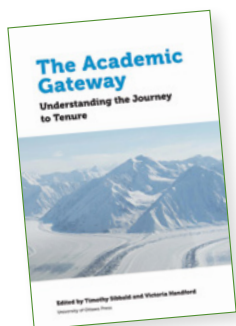
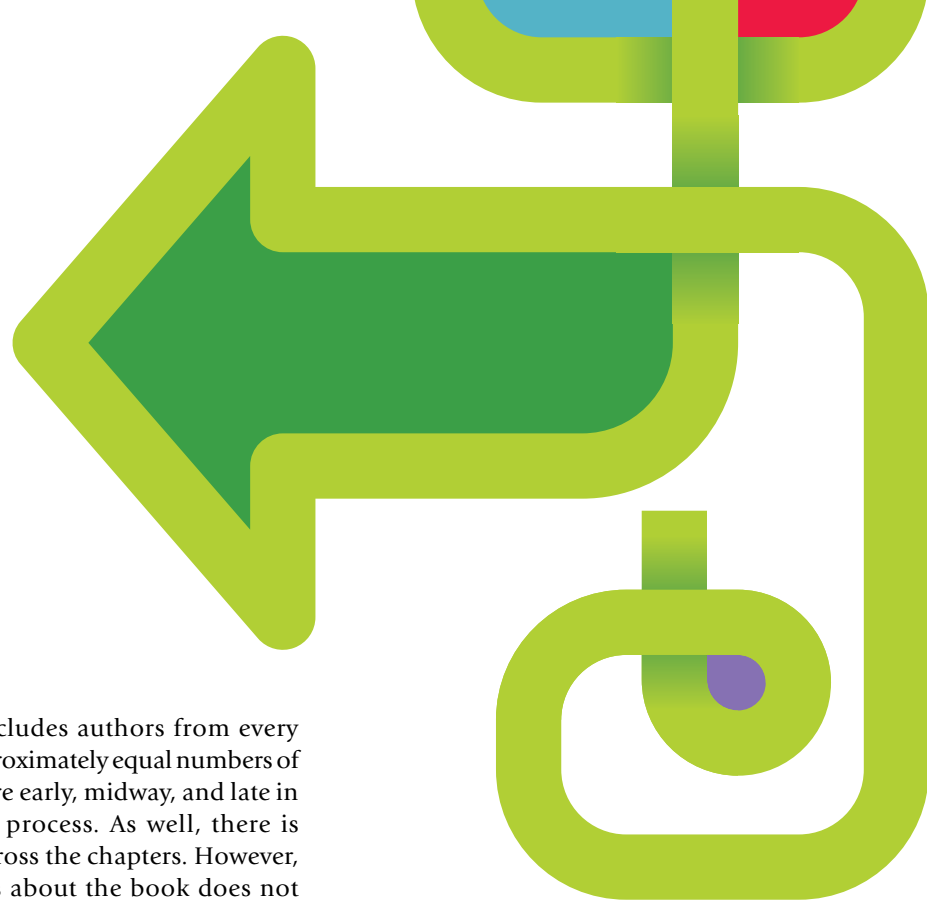
tion, seeing a description of what goes wrong during tenure-track and what leads to failed attempts at achieving tenure would have allowed us to self-evaluate throughout the process. Instead, tenure-track faculty are given specious advice from colleagues in building corridors, rather than factual evaluative data.

It is these kinds of considerations that make one wonder if the control of data collection, and the choices about what data are collected, are being manipulated to avoid revealing or facilitating forms of evaluation that might lead to demands for changes to institutional systems and governance.

This is problematic. For instance, data could be collected that enumerate the number of publications in various categories between successful and unsuccessful applications by subject area, immediately reinforcing notions of tenure being based upon “bean-counting,” and therefore helping to diminish academic freedom. How can faculty be innovative and responsive to the dynamic nature of research if they are ultimately focused on tabulations in fixed categories to achieve tenure? Academic freedom addresses this by ensuring that professors can make choices best suited to the innovations arising from their research. Tenure cannot be measured in terms of static, quantitative achievements that defy the dynamic university environment that is supposed to facilitate innovation.

These considerations led us to develop a book about tenure-track experiences. *The Academic Gateway: Understanding the Journey to Tenure¹*, brings together narratives of tenure-track experiences from across Canada. To make the task tractable, it focuses on faculty members in education, a discipline in which work experience as a teacher often informs the professorial role. This is not to suggest that other disciplines do not have overlap between careers and academia, only that education was viewed as sharing one meaningful overlap—teaching is 40 per cent of the assignment. In this way, we feel the narratives are likely more biased toward illustrating smoother transitions than might be found in other disciplines. However, they also demonstrate that there is a considerable diversity of experience.

One difficulty that arises with data collection is that it does not recognize the fine details that are part of lived experiences.



The book includes authors from every province, and approximately equal numbers of academics who are early, midway, and late in the tenure-track process. As well, there is gender equity across the chapters. However, knowing all this about the book does not allow one to evaluate the experiences that arise within its chapters. These differences are reflected in the varied ways in which the authors entered higher education—for example, some worked while pursuing graduate school and others took time away from their careers—and the experiences they had after joining the academy. The chapters speak to how lived experiences can be much more illuminating than simple quantitative data. In essence, it is the recognition of the importance of these varied experiences that obliges the use of peer evaluation, rather than just data, to make decisions about tenure.

Consider the following quotes from *The Academic Gateway* that show a variety of faculty experiences and speak to very different circumstances when entering academia: “My wife and I left permanent education positions, financial security, and family networks in Alberta... as I was offered a one-year term position as an assistant professor... This was not a tenure-track appointment, but I was told it would turn into one.” Another writer speaks of the emotional strain of moving: “In my new surroundings—living alone for the first time in 29 years, feeling lost and lonely, and having left family and friends behind—I began to seriously doubt my ability to cope.” There are also changes in stature noted: “I am on a steep learning path. Transitioning from a position where I had designated authority to a position as a junior faculty member means I am negotiating and navigating the bounds of my role.” While a critical eye can point to these experiences and say that none are relevant to tenure considerations, they are evidence of the deep changes that moving into academia can have on different individuals. For example, the following quote describes a formal application

within tenure-track and demonstrates issues with data-gathering and requiring applicants to conform to specific criteria:

This template featured categories and requirements entirely (and appropriately) geared to an academic career; but I found myself, for example, unable to list technical papers and publications I had written because policies in government had prohibited named authorship. This issue turned out to be the tip of the iceberg, as the methods, vocabulary, and rhetorical strategies I used to succeed in government did not transfer as readily as expected.

This issue may have become even more prevalent in recent decades, because graduate students tend to be older, taking longer to graduate as they work to pay for their education.

One difficulty that arises with data collection is that it does not recognize the fine details that are part of lived experiences. The absence of these details creates an environment in which academic supervision does not have to consider anything other than quantitative data. However, many authors in the book speak of circumstances in their lived tenure-track experience that needed a more nuanced analysis and a more personalized response in order to support their personal and professional growth. It is alarming to contrast the personal upheaval of moving into the academy with an impersonal, data-oriented approach that is oblivious to personal struggles.

The narratives in the book speak to the human side of tenure-track. While some may favour a quantitative approach to tenure decisions, the book’s qualitative narratives are demonstrable evidence that this approach does not work.

Innovative research is rarely clear-cut and frequently ill-defined, but here it speaks to the very core of what tenure is.

A MODEL OF TENURE: GRIEF AND SELF-DETERMINATION

After the book was published, we continued thinking about tenure and about the book as a source of cross-case narratives. This led to a model of tenure that comprises two components. The first component addresses the change in workplace culture when one moves into a tenure-track position: You need to learn the ways of a new institution, a sense of isolation occurs, and a loss of community becomes apparent when individuals move long distances. We address these essentially as a grieving process that models the sense of loss associated with prior workplace colleagues, community familiarity, and a loss of capacity to efficiently address fairly routine tasks.

It appears that, as tenure-track progresses well for individuals, this grief component of the model runs its course. Every professor who is on track to earn tenure, or believes that they are, will overcome any and all of the concerns that led to inclusion of this grief component. Of course, in some cases, individuals simply do not manifest any grief. For example, a person who moves directly from graduate studies to a tenure-track position at the same institution might transition quite seamlessly. We have also hypothesized that when an individual perceives that their tenure-track is not progressing in a successful fashion, the grief will manifest itself in more pronounced ways that reflect a continued sense of regret.

The other component of the model that endures after earning tenure is self-determination. This is a complex theory that reflects the wide array of university professors' skills and tasks.² Perhaps the most significant feature of this component is the aspect of autonomy—the capacity to self-direct one's work. It is not the absence of having to choose (i.e., one cannot use self-determination to justify being lethargic when they have tenure), but about choosing from the different options that arise. In terms of research, this includes academic freedom, but it also includes making choices when academic freedom is not a significant issue.

One of the consequences of this model is that self-determination theory is a grand theory that will not easily succumb to quantitative evaluation. How can one, for example, measure autonomy? Even if it could be measured, as one progresses through the tenure-track years, how much autonomy is required to grant tenure? In this sense, the research aspect of higher education requires freedom to explore where one's expertise suggests they should explore.

Within the academy, it is clearly important for tenured faculty to have intrinsic motivation. This is a cornerstone of self-determination theory, but runs contrary to having a strictly measured criterion for tenure success. There is a fundamental difficulty fostering intrinsic motivation if specific

details will be used to decide the ultimate outcome. The paradoxical nature of tenure decisions manifests itself in the need to assess how well an individual grows when their role is largely self-determined. To have a pre-defined measure would bias the self-determined element and thus defy the very definition of tenure. What is important within self-determination theory is the relatedness inherent within it, whereby circumstantial relationships contribute positive feedback that helps develop intrinsic motivation. Rather than suggesting the importance of using this data, it points to a need to further humanize the process and encourage social interaction through the tenure-track years as a way to support both personal and professional growth. Just as autonomy does not support post-tenure lethargy, intrinsic motivation does not support treating tenure-track faculty as if they live and work within a vacuum.

Fundamentally, self-determination theory operates within individuals as people. Furthermore, evaluation is a broadening of the notion of self that brings peers into the process of evaluation. Peer review is an approach to evaluation that shares the experience of developing one's self-determination. It is qualitative because of the small sample—the evaluation of a single individual—and facilitates the human capacity for exercising choice. Tenure is fundamentally about personal growth and developing niche expertise, neither of which is suitable for a data-collection approach. It is, however, exactly what some qualitative methods were developed to reveal.

CONCLUSION

It is disturbing that ideas for altering evaluative approaches do not seem to consider the theoretical grounding of what is being proposed. *The Academic Gateway* does not provide details about which authors have achieved tenure. This is left purposely unresolved so that readers can consider how they might assess each individual's merit for tenure. We doubt that anyone can come up with a data-collection approach that will successfully appraise the eventual tenure decisions of the book's authors. **AM**

Tim Sibbald is an Associate Professor in the Schulich School of Education at Nipissing University. Victoria Handford is an Assistant Professor in the Faculty of Education and Social Work at Thompson Rivers University.

1. TM Sibbald & V Handford (Eds.), *The Academic Gateway: Understanding the Journey to Tenure*, Ottawa: University of Ottawa Press, 2017. For the quotes contained in this article, see pp. 29–40, 93–110, 179–196, and 249–264.
2. EL Deci & RM Ryan, "The 'what' and 'why' of goal pursuits: Human needs and the self-determination of behavior," *Psychological Inquiry*, 11:4 (2000): 227–268; AV Broeck, DL Ferris, CH Chang, & CC Rosen, "A review of self-determination theory's basic psychological needs at work," *Journal of Management*, 42:5 (2016): 1195–1229.

The abuses and perverse effects of quantitative evaluation in the academy

Yves Gingras

The world of academic research is scored according to so-called “objective” measures, with an emphasis on publications and citations. But the very foundations of this approach are flawed. Is it time to abandon these simplistic ranking schemes?

Les professeurs et les chercheurs universitaires sont de plus en plus évalués à l'aide de mesures dites « objectives », qui mettent l'accent sur les publications et les citations. Mais le fondement même de cette approche est problématique. Le temps est-il venu d'abandonner ces méthodes de notation simplistes?



Since the neoliberal ideology of the “new public management” and its introduction of rankings in academia began in the 1990s, researchers and administrators have become increasingly familiar with the terms “evaluation,” “impact factors,” and “h-index.” Since that time, the worlds of research and higher education have fallen prey to a dangerous evaluation fever. It seems that we want to assess everything, including teachers, faculty, researchers, training programs, and universities. “Excellence” and “quality” indicators have proliferated in usage without anyone really understanding what these terms precisely mean or how they are determined.

Bibliometrics, a research method that considers scientific publications and their citations as indicators of scientific production and its uses, is one of the primary tools that informs the many “excellence indicators” that this administrative vision of higher education and research is attempting to impose on everyone. Whether ranking universities, laboratories, or researchers, calculating the number of publications and citations they receive often serves as an “objective” measure for determining research quality.

It is therefore important to understand the many dangers associated with the growing use of oversimplified bibliometric indicators, which are supposed to objectively measure researchers’ productivity and scientific impact. This paper focuses on analyzing two key indicators used extensively by both researchers and research administrators. It also examines the perverse effects that the oversimplified use of bad indicators has upon the dynamics of scientific research, specifically in the areas of social and human sciences.

THE IMPACT FACTOR: CORRUPTING INTELLECTUAL OUTPUT

A journal’s impact factor (IF) is a simple, mathematical average of the number of citations received in a given year (e.g., 2016) for articles published by a journal during the previous two years (in this case, 2014 and 2015). The IF has been calculated and published every year since 1975 in the Web of Science Journal Citation Reports. As early as the mid-1990s, experts in bibliometrics were drawing attention to the absurdity of confusing articles and journals. However, this did not stop decision-makers—who themselves are supposedly rational researchers—from using a journal’s IF to assess researchers and establish financial bonuses based directly on the numerical value of the IF. For example, as the journal *Nature* reported in 2006, the

Pakistan Ministry of Science and Technology calculates the total IF of articles over a year to help it establish bonuses ranging between \$1,000 and \$20,000. The Beijing Institute of Biophysics established a similar system: An IF of 3 to 5 brings in 2,000 yuan (\$375) per point and an IF above 10 brings in 7,000 yuan (\$1,400) per point.

However, in an editorial in the same issue, *Nature* criticized this system, noting that it is impossible for a mathematical journal to score an IF value as high as a biomedical research journal due to the substantially larger number of potential citers in the biomedical sciences. No sensible person believes that biomedical articles are superior to math articles, nor can they believe that this scoring system justifies granting one group of authors a larger bonus than another group. And, in another more recent (and ugly) example of the kind of intellectual corruption generated by taking the ranking race seriously, universities have contacted cited researchers who are working for other institutions and offered these researchers compensation for including the university as an affiliated body in the individual’s next article.¹ These fictitious affiliations, without real teaching or research duties, allow marginal institutions to enhance their position in university rankings without having to maintain real laboratories.

These extreme cases should be enough to warn university managers and their communications departments away from the use or promotion of such inaccurate rankings. In short, it is important to scrutinize the ranking system’s “black box,” rather than accepting its results without question.

The exploitation of these false rankings and indicators to promote institutional and individual achievement is a behaviour that reveals an ignorance of the system’s flaws. Only the institutions that benefit from association with these rankings, researchers who profit from incorrectly computed bonuses based on invalid indicators, and journals that benefit from the evaluative use of impact factors, can believe—or feign to believe—that such a system is fair, ethical, and rational.

THE H INDEX EPIDEMIC

In the mid-2000s, when scientific communities started devising bibliometric indices to make individual evaluations more objective, American physicist Jorge E. Hirsch, from the University of California in San Diego, came up with a proposition: the h index. This index is defined as being equal to the number N of articles published by a researcher that received at least N citations since their publication. For



“Excellence” and “quality” indicators have proliferated in usage without anyone really understanding what these terms precisely mean or how they are determined.

example, if an author has published 20 articles, 10 of which were cited at least 10 times each since their publication, the author will have an h index of 10. It is now common to see researchers cite their h index on their Facebook pages or in their *curricula vitae*.

The problematic nature of the h index is reflected in the very title of Hirsh’s article published in a journal that is usually considered prestigious, the *Proceedings of the National Academy of Sciences of the United States of America*, “An index to quantify an individual’s scientific research output.” In fact, this index is neither a measure of quantity (output) nor a measure of quality or impact: It is a combination of both. It arbitrarily combines the number of articles published and the number of citations received. In the eye of its creator, this index was meant to counter the use of the total number of articles published, a metric that does not take their quality into account. The problem is that the h index is itself strongly correlated with the total number of articles published, and is therefore redundant.

Furthermore, the h index has none of the basic properties of a good indicator. As Waltman and van Eck demonstrated, the h index is incoherent in the way it ranks researchers whose number of citations increases proportionally, and it therefore “cannot be considered an appropriate indicator of a scientist’s overall scientific impact.”²

This poorly constructed index also causes harm when it is used as an aid in the decision-making process. Let us compare two scenarios: A young researcher has published five articles, which were cited 60 times each (for a given period); a second researcher, of the same age, is twice as prolific and wrote 10 articles, which were cited 11 times each. The second researcher has an h index of 10, while the first researcher only has an h index of 5. Should we conclude that the second researcher is twice as “good” as the first one and should therefore be hired or promoted ahead of the first researcher? Of course not, because the h index does not really measure the relative quality of two researchers and is therefore not a technically valid indicator.

Despite these fundamental technical flaws, use of the h index has become widespread in many scientific disciplines. It seems as though it was created primarily to satisfy the ego of some researchers. Let us not forget that its rapid dissemination has been facilitated by the fact that it is calculated automatically within journal databases, making it quite easy to obtain. It is unfortunate to see scientists, who purportedly study mathematics, lose all critical sense when presented with this flawed and oversimplified number. It confirms an old English saying, “Any number beats no number.” In other words, it is better to have an incorrect number than no number at all.

A MULTIDIMENSIONAL UNIVERSE

What is most frustrating in the debates around research evaluation is the tendency to try to summarize complex results with a single number. The oversimplification of such an approach becomes obvious when one realizes that it means transforming a space with many dimensions into a one-dimensional space, thus realizing Herbert Marcuse's prediction of the advent of a *One-Dimensional Man*. In fact, by combining various weighted indicators to get a single number, we lose the information on each axis (indicators) within the multidimensional space. Everything is reduced to a single dimension.

Only by considering the many different initial indicators individually can we determine the dimensions of concepts such as research quality and impact. While postsecondary institutions and researchers are primarily interested in the academic and scientific impact of these publications, we should not ignore other impacts for which valid indicators are easily accessible. Think of the economic, societal, cultural, environmental, and political impacts of scientific research, for example.

In the case of universities, research is not the only mission and the quality of education cannot be measured solely by bibliometric indicators that ignore the environment in which students live and study, including the quality of buildings, library resources, or students' demographic backgrounds. For these dimensions to emerge, we must avoid the "lamppost syndrome," which leads us to only look for our keys in brightly lit places rather than in the specific (but dark) places where they are actually to be found. It is therefore necessary to go beyond readily accessible indicators and to conduct case studies that assess the impacts for each of the major indicators. It is a costly and time-consuming qualitative operation, but it is essential for measuring the many impacts that research can have.

The simplistic nature of rankings culminate in annual attempts to identify the world's "best" universities, as if the massive inertia of a university could change significantly every year! This in itself should suffice to show that the only aim of these rankings is to sell the journals that print them.

QUANTIFYING AS A WAY TO CONTROL

The heated arguments around the use of bibliometric indicators for assessing individual researchers often neglects a fundamental aspect of this kind of evaluation, which is the role of peers in the evaluation process. Peer review is a very old and dependable system that requires reviewers to have first-hand knowledge of the assessed researcher's field of study. However, in an attempt to assert more control over the evalua-

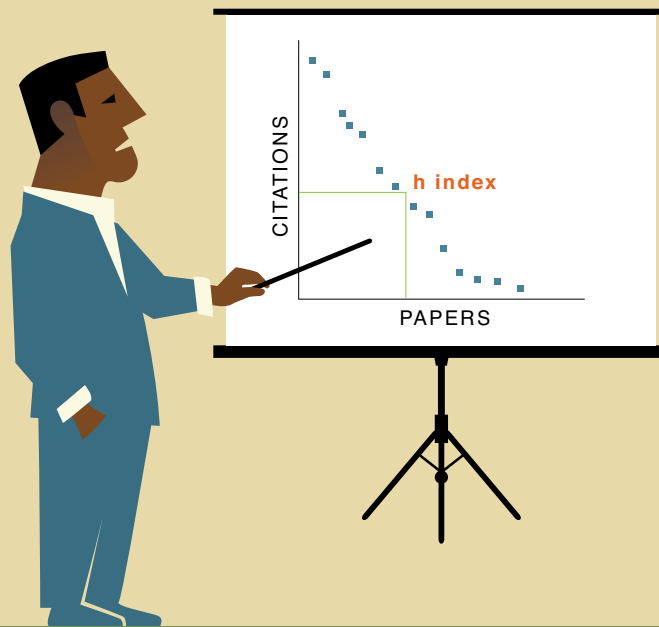
tion process, some managers in universities and granting agencies are pushing forward with a new concept of "expert review" in which an individual, often from outside the field of research being considered, is responsible for evaluating its merits. A standardized quantitative evaluation, such as the h index, makes this shift easier by providing supposedly objective data that can be used by anyone. It is in this context that we need to understand the creation of journal rankings as a means to facilitate, if not to mechanize, the evaluation of individuals. This constitutes a *de facto* form of Taylorization of the evaluation process—the use of a scientific method to de-specialize the expertise needed for evaluation.

Thus surfaces a paradox. The evaluation of a researcher requires appointing a committee of peers who know the researcher's field very well. These experts would already be familiar with the best journals in their field and do not need a list concocted by some unknown group of experts ranking them according to different criteria. On the other hand, these rankings allow people who don't know anything about a field to pretend to make an expert judgment just by looking at a ranked list without having to read a single paper. These individuals simply do not belong on an evaluation committee. Therefore, the proliferation of poorly built indicators serves the process of bypassing peer review, which does consider productivity indices but interprets them within the specific context of the researcher being evaluated. That some researchers contribute to the implementation of these rankings and the use of invalid indicators does not change the fact that these methods minimize the role of the qualitative evaluation of research by replacing it with flawed mechanical evaluations.

PSEUDO-INTERNATIONALIZATION AND THE DECLINE OF LOCAL RESEARCH

A seldom-discussed aspect of the importance given to impact factors and journal rankings is that they indirectly divert from the study of local, marginal, or less popular topics. This is particularly risky in human and social sciences, in which research topics are, by nature, more local than those of the natural sciences (there are no "Canadian" electrons). Needless to say, some topics are less "exportable" than others.

Since the most frequently cited journals are in English, the likelihood of being published in them depends on the interest these journals have in the topics being studied. A researcher who wants to publish in the most visible journals would be well advised to study the United States' economy rather than the Bank of Canada's uniqueness or Quebec's regional economy, topics that are of little interest to an American journal. Sociologists whose topic is international or who put forward more general theories are more



The simplistic nature of rankings culminate in annual attempts to identify the world’s “best” universities, as if the massive inertia of a university could change significantly every year!

likely to have their articles exported than those who propose an empirical analysis of their own society. If you want to study Northern Ontario’s economy, for example, you are likely to encounter difficulty “internationalizing” your findings.

Yet is it really less important to reflect on this topic than it is to study the variations of the New York Stock Exchange? As a result, there is a real risk that local but sociologically important topics lose their value and become neglected if citation indicators are mechanically used without taking into account the social interest of research topics in the human and social sciences.

CONCLUSION: NUMBERS CANNOT REPLACE JUDGEMENT

It is often said—without providing supporting arguments—that rankings are unavoidable, and that we therefore have to live with them. This is, I believe, a false belief and, through resistance, researchers can bring such ill-advised schemes to a halt. For example, in Australia, researchers’ fierce reaction to journal rankings has succeeded in compelling the government to abandon the use of this simplistic approach to research evaluation.

In summary, the world of research does not have to yield to requirements that have no scientific value and that run against academic values. Indeed, French-language journals and local research topics that play an invaluable role in helping us better understand our society have often been the hardest hit by these ill-advised evaluation methods, and so fighting back against this corruption is becoming more important every day. **AM**

Yves Gingras is a Professor in the History Department and Canada Research Chair in History and Sociology of Science at the Université du Québec à Montréal.

NOTE:

This article is a translation of a revised and shorter version of the essay, « Dérives et effets pervers de l’évaluation quantitative de la recherche : sur les mauvais usages de la bibliométrie », in *Revue internationale PME* 28;2 (2015): 7-14. For a more in-depth analysis, see: Yves Gingras, *Bibliometrics and Research Evaluation: Uses and Abuses*, Cambridge: MIT Press. 2016.

1. Yves Gingras, “How to boost your university up the rankings,” *University World News*, (2014) July 18;329, <http://www.university-worldnews.com/article.php?story=20140715142345754>. Refer also to the many responses in *Science*, (2012), March 2;335: 1040-1042.
2. L Waltman and NJ van Eck, “The inconsistency of the h-index,” 2011, <http://arxiv.org/abs/1108.3901>.

OCUFA Report

Stay up to date with the latest developments and analysis of issues in Ontario postsecondary education.



OCUFA Report is a weekly resource for Ontario's professors and academic librarians. Learn about new initiatives in higher education, collective bargaining issues, and OCUFA conferences and events.

Sign up for the weekly OCUFA Report email newsletter today:

www.ocufa.on.ca/report

OCUFA

Ontario Confederation of University Faculty Associations
Union des Associations des Professeurs des Universités de l'Ontario

Dérives et effets pervers de l'évaluation quantitative de la recherche

Yves Gingras

Les professeurs et les chercheurs universitaires sont de plus en plus évalués à l'aide de mesures dites « objectives », qui mettent l'accent sur les publications et les citations. Mais le fondement même de cette approche est problématique. Le temps est-il venu d'abandonner ces méthodes de notation simplistes?

The world of academic research is scored according to so-called "objective" measures, with an emphasis on publications and citations. But the very foundations of this approach are flawed. Is it time to abandon these simplistic ranking schemes?



Avec l'arrivée en milieu universitaire de l'idéologie néolibérale adossée aux techniques du nouveau management public avec ses « tableaux de bord », surtout depuis les années 1990, les chercheurs et les administrateurs utilisent de plus en plus souvent les mots « évaluation », « facteurs d'impact », « indice h ». Le monde de la recherche et de l'enseignement supérieur, est ainsi la proie d'une véritable fièvre de l'évaluation. On veut tout évaluer: les enseignants, les professeurs, les chercheurs, les programmes de formation et les universités. Les indicateurs « d'excellence » et de « qualité » se multiplient sans que l'on sache toujours sur quelles bases ils ont été construits.

Parmi les outils utilisés pour mettre au point les nombreux « indicateurs d'excellence » qu'une vision gestionnaire de l'enseignement supérieur et de la recherche tente d'imposer à tous comme une évidence, une place de choix est aujourd'hui accordée à la bibliométrie—méthode de recherche qui consiste à utiliser les publications scientifiques et leurs citations comme indicateurs de la production scientifique et de ses usages. Que ce soit pour classer les universités, les laboratoires ou les chercheurs, le calcul du nombre de publications et des citations qu'elles reçoivent sert souvent de mesure « objective » de la valeur des résultats de recherche des uns et des autres.

Il est donc important de rappeler, même brièvement, les nombreux dangers que comportent l'usage simpliste qui tend à se répandre de l'utilisation mécanique d'indicateurs bibliométriques supposés mesurer de façon « objective » la productivité et l'impact scientifique des chercheurs. Nous nous limiterons ici à analyser les usages des deux principaux indicateurs amplement utilisés tant par les chercheurs que par les administrateurs de la recherche. Nous nous pencherons aussi sur les effets pervers des usages simplistes de mauvais indicateurs sur la dynamique de la recherche scientifique particulièrement dans les domaines des sciences sociales et humaines.

LES MAUVAIS USAGES DU FACTEUR D'IMPACT

Calculé et publié chaque année depuis 1975 dans le *Journal Citation Reports du Web of Science* (maintenant propriété de Clarivate Analytics) le facteur d'impact (FI) d'une revue consiste en une simple moyenne arithmétique du nombre de citations obtenues une année donnée (disons 2016) par les articles publiés par une revue au cours des deux années précédentes (soit 2014 et 2015). Bien que, dès le milieu des années 1990, des experts en bibliométrie n'aient cessé d'attirer l'attention sur l'absurdité de confondre ainsi les arti-

cles et les revues, cela n'a pas empêché les « décideurs » et, il faut le souligner, de chercheurs supposément rationnels, d'utiliser le facteur d'impact des revues pour évaluer les chercheurs et instituer des systèmes de primes fondés directement sur la valeur numérique du facteur d'impact des revues! Comme le rapportait la revue *Nature* en 2006, le ministère de la Science du Pakistan calcule la somme des facteurs d'impact des articles sur une année pour fixer une prime variant entre 1 000 et 20 000 dollars! En Chine, l'Institut de biophysique de Beijing a établi un système semblable : un FI entre 3 et 5 rapporte 2 000 yuans (375 \$) par point, et 7 000 yuans (1,400 \$) par point si le FI est au-dessus de 10.

Dans un éditorial du même numéro, la revue dénonçait cette absurdité. Or, il est impossible que le FI d'une revue de mathématiques (par exemple) ait jamais la valeur de celui d'une revue de recherche biomédicale! Pourtant, aucune personne sensée ne peut croire que les articles de médecine sont tous supérieurs aux articles de mathématiques et justifient donc d'accorder à leurs auteurs une prime plus importante. Dernier exemple montrant le genre de corruption intellectuelle engendrée par la course aux classements : certaines universités contactent des chercheurs très cités qui sont employés par d'autres institutions et leur offrent d'ajouter leur adresse dans leurs publications en échange d'une rémunération¹. Ces affiliations factices, auxquelles aucune tâche d'enseignement ou de recherche n'est attachée, et dont les chercheurs qui y participent sont complices, permettent à des institutions marginales d'améliorer facilement leur position dans les classements des universités sans avoir à créer de véritables laboratoires.

Ces cas extrêmes devraient suffire pour mettre en garde les gestionnaires d'université, ou leurs chargés de communication, contre les usages médiatiques de tels classements douteux. En somme, mieux vaut regarder à l'intérieur de la « boîte noire » des classements plutôt que de l'accepter telle quelle comme si elle contenait un beau cadeau de bienvenue...

L'usage abusif de classements et d'indicateurs faussement précis constitue en somme un comportement qui trahit l'ignorance des propriétés des indicateurs utilisés. Seul l'opportunisme des chercheurs, qui profitent de primes mal calculées, et des revues, qui profitent de l'usage évaluatif des facteurs d'impact, peut les amener à croire, ou à feindre de croire, qu'un tel système est juste et rationnel.

L'ÉPIDÉMIE DE « L'INDICE H »

Il est devenu courant de voir des chercheurs indiquer sur leur page face book ou dans leur curriculum vitae leur



Les indicateurs « d'excellence » et de « qualité » se multiplient sans que l'on sache toujours sur quelles bases ils ont été construits.

« indice h ». Au milieu des années 2000, alors que les milieux scientifiques avaient commencé à concocter des indices bibliométriques pour rendre les évaluations individuelles plus « objectives », le physicien américain Jorge E. Hirsch, de l'université de Californie à San Diego, y est allé de sa proposition : l'indice h. Cet indice est défini comme étant égal au nombre d'articles N qu'un chercheur a publiés et qui ont obtenu au moins N citations chacun depuis leur publication. Par exemple, un auteur qui a publié 20 articles parmi lesquels 10 sont cités au moins 10 fois chacun aura un indice h de 10.

Le caractère improvisé de cet indice se voit déjà au titre même de l'article paru dans une revue pourtant considérée comme « prestigieuse », *les Proceedings de l'Académie nationale des sciences des États-Unis* : « un indice pour quantifier la production (*output*) scientifique d'un chercheur ». En fait, cet indice n'est ni une mesure de quantité (*output*), ni une mesure de qualité ou d'impact, mais un composite des deux. Il combine de façon arbitraire le nombre d'articles publiés et le nombre de citations obtenues. Cet indice est supposé contrer l'usage du seul nombre d'articles, lequel ne tient pas compte de leur « qualité ». Le problème c'est qu'il a rapidement été démontré que l'indice h est lui-même très fortement corrélé au nombre total d'articles et se révèle ainsi redondant!

Pis encore, il n'a aucune des propriétés de base que doit posséder un bon indicateur. Comme l'ont montré Ludo Waltman et Nees Jan van Eck, l'indice h est en réalité inco-

hérent dans la manière dont il classe des chercheurs dont le nombre de citations augmente de façon proportionnelle. Ces auteurs en concluent que l'indice h « ne peut être considéré comme un indicateur approprié de l'impact scientifique global d'un chercheur »².

Cet indice mal construit est même dangereux lorsqu'il est utilisé comme aide à la prise de décisions car il peut générer des effets pervers. Un exemple simple suffit à le démontrer. Comparons deux cas de figure : un jeune chercheur a publié seulement cinq articles, mais ceux-ci ont été cités 60 fois chacun (pour une période de temps donnée) ; un second chercheur, du même âge, est deux fois plus prolifique et possède à son actif 10 articles, cités 11 fois chacun. Ce second chercheur a donc un indice h de 10, alors que le premier a un indice h de 5 seulement. Peut-on en conclure que le second est deux fois « meilleur » que le premier et devrait donc être embauché ou promu? Bien sûr que non... On voit ici que l'indice h ne mesure pas vraiment la qualité relative de deux chercheurs et est donc un indicateur techniquement invalide.

Malgré ces défauts techniques réhilitoires, l'usage de l'indice h s'est généralisé dans plusieurs disciplines scientifiques. Il semble taillé sur mesure pour satisfaire d'abord le narcissisme de certains chercheurs. N'oublions pas que sa diffusion rapide a aussi été facilitée par le fait qu'il est calculé directement dans toutes banques de données et s'obtient donc sans aucun effort! Il est tout de même navrant de constater que des scientifiques pourtant supposés avoir fait des

études en mathématiques perdent tout sens critique devant un chiffre simpliste—cela vient confirmer un vieil adage anglais qui a toutes les apparences d'une loi sociale : « *Any number beats no number.* » En d'autres termes, mieux vaut un mauvais chiffre que pas de chiffre du tout...

UN UNIVERS À PLUSIEURS DIMENSIONS

Le plus irritant dans les débats sur l'évaluation de la recherche est la tendance à vouloir tout résumer par un seul chiffre. Le simplisme d'une telle démarche devient patent quand on observe que cela revient à transformer un espace à plusieurs dimensions en un espace de dimension zéro ! En effet, un nombre, considéré ici comme un point, est de dimension zéro et combiner différents indicateurs pondérés pour obtenir un seul chiffre fait perdre l'information sur chacun des axes (indicateurs) d'un espace à plusieurs dimensions. Au mieux, si on considère que le point est sur une ligne, on a quand même réduit le tout à une seule dimension.

Or, seule la prise en compte de plusieurs indicateurs différents permet de tenir compte des différentes dimensions d'un concept, tel ceux de qualité et d'impact de la recherche. Ainsi, le milieu académique est d'abord intéressé par l'impact scientifique des publications, mais on ne saurait négliger d'autres types d'impacts pour lesquels on trouve plus ou moins facilement des indicateurs valides. Pensons aux impacts économiques, sociétaux, culturels, environnementaux, politiques de la recherche scientifique.

Ainsi, dans le cas des universités, la recherche n'est qu'une fonction de l'institution, et la qualité de l'enseignement ne se mesure pas à l'aune de la recherche, en faisant abstraction de l'environnement dans lequel baignent les étudiants (qualité des édifices, ressources bibliothécaires, etc.). Si l'on veut faire émerger ces dimensions, il faut dépasser le « syndrome du lampadaire » (« *lamp-post syndrome* »), qui porte à chercher ses clés dans une zone éclairée plutôt qu'à l'endroit précis (mais sombre) où elles ont en fait été égarées. Il est donc nécessaire d'aller au-delà des indicateurs facilement accessibles et de faire des études de cas afin d'évaluer la présence de certains de ces impacts pour chacun des grands indicateurs. C'est une démarche qualitative coûteuse mais indispensable lorsqu'on a l'ambition de mesurer les impacts de la recherche dans plusieurs secteurs.

Le simplisme des classements atteint son paroxysme avec la publication annuelle des classements des universités, censés identifier les « meilleures » universités au niveau mondial.

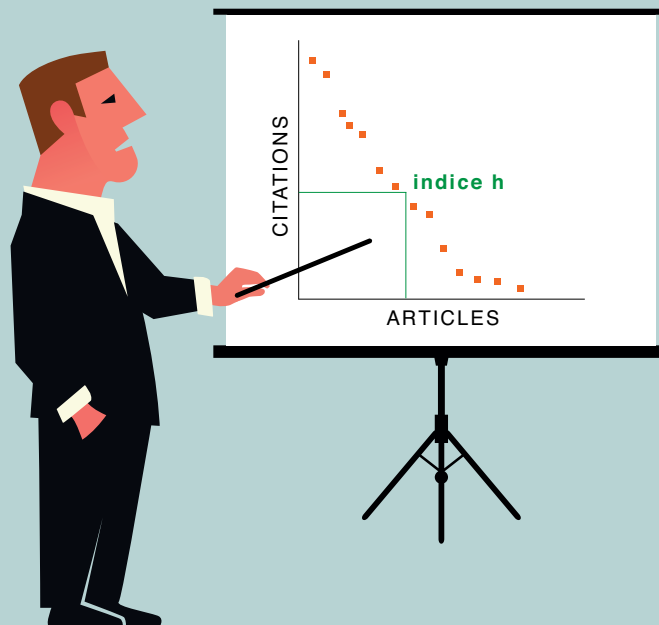
QUANTIFIÉ POUR CONTRÔLER

Les discussions animées entourant l'utilisation d'indicateurs bibliométriques dans l'évaluation des chercheurs laissent le plus souvent dans l'ombre un aspect pourtant fondamental de l'évaluation, à savoir le rôle de l'expertise des chercheurs dans le processus d'évaluation. La volonté de mieux contrôler le système très ancien d'évaluation par les pairs (*peer review*), qui repose sur une connaissance de première main du domaine de recherche du chercheur évalué, fait lentement place à l'idée d'évaluation par des experts (*expert review*) lesquels sont souvent externes au domaine de recherche considéré. L'évaluation quantitative normalisée facilite ce déplacement en fournissant des données soi-disant « objectives » qui peuvent alors être utilisées par n'importe qui. C'est dans ce contexte qu'il faut comprendre la création de classement des revues en A, B et C pour faciliter, sinon mécaniser, l'évaluation individuelle. Cela constitue *de facto* une forme de taylorisation de l'évaluation, une déqualification de l'expertise nécessaire à l'évaluation.

On est ainsi face à un paradoxe. L'évaluation d'un chercheur exige la constitution d'un comité de pairs qui connaissent bien le domaine. Ces experts savent déjà, par définition, quelles sont les bonnes revues dans leur domaine et n'ont pas besoin d'une liste préétablie par on ne sait quel groupe d'experts les classant en A, B et C. Par contre, ces classements permettent à des personnes ignorant tout d'un domaine de prétendre quand même porter un jugement autorisé. Mais alors ils ne devraient justement pas faire partie d'un comité d'évaluation ! La multiplication d'indicateurs mal construits sert donc en fait un processus de contournement de l'évaluation par les pairs, évaluation qui doit prendre en compte des indices de productivité, mais qui doit les interpréter dans le contexte spécifique de l'évaluation. Que certains chercheurs contribuent à la mise en place de ces classements, comme à l'utilisation d'indicateurs pourtant invalides, ne change rien au fait que ces méthodes ont pour effet de minimiser le rôle de l'évaluation qualitative de la recherche en la remplaçant par des évaluations mécaniques.

PSEUDO-INTERNATIONALISATION ET DÉCLIN DES RECHERCHES LOCALES

Un aspect peu discuté de l'importance accordée aux facteurs d'impact et au classement des revues est qu'elle détourne indirectement de l'étude de sujets locaux, marginaux ou peu à la mode. Cela est particulièrement dangereux dans les sciences humaines et sociales, dont les objets sont par nature plus locaux que ceux des sciences de la nature. Il va de soi que certains sujets sont moins « exportables ».



Le simplisme des classements atteint son paroxysme avec la publication annuelle des classements des universités, censés identifier les « meilleures » universités au niveau mondial.

Les revues les plus citées étant anglo-saxonnes (et non pas « internationales »), les chances d’y accéder dépendent de l’intérêt que ces revues portent aux objets étudiés. Un chercheur qui veut publier dans les revues les plus visibles a intérêt à étudier l’économie des États-Unis plutôt que les spécificités de la Banque du Canada ou l’économie régionale du Québec, sujet de peu d’intérêt pour une revue américaine. Le sociologue dont l’objet est « international », donc délocalisé, ou qui fait de la théorie a plus de chances d’exporter ses articles que celui qui propose l’étude empirique d’un aspect précis de sa propre société. Mais, si on souhaite étudier l’économie du nord de l’Ontario on risque aussi d’avoir plus de problèmes à « internationaliser » les résultats.

Or est-ce vraiment moins important de se pencher sur cet objet que d’étudier les variations du New York Stock Exchange? Il y a donc un danger réel que les objets locaux mais sociologiquement importants soient dévalorisés et donc, à terme, négligés si les indicateurs de citations sont utilisés mécaniquement sans que l’on tienne compte de l’intérêt social des objets de recherche en sciences humaines et sociales.

CONCLUSION : JUGER PLUTÔT QUE COMPTER

On entend souvent dire que ces classements sont inévitables et qu’il faut « vivre avec ». Cela est tout à fait faux. La résistance des chercheurs est tout à fait capable de bloquer de

tels projets malavisés. En Australie, notamment, la vive réaction des chercheurs au classement des revues a réussi à faire plier le gouvernement, qui a abandonné l’usage de ces classements pour l’évaluation de la recherche. En somme, le monde de la recherche n’a pas à céder devant des exigences qui n’ont rien de scientifique et appartiennent à des logiques qui lui sont étrangères. D’autant plus que ce sont en fait les revues francophones et les objets de recherche locaux mais très importants pour la société qui sortiront perdantes de ces dérives de l’évaluation. **AM**

Yves Gingras est professeur au département d’histoire et titulaire de la Chaire de recherche du Canada en histoire et sociologie des sciences de l’Université du Québec à Montréal.

NOTE :

Ce texte est une version plus courte d’un article intitulé « Dérives et effets pervers de l’évaluation quantitative de la recherche : sur les mauvais usages de la bibliométrie », paru dans la *Revue internationale PME* 28;2 (2015) : 7-14. Pour une analyse plus approfondie, voir: Yves Gingras, *Bibliometrics and Research Evaluation : Uses and Abuses*, Cambridge : MIT Press, 2016.

1. Yves Gingras, “How to boost your university up the rankings,” *University World News*, (2014) July 18;329, <http://www.universityworldnews.com/article.php?story=20140715142345754>. Voir aussi les nombreuses réactions dans *Science*, (2012), March 2;335: 1040-1042.
2. L Waltman and NJ van Eck, “The inconsistency of the h-index,” 2011, <http://arxiv.org/abs/1108.3901>.

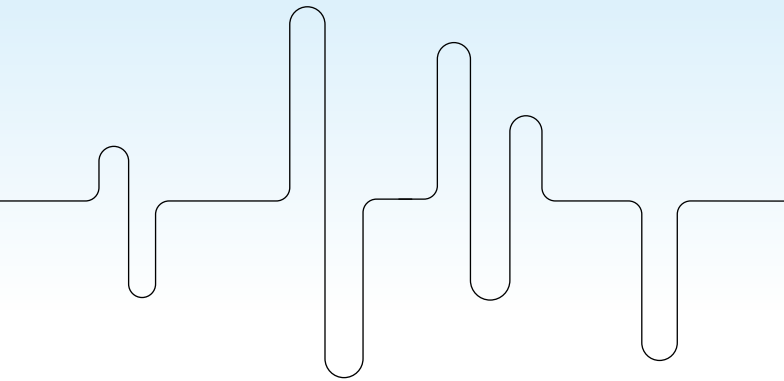
Collecting data *from students with students*

Ruth Childs



Gathering data on university students can provide important information about how they interact with the postsecondary education system, but it is also important to consult students to determine what data are collected and how.

La collecte de données sur la population étudiante des universités peut fournir d'importants renseignements sur la façon dont les étudiants interagissent avec le système d'éducation post-secondaire, mais il est également important de consulter les étudiants afin de déterminer quelles données sont recueillies, et comment.



If faculty, staff, and students have a common purpose in improving universities' programs, why don't we work together to develop better ways to collect data.

A few years ago, I was part of an admissions committee that developed a short, voluntary survey for one of our academic programs. The survey responses would not be part of admission decisions, but we hoped to determine if we were making effective changes both to the application process and to our outreach to communities facing barriers accessing educational opportunities.

I teach courses in survey development and measurement theory in which I emphasize the importance of checking that respondents understand the questions as intended and are able and willing to answer them. This can be done by recording a few respondents thinking aloud as they read the instructions and respond to the questions, and by testing the questions with a small sample of respondents. Even better, a small group of respondents might be involved throughout the design process to make sure that the questions asked are appropriate, the terms used are familiar, and the intended uses of the data are clearly explained and acceptable to respondents.

Did we involve students in developing our survey? I am embarrassed to admit that it didn't occur to us to do that. As faculty and staff who work daily with students, we were confident we knew how applicants would interpret the questions. And as a committee that focuses on equity in the admissions process, we were certain that applicants would believe our assurances about how we would and would not use the data.

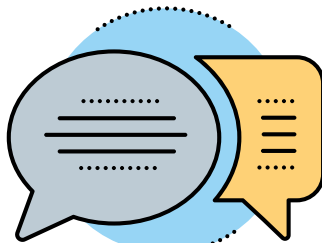
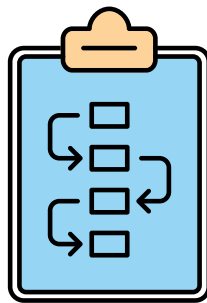
Only about half of the applicants responded to the survey—a response rate that would be enviable

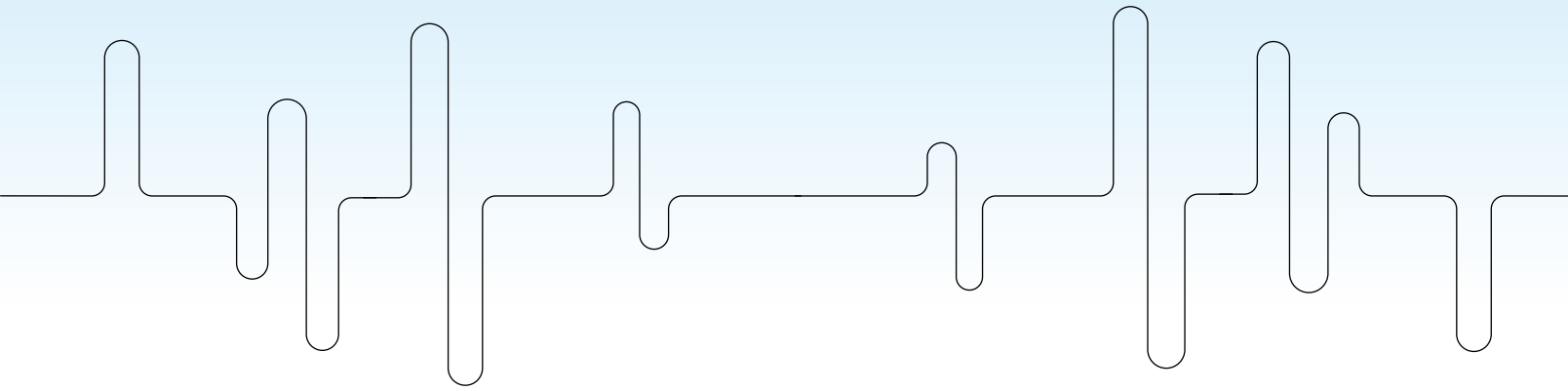
in much social science research, but was not what the admissions committee needed to evaluate the changes it was making.

Fortunately, this story has a happy ending: Within a couple of years, the response rate increased to more than 90 per cent. We were able to compare the demographics of applicants to the program with the demographics of the wider community and, when we made changes to the application process or to how we made admissions decisions, we were able to see who was affected. That program has since closed, but we are beginning to apply what we learned from that experience to other programs.

All of the credit for this happy ending goes to students. A Master's student who was interested in equity in education decided to make the survey the focus of her thesis research—not the results of the survey, but the survey itself. She went directly to the students who were currently in the program to find out what they thought of the survey.

She led discussions with groups of students and used an anonymous online survey to find out how individual students interpreted the questions and how they believed the responses were used. Based on what she learned, this student worked with the admissions committee to revise the survey's title, reorder and reword the questions, and rewrite the explanation of how the responses would and would not be used. Other students helped us analyze the data and, over time, suggested further revisions. The eventual success of the survey was due to their work.





Large-scale surveys, such as the National Survey of Student Engagement and the National College Health Assessment, can provide important data about students' identities, experiences, and perceptions, but there will always be a need to develop surveys for specific contexts. If faculty, staff, and students have a common purpose in improving universities' programs, why don't we work together to develop better ways to collect data from the students in those programs?

Time is one reason, I suspect. Even if our admissions committee had not been confident in its ability to develop a survey that students would want to answer, we had not allowed enough time to involve students. Finding students who are interested in being involved can be a lengthy process. Depending on how we want them to be involved, there will need to be time to mentor the students in survey development and include them in meetings to develop the questions, organize the collection and the analysis of initial test responses, and revise and retest the items if necessary.

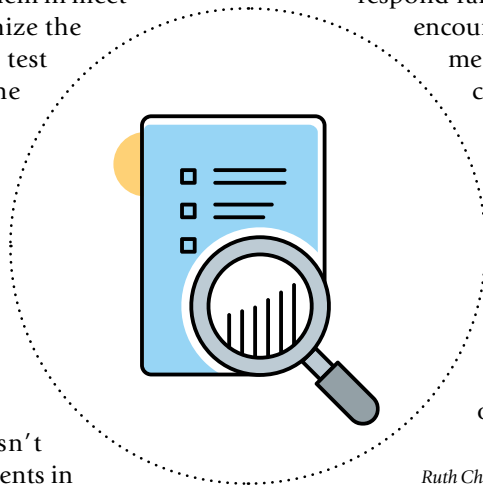
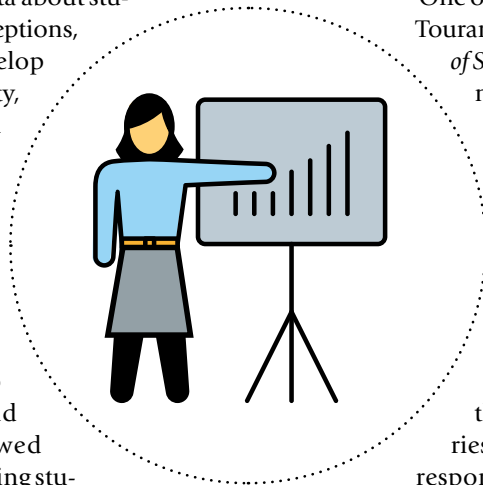
Money is another reason we often don't involve students. For our work, however, we have been fortunate to have access to a small amount of money to hire students. As well, some students have been interested in contributing to the development of surveys as a way to gain research experience.

I wonder, though, if there isn't another reason we don't involve students in developing these surveys: We believe we know how students think. Or perhaps we don't believe we know how

they think, but we believe that students won't mind making the effort to understand what we mean by the questions.

One of my favourite books on survey design is Tourangeau, Rips, and Rasinski's *The Psychology of Survey Response*. Based on research by cognitive psychologists and market researchers, the authors list 13 steps respondents might take when answering a survey question, beginning with "Attend to the questions and instructions" and including "Identify what information is being sought," "Retrieve specific and generic memories," "If the material retrieved is partial, make an estimate," and "Map the judgment onto the response categories." These steps assume, of course, that respondents want to provide as accurate an answer as possible. If respondents judge the questions to be unimportant or to require too much effort, however, they may choose not to respond or, worse, may respond randomly. The authors' findings are not encouraging: Reading the book always leaves me marvelling that anyone ever manages to collect useful survey data.

Nevertheless, surveys are the best tool we have for learning about students' identities, experiences, and perceptions. We need such data if we are to improve programs. We owe it to our students to create the best surveys we can so that the time and effort they spend responding to them is not wasted. That means collecting data not only *from* students, but *with* students. **AM**



Ruth Childs is Ontario Research Chair in Postsecondary Education Policy and Measurement at the Ontario Institute for Studies in Education, University of Toronto.

WAKING UP TO THE REALITY OF Canadian higher education

Claire Polster and Sarah Amsler



Higher-education systems in Canada and the United Kingdom share much in common, but there are important differences that faculty on both sides of the Atlantic should appreciate. The UK experience can wake Canadian academics up to the urgency of resisting university corporatization and to the opportunities for resistance that remain.

Les systèmes d'enseignement supérieur au Canada et en Grande-Bretagne ont beaucoup en commun, mais il existe des différences importantes que les professeurs des deux côtés de l'Atlantique devraient apprécier. L'expérience de la G.-B. peut sensibiliser les universitaires canadiens à l'urgence de résister à la corporatisation des universités et aux possibilités de résister qui subsistent.

Wake up! That is the call of mindfulness practice to ground us in the present moment, and help us see and use all available resources to meet the challenges that confront us. In this article, we urge academics to become more mindful of the harm that corporatization is causing to Canada's universities, and especially of the opportunities to resist it.

During the winter of 2017, while conducting interviews at universities across the UK, we realized that many academics—on both sides of the Atlantic—are “asleep,” albeit in different ways. Many British faculty have been so traumatized by the outrageous burdens, irrationalities, and indignities within their higher-education system that they can no longer see—or muster the resources to face—what is horrifyingly clear to an outsider: namely, how aberrant and destructive that system has become. Meanwhile, many Canadian faculty have grown so resigned to the corporate values and practices that have overtaken their universities that they no longer appreciate the numerous resources and opportunities they have to resist corporatization.

We believe that academics in both countries need to wake up to their respective realities, and that they can help one another to do this. Canadian academics can help their British colleagues recover the sense that they are experiencing a situation that is neither normal nor acceptable. In turn, British academics can help Canadian colleagues shake off their complacency or despondency by helping them appreciate the precious opportunities to resist corporatization they still have.

In this article, we use the British context as a backdrop to help bring sharper relief to those unrecognized or underappreciated features of Canadian higher education that make it possible for Canadians to resist—and possibly reverse—the corporatization of their universities. We begin by briefly describing some features of British universities, which provide a glimpse into the reality that awaits Canadian academics should they fail to seize these opportunities.

UK HIGHER EDUCATION TODAY

Many Canadian academics are aware of well-publicized developments in the British higher-education system, such as the establishment of the Research Excellence Framework (REF; formerly the Research Assessment Exercise) through which academics' research is ranked and their universities funded accordingly, and the replacement of most direct government grants to universities with student loans to cover the resulting tuition fees. They are also aware of some of the financial, personal, and interpersonal costs that the REF and other performance measures impose, and



British higher education is being progressively privatized, as for-profit higher-education providers undermine and displace public universities.

the ways that they, along with recent funding changes, entrench and advance corporate thinking and practice within British universities. However, what Canadians may not know, and what they may find shocking, is how severely many British academics are disciplined within their institutions and how thoroughly public practices, values, and interests are being excised from British higher education.

For example, unreasonable performance expectations are regularly imposed on British faculty, such as the requirement to meet annual “performance targets” for securing external research funding, despite the reality that almost two-thirds of that funding is awarded to an elite 12 per cent of the country's universities.¹ Tenure was effectively abolished in 1988, and now many faculty are not only relentlessly micro-managed and monitored by administrators, but routinely disciplined—often unjustly and with no recourse—for displeasing management in any number of ways.²

Further, British higher education is being progressively privatized, as for-profit higher-education providers undermine and displace public universities. The market increasingly shapes which academic courses are provided and at what cost, and higher-education institutions use public money to enrich shareholders at the expense of students and taxpayers. Meanwhile, the British government has passed draconian legislation that cripples the ability of academic unions to engage in “political” activities or to strike, and has authorized vicious attacks against citizens who demonstrate in defence of public-serving higher education.³

These developments are causing widespread and acute distress, particularly, but not exclusively, among British academics. They are experiencing alarming rates of physical and mental illness and are abandoning, in growing numbers, a cherished profession that has become intolerable.

It is difficult to convey how thoroughly irrational, destructive, and inhumane the British higher-education system seemed to the Canadian author of this article. The intense shock she felt acted as a very powerful wake-up call, not only to the necessity and urgency of resisting university corporatization in Canada, but also to the considerable space and opportunity that still exist to achieve this. In what follows, we briefly describe some of these possibilities for resistance. If Canadian academics and others approach these mindfully, free from any distracting memories of the past or assumptions about the future, they can gain a renewed sense of their ability and power to act, as well as a clearer idea of how they might do so.

EXPLOITING OUR ADVANTAGES

To bring some of the opportunities for resistance more clearly into view, it is helpful to compare the Canadian and British university systems and identify key advantages of the Canadian system. Three of these are closed shop unions (i.e., unions to which all employees *must* belong), tenure, and decentralization.

Closed shop unions, particularly in a context of less restrictive labour legislation, provide Canadian academics with more resources and freedom to organize and oppose corporatization both at the bargaining table and within broader society. This is because Canadian faculty associations do not need to spend substantial energy and resources recruiting and retaining members, as is the case for the UK’s University and College Union (UCU). Canadian faculty unions also face far less severe consequences should they alienate some union members, members of the public, and/or members of government when using the broader range of available means to oppose corporatization. (Stunningly, British legislation allows people who do not even belong to

the UCU to launch complaints against it, which can result in fines for the union. The British government also has a lot of leeway to define those “political” activities that unions are prohibited, by law, from undertaking.)

Along with strong faculty associations, the institution of tenure affords many (albeit too few) Canadian academics additional protections against the kinds of retaliation and reprisal routinely faced by British academics who dare oppose corporatization. These protections may not only support, but can also motivate resistance—if tenured Canadian academics realize how relatively little they risk in return for potentially substantial gains.

The decentralization of the Canadian higher-education system, with postsecondary education under the purview of the provinces, is a further advantage for academics seeking change. When public values and services are under attack, no government can single-handedly impose system-wide changes as dramatic and consequential as those the British government has recently put into place. Decentralization also makes those politicians responsible for higher education more accessible and accountable to Canadian citizens (at least in theory).

Most Canadian academics have not fully leveraged these advantages; however, when they are used, they can deal effective blows to the corporatization process that may reverberate throughout the postsecondary system. The recent successful strike in which the University of Manitoba Faculty Association (UMFA) prioritized opposing corporatization over traditional bread-and-butter issues is a case in point. In forgoing a pay increase and seeking instead to restrict managerial control and the use of performance metrics, UMFA not only prevented its own administration from advancing some of its damaging agenda, but provided an inspiring example that can embolden other Canadian faculty associations to do the same.⁴

WE DON’T WANT WHAT THEY’RE HAVING

Additional opportunities for resistance can be uncovered by identifying negative features of the British higher-education system that are less developed within Canadian universities. Primary among these is the infamous UK audit culture, which involves the use of various institutional and system-wide metrics to measure, rank, and reward—or punish—academics and universities. This culture has produced tremendous waste and dysfunction, as well as acute personal and social harm.

Because the audit culture is less developed and less coordinated in Canada, academics have considerably more autonomy in their work and their institutions than do many of their British counterparts. This autonomy can be used to



Canadian academics can deal effective blows to the corporatization process that may reverberate throughout the postsecondary system.

defend and preserve professional and public-serving values and practices as well as defy the corporate logic and processes being incorporated into Canada's universities.

Lower average tuition fees are another feature of Canadian higher education that helps limit corporatization. As was demonstrated in Quebec's "Maple Spring" student protests, low tuition fees (and, in the case of the province's CEGEP students, no tuition fees) can support the ability and willingness of students to organize and sustain anti-corporatization campaigns. Lower fees may also dissuade for-profit higher-education providers from entering the Canadian university marketplace, because they will not be able to effectively compete and make a profit. This is extremely important, for when large numbers of for-profits enter the higher-education sector, as they have in the UK, corporate practices and dynamics are entrenched and advanced within all higher-education institutions in a viciously circular way.

Finally, there is far less hierarchy among Canadian universities than there is in the UK, and there are fewer established university "clubs," such as the UK's Russell Group, that prioritize member universities' interests over all over others. These features of the Canadian system increase the potential to generate inter-university solidarity and united action to oppose corporatization, and restrict opportunities to employ "divide-and-conquer" tactics in order to advance or legitimize the process.

It is important to emphasize that Canada is not without a developed audit culture, unacceptably high tuition fees, or hierarchy and division within its university system, all of which need to be arrested and reversed. However, because these problems are less developed than they are elsewhere, Canadians are better able to resist them, particularly when one takes into account the advantages addressed above.

STRENGTHS TO BUILD ON

One final way to highlight Canadians' opportunities for resistance is to note positive features of British higher education that are equally if not more developed in Canada. Perhaps the most significant of these is the presence of public-serving bodies and institutions that can be used to challenge corporatizing policies in higher education and champion alternatives to them.

On the one hand, Canada has many formal organizations that produce solid critiques of corporatizing policies and develop credible and innovative alternatives to them. These include national and regional organizations that advocate on behalf of faculty, such as the Canadian Association of University Teachers and the Ontario Confederation of University Faculty Associations, and progressive think tanks that prioritize education issues, such as the Canadian Centre for Policy Alternatives.

Canada also has more robust channels and bodies through which progressive ideas can be promoted. There is more institutional democracy within Canadian universities than there is in most British universities, particularly those that have been established post-1992. There is also more opportunity to revitalize neglected collegial bodies (such as academic senates) and to expand academics' (and others') representation on governing bodies. These not only allow academics to better defend against corporatizing policies, but also to put forward and move forward alternatives to them.

Canada is also home to a relatively large number of established professional and labour organizations, confederations, and coalitions representing academics and other higher-education workers (such as the Federation for the Humanities and Social Sciences and the coalition of univer-

sity workers, students, and civil society organizations currently promoting the creation of a national postsecondary education act). These organizations' substantial resources and networks make it possible for them to connect with a wide range of individuals and groups—including their members, political representatives, and ordinary citizens—to generate opposition to corporate values and practices, and to build support for publicly oriented values and practices within and beyond Canadian higher education.

To be sure, many of these bodies and institutions could be further strengthened and their actions better coordinated. For this to happen, Canadian academics and others need to recognize the latent potential that exists within them, just as they must recognize the potential that exists in the other places and spaces that have been addressed here.

MOVING FORWARD

Most readers will not be surprised by anything we have said about Canada's higher-education system. Our aim is not to present anything new, so much as to encourage a more mindful or wakeful way of looking at what exists and what we can do with it.

Many forces currently conspire to keep Canadian academics and others from adopting this perspective. The corporatization process itself dulls awareness by rendering people ever more isolated, frantic, and insecure. Vision is also clouded by memories of past defeats and disillusion, as well as fears of failure that extinguish inspiration or initiative. However, as we have tried to demonstrate, Canadians can and should shake off their sleepy state and approach the present moment with greater clarity and focus. This will help them appreciate the many opportunities and resources that exist to challenge corporatization and to recover a sense of their power and energy to use them.

In closing, we emphasize that we are not calling for naive optimism nor suggesting that defeating corporatization is merely a matter of shifting perspective. On the contrary, we are promoting a more realistic appraisal of the challenges facing the Canadian higher-education system and the possibilities to respond to them. While resisting and reversing corporatization is no simple task, it is, we believe, an achievable and ever more urgent one. The nightmare being lived by British academics highlights both of these truths and offers an excellent reason and motivation to act. **AM**

Claire Polster is a Professor in the Department of Sociology and Social Studies at the University of Regina. Sarah Amsler is an Associate Professor in the School of Education at the University of Nottingham.



Canadians can and should shake off their sleepy state and approach the present moment with greater clarity and focus.

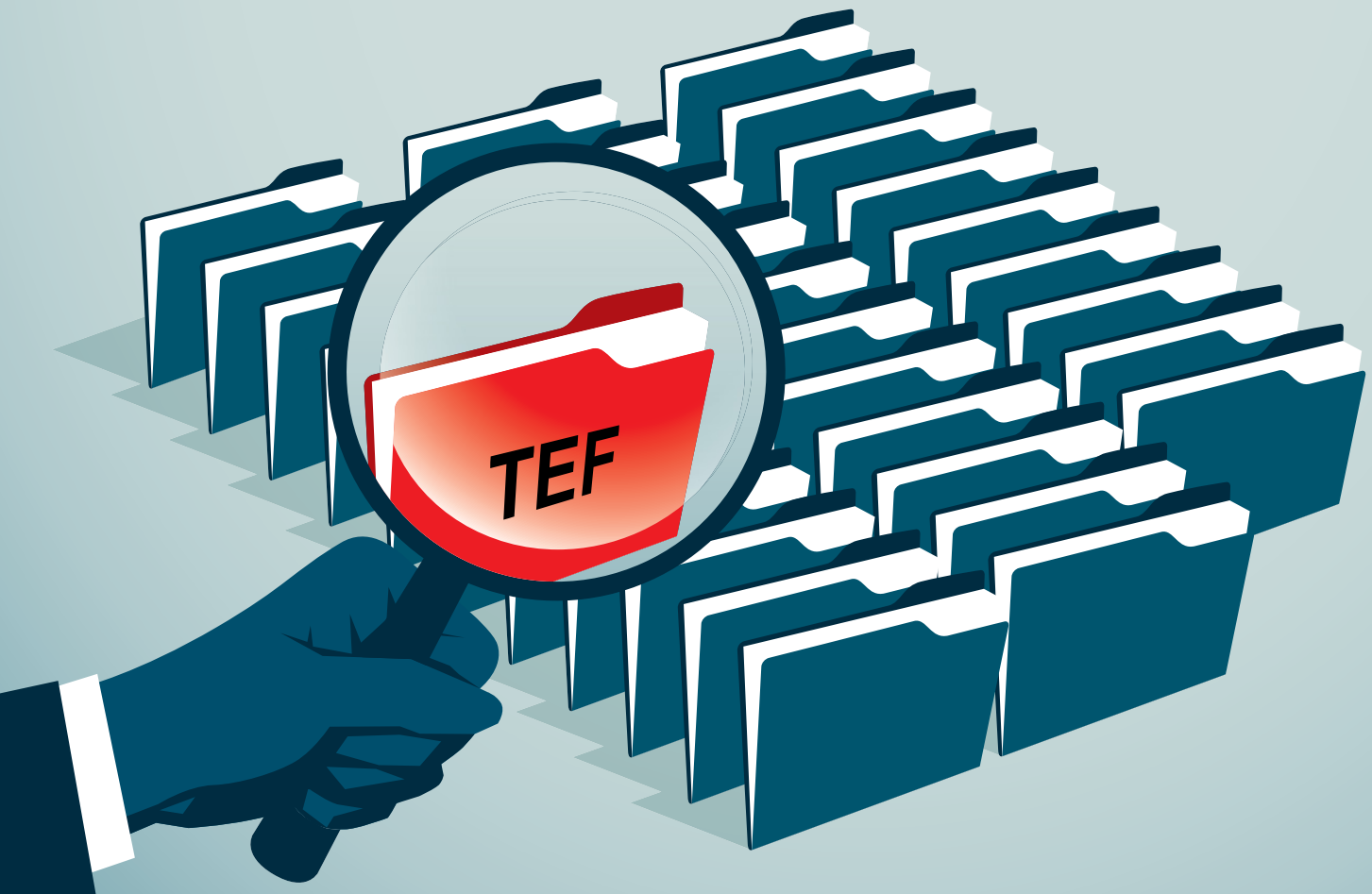
1. The Russell Group of Universities, *Submission of Evidence for the House of Commons Science and Technology Select Committee Inquiry: The Impact of Government Spending Cuts on Science and Scientific Research*, 2010 Jan., <https://www.russellgroup.ac.uk/media/5260/evidence-to-commons-science-and-technology-committee-inquiry-on-the-impact-of-funding-cuts-to-scientific-research.pdf>.
2. Liz Morrish, "Why the Audit Culture Made Me Quit," 2017, <http://www.timeshighereducation.com/features/why-audit-culture-made-me-quit>; for further analysis, see Morrish's blog, "Academic Irregularities," <https://academicirregularities.wordpress.com/>.
3. For elaboration on these issues, see Andrew McGettigan, *The Great University Gamble: Money, Markets and the Future of Higher Education*, London: Pluto Press, 2013; Clare Solomon and Tania Palmieri, *Springtime: The New Student Rebellions*, London: Verso, 2011.
4. Jen Hedler Phillis, "Beyond Bread and Butter," 2016, <https://www.jacobinmag.com/2016/12/university-manitoba-strike-union-umfa-bargaining>.


Understanding the United Kingdom's Teaching Excellence Framework and its implications

Rob Copeland

The UK's new metrics-based teaching evaluation framework is methodologically and politically flawed. What will this mean for the country's universities and faculty?

Le nouveau cadre d'évaluation de l'enseignement de la G.-B. fondé sur des mesures est méthodologiquement et politiquement défectueux. Quelles seront les retombées pour les universités et les professeurs de ce pays?





It is time for the UK's faculty unions to put forward alternative approaches for valuing and supporting teaching in higher education.

In 2017, a new higher-education assessment system—known as the Teaching Excellence Framework (TEF)—was launched in the United Kingdom. Based heavily on metrics, the TEF seeks to “recognize and reward excellence in teaching and learning, and help inform prospective students’ choices for higher education.” However, the TEF is both methodologically and politically flawed. It is time for the UK’s faculty unions to put forward alternative approaches for valuing and supporting teaching in higher education.

FROM REF TO TEF

The UK has often led the way in developing performance-based assessment systems in higher education. For example, in the mid-1980s, it was the first country to bring in a nationwide research evaluation process, which became known as the Research Assessment Exercise (RAE). Thirty years later, the RAE has morphed into the Research Excellence Framework (REF). As with its predecessor, the REF is based on a peer review process but also uses key metrics on research income, citations, and a qualitative assessment of research impact. The new TEF is a conscious attempt to ape the language and logic of the REF, including the link between awards and additional fee income. But the methodology is different: the TEF relies largely on institutional metrics, combined with a short written submission from universities, which is then evaluated by an “expert panel” and individually benchmarked against the types of students who attend each institution.

The origins of the TEF lie in the UK Conservative Party’s 2015 manifesto commitment to “introduce a framework to recognize universities offering the highest teaching quality.” This proposal was based on a perception that official “accountability” mechanisms and commercial university rankings were too focused on research outcomes and were therefore of limited value to potential students in choosing where and what to study. Government ministers had also picked up on a widely shared view that university teaching lacked the same status as research (ironically, due largely to the REF) and so believed that the answer lay in a REF for teaching. The Conservative’s surprise electoral victory in 2015 meant that the policy had to be implemented and, following a year-long technical consultation exercise, the TEF came into being.

KEY ELEMENTS OF THE TEF

In essence, the TEF is an official process for measuring the undergraduate student experience in higher-education institutions. The TEF panel—comprising academics, students, and employers—considers evidence from a set of metrics using national data as well as a written statement submitted by the institution. The metrics cover retention rates, student satisfaction, and employment outcomes. These data are then benchmarked to take account of differences in students’ characteristics, entry qualifications, and subjects studied.

Therefore, unlike other rankings and evaluations, the TEF provides a judgment of relative, rather than absolute, performance through its data-benchmarking process. This means that elite, well-funded, “research-intensive” universities are not compared directly with newer “access-oriented” institutions on key indicators such as dropout rates. In some ways, this makes the TEF a fairer measure of performance than commercial university rankings are, but it also makes it more difficult to demonstrate student outcomes. Ironically, the new process also makes it much harder for the government to claim that the TEF will aid student choice, especially because the TEF is currently an institutional award that tells students nothing about subject-level provision.

Participation in the TEF is also a voluntary process, and consequences vary for the different nations of the UK. This reflects the increasingly different funding and regulatory systems within the country. For example, in England, partici-

pation in the TEF is linked to an ability to increase tuition fees, whereas in Scotland, which currently has no tuition fees for Scottish students, participation in the TEF is linked to potential reputational advantage. Given the absence of additional financial benefits, it is no surprise that the majority of Scottish universities opted out of the latest TEF.

In June, the TEF panel published the results of its first assessment. Using an Olympics-style classification system (gold, silver, and bronze), half of the participants were awarded silver, 26 per cent obtained a gold award, and 24 per cent received a bronze. In terms of institutional classifications, the results defy easy schematization. For example, a number of access-oriented universities achieved the highest award, while a few research-intensive universities were awarded a bronze. In fact, a disproportionate amount of press attention focused on the fact that the prestigious London School of Economics achieved the lowest grade, although the overall results were less disruptive of traditional university hierarchies than many predicted.

FACULTY VOICES

Irrespective of the results, the UK's University and College Union (UCU) has consistently opposed the TEF, both on methodological grounds and because of its potential impact on institutions, staff, and students. This opposition is based on a number of factors. First, the core metrics—student satisfaction expressed through the National Student Survey (NSS), retention rates, and graduate outcomes—are flawed for the purposes of assessing teaching quality. To a significant extent, these metrics are influenced by external factors such as social background, gender, and, in terms of jobs, the region in which the university or college is located. Above all, they are

poor proxies for measuring teaching excellence; indeed, even the chair of the TEF panel admitted that this was the case for student satisfaction scores.

Rather than focusing on improving teaching practice *per se*, universities are likely to concentrate on targeting better survey results, higher completion rates, and graduate outcomes. That is the nature of metrics and quantitative measures when they end up becoming targets. But this, in turn, can have detrimental consequences for the composition of the student body. For example, some institutions are already talking about increasing their student-entry requirements and cutting student numbers on specific courses in a bid to reduce their dropout rates. Another serious concern is that universities may seek to improve their rating on graduate outcomes by altering their subject mix, such as moving away from creative arts courses, which score lower on short-term labour market outcomes.

As a faculty union, the UCU knows that the TEF has already been cited as a reason for job cuts by some universities, and we are concerned that other institutions may follow suit. There is also a legitimate concern that, alongside the REF, the TEF will lead to a further fragmentation of academic roles into teaching-only and research-only positions.

Despite the preoccupation with the choice of metrics, the most controversial part of the TEF is its link to tuition fee increases. England already has the highest public tuition fees in the industrialized world and the TEF allows institutions in England to increase these further (by the rate of inflation). The government's ultimate objective is to allow for fee differentiation on the basis of TEF results, although these plans have been delayed until 2020 at the earliest. However, the extent to which the government will be able to deliver on this fee-differentiation agenda remains open to question. The Labour Party's pledge to abolish tuition fees during the 2016 general election was instrumental in increasing its popularity amongst younger voters. As a result, fees have become an increasingly toxic issue for the Conservative Party.

REACTIONS FROM THE SECTOR

Because of the TEF's role in increasing tuition fee levels, the National Union of Students (NUS), along with the UCU, has also opposed key aspects of the TEF. In an explicit attempt to undermine the process, it urged its local student unions to support a boycott of the NSS. The UCU supported this initiative and it had some effect at universities such as Manchester, Bristol, and Cambridge, where NSS returns dropped below the required threshold. Partly for this reason,



the government has said that future TEFs will rely less heavily on NSS data than other metrics, although this decision also reflects pressure from the larger research-intensive universities, who tend to score poorly on student satisfaction.

Unfortunately, the response from university leaders has been largely one of self-interest. Although critical of aspects of the TEF, they were willing to go along with the system as a means of increasing institutional revenue (via higher fees). Those universities who struck gold were the first to pop open the champagne, while those who secured bronze were more likely to appeal the decision or dismiss the exercise as “meaningless.” These responses reflect one of the structural problems of the UK higher-education sector: namely, the division of the sector into competing groups who lobby on behalf of their own type of institution (e.g., research-intensive, business-focused, access-oriented, etc.), sometimes at the expense of the system as a whole. In many ways, both the TEF and the REF reflect and reinforce these divisions *within* the higher-education sector.

The shift to a more market-oriented model was also evident in the passing of a new *Higher Education and Research Act*. The legislation establishes a new regulatory body in England—the Office for Students—that will be responsible for overseeing future developments of the TEF and encouraging greater competition between providers. For example, the new law makes it easier for private providers to obtain degree-awarding powers and the university title, and access the student-loan system. In the interests of ensuring academic quality and protecting the public purse, the UCU has major concerns about the further entry of private, for-profit providers into the English system. The quicker and easier it is to become a university and award degrees, the more vulnerable the higher-education sector becomes to for-profit organizations pursuing financial gain rather than providing high-quality educational experiences. The UCU has consistently warned that the UK government is opening the floodgates to a repetition of US-type scandals involving for-

profit providers. As a union, we will continue to call for additional regulation for these types of risky providers.

NEXT STEPS

What is the specific future for the TEF in this new regulatory environment? First, after a cursory lessons-learned exercise, the government announced that it will be making “no changes to the overall structure or methodology of the TEF.” However, in future exercises there will be a reduced role for benchmarking and student surveys in the evaluation process—changes that are likely to benefit research-intensive universities. New supplementary metrics on both graduate earnings and grade inflation have also been proposed by the government. The former reinforces the controversial link between the TEF and labour market outcomes, while the latter seems more of a populist response to media concerns about the growing proportion of students achieving top degree classifications. Despite their different origins, both proposals highlight the politicized nature of metrics in higher-education policy-making.

Second, the government remains committed to introducing a subject-level TEF exercise in the future. A series of pilots will begin later this year. Although no ratings will be attached to the pilots, the whole process is likely to be complex and contentious. If introduced as part of a full subject-based TEF, the impact on faculty will be considerable, particularly in terms of increased workload and student expectations. Whatever the outcomes, the UCU will continue to challenge the use of the TEF as a crude performance-management tool.

SUPPORT AND RECOGNITION FOR TEACHING

Finally, despite the obvious flaws with the TEF, there is a need for a proper discussion about how universities can better support and recognize teaching in higher education. There is a key role for faculty unions in this process and an opportunity to build broader alliances with student organizations. The UCU has long argued that faculty working conditions are effectively student learning conditions. In addition to campaigns on public funding, better job security; reduced workloads; access to appropriate training, support, and professional development; and fairer promotion criteria need to be at the heart of the debate on good-quality teaching. **AM**

Rob Copeland is a Policy Officer for the University and College Union, which represents over 100,000 faculty and staff at universities and colleges across the UK.



Academic Matters is ONE CLICK AWAY



BLOGS
JOB LISTINGS
FEATURES
COMMUNITY
EVENTS
RESOURCES

Academic Matters, an engaging journal of commentary, analysis, and research on trends in higher education, is available online. The website contains features from the popular print magazine, web-exclusive articles, blogs, and many other resources of interest to academics and the higher education community. Access is always free. Visit us today!

www.academicmatters.ca

WIN
a trip for 2

to
Italy

valued at over \$5,500!

Enter NOW
OTIP.com/Dream



OTIP is a proud
partner of
Edvantage

Sarracini
Travel
The travel people.™

OMEGA
TOURS

Restrictions apply. Win an eight day trip for two to Italy from Sarracini Travel and Omega Tours valued at over \$5,500 CAD. Go to otip.com/dream to enter the OTIP "Taste of Italy" Contest. No purchase necessary. Open to all active and retired education members, Ontario residents, 18 years of age or older. Odds of winning depend on entries received by closing date. Winner must correctly answer skill-testing question. Contest closes August 31, 2018. Full rules at otip.com/contest-rules.