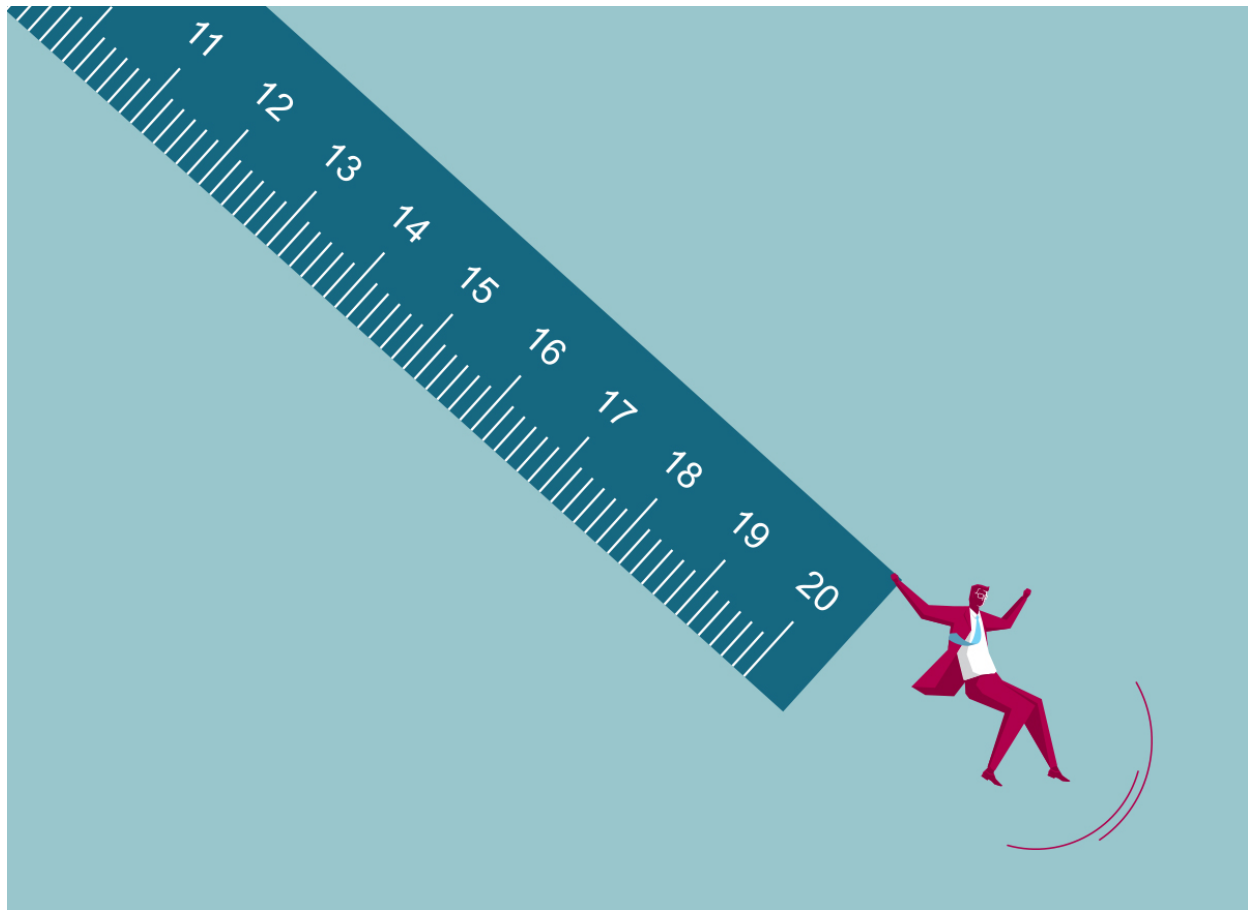


Unintended consequences: The use of metrics in higher education

AM academicmatters.ca/2017/11/unintended-consequences-the-use-of-metrics-in-higher-education/

By: Gavin Moodie | From the Winter 2017 Issue

11/29/2017



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.

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.

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.

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

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.

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

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 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 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 performance 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. AM

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.