

2018

Numbers Games: Review of *The Tyranny of Metrics* by Jerry Z. Muller (2018)

Joel Best

University of Delaware, joelbest@udel.edu

Follow this and additional works at: <https://scholarcommons.usf.edu/numeracy>



Part of the [Sociology Commons](#)

Recommended Citation

Best, Joel. "Numbers Games: Review of *The Tyranny of Metrics* by Jerry Z. Muller (2018)." *Numeracy* 11, Iss. 2 (2018): Article 13. DOI: <https://doi.org/10.5038/1936-4660.11.2.13>

Authors retain copyright of their material under a [Creative Commons Non-Commercial Attribution 4.0 License](#).

Numbers Games: Review of *The Tyranny of Metrics* by Jerry Z. Muller (2018)

Abstract

Jerry Z. Muller. 2018. *The Tyranny of Metrics* (Princeton, NJ: Princeton University Press). 220 pp. ISBN 978-0-691-17495-2 (also available as an e-book).

The historian Jerry Z. Muller argues that quantitative metrics to assess performance have spread throughout many institutions. This trend poses problems when people begin to game the numbers, to focus on achieving good scores rather than on meeting the institution's goals.

Keywords

accountability, metrics, numbers games

Creative Commons License



This work is licensed under a [Creative Commons Attribution-Noncommercial 4.0 License](https://creativecommons.org/licenses/by-nc/4.0/)

Cover Page Footnote

Joel Best is a professor of sociology and criminal justice at the University of Delaware. His books include *Damn Lies and Statistics* (University of California Press, 2001), *More Damned Lies and Statistics* (University of California Press, 2004), *Flavor of the Month: Why Smart People Fall for Fads* (University of California Press, 2006), *Stat-Spotting: A Field Guide to Identifying Dubious Data* (University of California Press, 2008), *The Stupidity Epidemic: Worrying about Students, Schools and America's Future* (Routledge, 2011), and *American Nightmares: Social Problems in an Anxious World* (University of California Press, 2018). His five papers in *Numeracy* include a perspective ("Birds—Dead and Deadly: Why Numeracy Needs to Address Social Construction") in the journal's first issue (Jan. 2008).

Years ago, when I lived in Central California, I taught some evening criminology classes for sheriff's deputies in a nearby county. The sessions usually started with me lecturing for about twenty minutes, then my students would start telling me how things worked in the real world. I like to think we all learned something—certainly I did. One time, I was talking about the problems in measuring crime rates, and they told me a story. A few years earlier, a semi-trailer truck had been stolen a couple of counties away. This truck was a big-ticket item, a six-figure theft. County sheriffs reported departmental statistics to their respective counties' board of supervisors, and they usually reported the total value of the goods stolen in the county the previous year, but they also reported the value of the goods recovered. Law enforcement could not prevent all thefts, but they could be considered successful if they recovered much of the stolen property, so a report showing that a high proportion of stolen goods had been recovered was taken to mean that the sheriff's department was doing a good job. The stolen truck was recovered by the deputies from the county where I was teaching, and their sheriff's department was delighted because while they did not have to count the theft, they could count the high-value recovery, which would make their department look pretty good. The sheriff in the county where the theft had occurred had to report the stolen truck's value in his report, but he also reported the value of the recovery. In other words, the truck was reported stolen in one county, but recovered in two. As one of the deputies explained to me: "It's just a numbers game."

This story's particulars may be unfamiliar, but most of us know about—and probably have personal experience with—analogue numbers games. Professors demonstrate the quality of their work by reporting numbers of publications and presentations, the dollar value of grants received, and scores on teaching evaluations. There are increasing pressures on public school teachers and principals to demonstrate that their students attain high standardized test scores. And, as Jerry Z. Muller demonstrates, such quantitative evaluation is common in many institutions, including medicine, business, law enforcement, the military, and so on. Perhaps the participants in mom-and-pop stores and other small operations are familiar enough with one another's performances to judge whether everyone is holding up their end, but once the workplace grows to include dozens, hundreds, or thousands of workers, reliance on quantitative measures to evaluate and manage organizations seems inevitable. Quantification seems to promise a way to ensure accountability. Organizations need performance metrics to identify problems, or at least make sure that things are going well.

Muller found his subject after serving as chair of the history department at his university while it was adopting "assessment" (an institutional fad that spread through much of higher education over the past 25 years): "I found my time increasingly devoted to answering queries for more and more statistical information about the activities of the department There were new scales for evaluating

the achievements of our graduating majors—scales that added no useful insights to our previous measuring instrument, namely grades” (p. 10). Muller began gaming the assessment metric, which led him to think about the ways analogous numbers games are played in other institutions.

Performance metrics invite people to game the system. When people know that they will be evaluated by some numeric score, they will be encouraged to perform in ways that will produce better scores. In part, this tendency may be expected and even considered desirable by the people devising the metric: employers may seek to increase workers’ productivity by rewarding those who produce more, and they presumably hope workers will become more productive in order to receive those rewards. But problems emerge when people start gaming the system, when they strive to attain high scores without necessarily fulfilling the goals the metrics are designed to foster.

As Muller notes, there are various aspects to this problem. There are overt cases of fraud that, if they become public, can create scandals: police officers failing to record some crime reports correctly in order to keep the local crime rate low; educators altering standardized-test answer sheets as a way of demonstrating that their students are showing improvement; scientists who falsify research results in hopes of getting their funding renewed; and so on. But gaming need not involve outright fraud. Often, there is no direct way of determining whether an organization is meeting its goals. How can military commanders know whether their guerilla opponents are gaining or losing strength? How can drug enforcement officials tell whether drug use is rising or falling? There may be no way to accurately measure whether these goals are being met, which creates a temptation to substitute metrics of things that can be counted, such as measures of organizational activity. Thus, we are asked to believe that progress in the war on drugs can be measured by the number drug-related arrests, or the value of the drugs seized, while generals in the Vietnam War reported high body counts as proof that the enemy was being defeated. But, of course, it isn’t clear what such metrics measure—are more drugs being seized because officials are effectively reducing drug imports, or because the drug trade is expanding and importing more drugs? And, there are other ways these numbers can be gamed. When drug shipments are seized, officials often report the eventual “street value” of the drugs—a figure likely to be much higher than what the drugs had cost the people who had them when the seizure occurred. And activity measures can be falsified. There is considerable evidence that officers in the chain of command in Vietnam were under considerable pressure to report high body counts, which led to reported enemy casualties being inflated and reinflated as they passed up the chain of command.

Of course, the impulse behind metrics is often admirable. People want to have accurate understandings of how organizations operate so that they can know what’s working well, and what ought to be improved. Quantitative measures of

performance seem to be the best way to collect that knowledge. But Muller notes that problems occur when these performance scores become high-stakes, when scores become the basis for rewarding those with good scores and punishing those with lower scores. In such cases, there are powerful incentives for gaming: fudging one's results, presenting the data in the most favorable light, teaching to the test, and so on. This is when the metric stops being a useful tool and becomes a tyrant.

The Tyranny of Metrics should interest students of numeracy. Much of the numeracy literature assumes that our problems lie in people's difficulties in performing calculations properly or in understanding what numbers represent. But Muller's point is not that people aren't counting competently, but rather that they are often encouraged to count the wrong things for the wrong reasons, and this in turn may lead them to produce numbers that may be properly calculated, yet are nonetheless inaccurate or meaningless. *The Tyranny of Metrics* reminds us that thinking critically about numbers is about more than getting the sums right.