

Our Misguided Trust in Quantitative Measurement

August 13, 2018

by Michael Edesess

The belief that detailed quantitative measurement will make performance easier to evaluate, manage efficiently, and improve has survived repeated failures of the doctrine. In fact, the failures serve only to bring forth calls for *more* of the treatment. Only very rarely is it admitted that quantification doesn't work and should be scrapped.

Quantitative measurement has gripped the management of everything in the last 60 years. It is now so firmly entrenched that nothing could shake it loose. Going back to the old methods of subjective, non-quantitative evaluation – even in small ways – is virtually unthinkable. For that reason, history professor Jerry Z. Muller, in his very important book *The Tyranny of Metrics*, says “Because belief in its efficacy seems to outlast evidence that it frequently doesn't work, metric fixation has elements of a cult.”

Throughout society, the reliance on fixed beliefs has distorted outcomes and misallocated resources. But the best example of how this type of overreliance led to a horrendous result occurred over 60 years ago, among the followers of an obscure cult.

In a 2010 Wired article, writer Jonah Lehrer retold psychologist Leon Festinger's account of what happened after a cult whose leader predicted the end of the world on December 20, 1954, realized that it hadn't occurred. Festinger was the originator of the concept of cognitive dissonance. He was curious what would transpire when the cult discovered that the belief they had committed their lives to turned out to be false. He infiltrated the cult in order to observe.

The cult leader, Dorothy Martin, to whom Festinger gave the pseudonym Marion Keech, had been in constant contact with aliens of the planet Clarion. She told her cult members that these aliens would pick up her cult in a flying saucer at 12:01 AM on that date just before the world was destroyed by massive flooding. In Lehrer's retelling:

On the night of December 20, Keech's followers gathered in her home and waited for instructions from the aliens. Midnight approached. *When the clock read 12:01 and there were still no aliens, the cultists began to worry. A few began to cry. The aliens had let them down.* But then Keech received a new telegram from outer space, which she quickly transcribed on her notepad. *"This little group sitting all night long had spread so much light," the aliens told her, "that god saved the world from destruction. Not since the beginning of time upon this Earth has there been such a force of Good and light as now floods this room."* In other words, it was their stubborn faith that had prevented the apocalypse. Although Keech's predictions had been falsified, the group was now more convinced than ever that the aliens were real. They began proselytizing to others, sending out press releases and recruiting new believers. This is how they reacted to the dissonance of being wrong: by becoming even more certain that they were right.

This analogy may seem extreme, but it is not.

Imposing a strict system of quantified metrics to evaluate and reward performance has serious deleterious unintended consequences. It induces gaming of the system, a kind of rent-seeking behavior that adds nothing to productivity and often detracts. It siphons attention toward goals whose achievement can be measured and away from goals whose achievement is difficult or impossible to measure but may be of greater importance. It brings forth a large bureaucracy of data-recording and analysis, performance measurement and evaluation, thus undermining one of the main goals of quantitative measurement, fiscal efficiency. It poisons employees' desire to do their job by substituting external and often arbitrary-seeming requirements for internal motivation.

A litany of failures

Muller runs through a list of situations in which quantitative metrics have been applied to performance measurement and evaluation, and documents what has gone wrong in each of them. He reviews their applications to colleges and universities, K-12 education, medicine and hospitals, policing, the military,

business and finance, and philanthropy and foreign aid.

In each case efforts to hold individuals and institutions accountable by imposing quantitative metrics to measure and reward performance resulted in consequences that were worse than the problems they were intended to solve.

In higher education, efforts to evaluate departments and rank universities using quantitative metrics led to assessing individual and departmental quality by the number of citations their journal articles received. This led both to gaming and to a reduction in the quality of written material while its quantity increased.

Some professors in the same field formed informal citation circles, in which when any of the professors published an article she would cite numerous articles published by the others. Some journals, to increase their "impact factor" - the number of times they were cited - even requested submitters of articles to add citations referencing their own journal.

Along with the need for reporting to meet equal opportunity requirements, this led to the ballooning of administrative staff at universities - many of whom are paid more than the professors whose research is supposed to be the core product of the university.

In K-12 education, government programs such as No Child Left Behind (NCLB) led to extensive gaming and even cheating. Teachers' schools' funding, and their own employment, was dependent on their students' performance on standardized, government-mandated math and English exams. Hence, many teachers spent much of their class time "teaching to the test." This moved the emphasis toward test-taking and away from arguably more important activities with unmeasurable results, such as cultivating students' capacity for intellectual curiosity, good behaviour, and creative thought and innovation.

It was later discovered that under great pressure, many teachers cheated by changing the answers to students' test questions from incorrect to correct. Some schools engaged in a practice called "creaming" by redefining underperforming students' status in such a way that their test scores would not be counted.

And yet, when programs like NCLB were evaluated by third-party consultants, it was found that they hadn't even added appreciably to students' test scores.

In medicine, government-mandated metrics resulted in extensive creaming. A government requirement that hospitals report readmission rates - on the assumption that if a patient was readmitted soon after being treated and released, the treatment had been ineffective - resulted in hospitals redefining those who returned a short time after being released as out-patient or emergency room cases, causing them to be left out of the count. A requirement to report the rate of deaths from surgeries resulted in surgeons accepting only easy cases and declining to treat difficult ones, for fear that the patients might die and degrade their statistics.

In policing, to show reduction in their crime rates, many police departments defined crimes down to a lower level - a felony becoming a misdemeanor for example - thus making it look like crime had decreased.

In the military, the Vietnam War example that should have killed metricization once and for all was Department of Defense Secretary Robert McNamara's insistence that progress in the war should be measured by the "kill ratio," the ratio of dead bodies of Vietnamese enemy soldiers to those of American and allied troops. This resulted in widespread lying about the kill ratio, but also led to American soldiers exposing themselves to hazardous conditions by searching the battlefield for dead bodies of Vietnamese enemies to add to the count.

Says Muller, 'McNamara's Pentagon was characterized by what the military strategist Edward Luttwak called "the wholesale substitution of civilian mathematical analysis for military expertise. The new breed of the 'systems analysts' introduced new standards of intellectual discipline and greatly improved bookkeeping methods, but also a trained incapacity to understand the most important aspects of military power, which happen to be nonmeasurable.'"

In business, Muller chronicles two incidents in which the imposition of external quantitative measures, meant to enhance performance, actually created conditions that destroyed the reputation of the corporation. In one, the case of Mylan pharmaceuticals, the top executives were highly incentivized to increase corporate profits. Finding themselves with a widely-used product for which there were few competitors - the EpiPen, which injects epinephrine into allergic individuals to counteract severe allergic shock - Mylan executives devised a strategy to increase sales and increased the price of the product by 500% over a five-year period. The result was that the executives were paid more over that period than the executives of any other pharmaceutical company. But when attention was called to their practices resulting in congressional hearings and Justice Department investigations, Mylan's stock price dropped by half and

its reputation was severely damaged.

In the well-known case of Wells Fargo, management imposed quotas on employees for “cross-selling” – the signing up of banking customers for additional (more profitable) services like overdraft coverage and credit cards. Under pressure to perform, like the K-12 teachers who cheated by changing students’ answers on the tests, many Wells Fargo employees signed clients up for these additional services without their knowledge or consent. Wells Fargo fired 5,300 employees when it discovered the fraud, but as Muller says, “the spate of fraud was a predictable response to the performance quotas that the company’s managers had set for their employees.” The result was disastrous for the company, which instead of enhancing its profits suffered a decline in its reputation from which it will be difficult to recover.

In finance, the gaming of a particular set of quantitative requirements – the Basel rules – was a principal cause of the financial crisis. The rules were designed to require that banks held sufficient capital, but the requirement was “risk-adjusted” depending on the investments that the bank held. Because “diversified” investments such as collateralized debt obligations (CDOs) were deemed less risky than individual mortgages and thus required less reserve capital, banks sold their mortgages to packagers like Merrill Lynch and Goldman Sachs and then purchased the CDOs they created. It was this gaming of the rules for risk-adjusting reserve capital that led to the downgrading of mortgage quality, and the explosion of CDOs in the early 2000s.

For philanthropic organizations outcomes are hard to measure, so the emphasis is on measuring inputs. For example, a lot of attention is given to what percentage of the budget is spent on administration, including fundraising. It is assumed that the lower that percentage is, the better. “In response,” Muller says, “the leaders of charitable organizations often end up trying to game the figures: by reporting that the time of leading staff members is devoted almost entirely to programs, or that there is no spending on fundraising. That response is understandable. But it feeds the expectations of funders that low overhead is the measure they should be looking at to hold charities accountable. Thus the snake of accountability eats its own tail.”

“The snake of accountability eats its own tail” could be shorthand for a lot of the ill effects caused by excessive quantitative measurement. As Muller explains, part of the reason for the increase in metricization is that employees and executives are not trusted. Therefore, quantitative requirements are imposed on their behavior to keep them in line. In response, they behave in precisely those ways that the quantification model expects them to behave. And by gaming the quantification system and cheating, they make it appear that the measurement system is working. And yet, the ultimately desired results are not improved and are often even made worse.

Bullst Jobs¹**

I participate regularly in a book club. We recently had to decide what we should read and discuss next. One participant suggested the book *Bulls**t Jobs* by David Graeber. I had not read that book yet but I had read *The Tyranny of Metrics*. I suggested *Metrics* and said that I think it subsumes *Bulls**t Jobs*.

Having now read *Bulls**t Jobs* I think indeed I was right. Having read Graeber’s book, which is written from a peculiar viewpoint but is spot-on in many ways and often funny, I believe that the reason that most of those bulls**t jobs exist is because of the cult of metrics. Many of those jobs involve bean-counting, reporting, form-filling-out, and the like in seemingly senseless ways (If they even involve any work at all, which many of them, oddly enough, don’t).

*Bulls**t Jobs* is in large part an account of the explosion of nonsensical administrative tasks that has been the result of the obsession with creating metrics to assess, control and reward – or penalize – good or bad performance.

This leads me to wonder whether the cult of metrics and the bulls**t jobs it creates is somehow a naturally-occurring spread-the-wealth phenomenon in an age when the bulk of profits is actually created by a very few. Perhaps if it weren’t for bulls**t jobs – created and facilitated by the obsession with quantitative metrics and their recording in databases, analysis, and reporting – income inequality would be even worse than it is. When we speak of growth of the service industry that has replaced much of the decline in bricks and mortar industries, how much of that “service” is actually devoted to the pointless and unproductive occupation with metrics, in a whole gigantic industry of bulls**t jobs?

Metrics obsession and bullst jobs in investment finance**

We don’t have to look far to see the effect of the tyranny of metrics and the bulls**t jobs it creates in the finance industry, especially investment management and advice. Analysts spew out reams of numbers, generated by computer programs from databases containing terabytes of financial data, and embodied in endless graphs and tables. But no overall improvement in the achievement of the goals of investment

management can be discerned.

Investment management could be the poster child for the entire cult of metrics phenomenon, but it creates jobs and spreads the wealth of those who actually created the wealth – not among the many but among an almost-as-wealthy and sometimes even-more-wealthy sub-tier.

Economist and mathematician Michael Edesess is adjunct associate professor and visiting faculty at the Hong Kong University of Science and Technology, chief investment strategist of Compendium Finance, adviser to mobile financial planning software company Plynty, and a research associate of the Edhec-Risk Institute. In 2007, he authored a book about the investment services industry titled The Big Investment Lie, published by Berrett-Koehler. His new book, The Three Simple Rules of Investing, co-authored with Kwok L. Tsui, Carol Fabbri and George Peacock, was published by Berrett-Koehler in June 2014.

¹ The actual title of this book has the asterisks replaced by the letters “h” and “i.”