

MEDICINE AND SOCIETY

Performance Anxiety — What Can Health Care Learn from K–12 Education?

Debra Malina, Ph.D.

When you entered this field, you considered it a calling. You had to master both an art and a science — you aimed to effect critical changes in other people, who were infinitely variable and over whom you had limited influence, but if you established relationships with them, you and they often triumphed together. Nowadays, you're increasingly assessed on the basis of how well those people, now considered your customers, do on a few narrow tests. Although you see this as inapt quantification that breeds constricting standardization, society demands services of more consistent quality. So policymakers, applying business principles to your field, insist on measuring your performance in whatever ways it is easily measurable and then rewarding or punishing you accordingly.

Though this scenario may sound familiar, the “you” here is not a physician but a U.S. schoolteacher. Elementary and high-school education, like health care, has been in the throes of an “accountability movement” whose impetus may be honorable but whose product is widely considered misguided, shortsighted, counterproductive, and even dangerous. Some teachers, demoralized by such reforms, have abandoned their vocation. A better solution, for education and health care, may lie in new bottom-up performance-improvement systems that treat an essential human service field as a profession rather than an industry (Table 1).

INDUSTRIAL MODEL, PROFESSIONAL CRITIQUE

In factories, process standardization eliminates variation in quality, and the quality of U.S. education and health care is unacceptably spotty. In the 1980s, industry strategies began seeping into both fields. Writing in 1989, Donald Berwick, founder of the Institute for Healthcare Improve-

ment, touted the Theory of Continuous Improvement used by Japanese corporations, whereby data on production processes are analyzed and used to revise those processes and all workers “are enlisted in the pursuit of better ways.”¹ Berwick preferred continuous improvement to the Theory of Bad Apples, whereby individuals are blamed (and punished) for the poor performance of a system, an approach he discerned behind developments such as the publication of hospital mortality data, which bred fear and defensiveness.

Unfortunately for education, the quick fix promised by the bad-apples theory appeals to politicians. Federal education policymakers sought to apply free-market principles to schools — recommending competition from charter schools, merit pay based on standardized test scores, firing of “bad” teachers, and the shuttering of “failing” schools.² By 2001, these principles were embodied in the No Child Left Behind Act, the education-reform law aiming to hold schools accountable for student achievement (as measured by standardized tests) so that low-performing schools would somehow mend their ways — or be closed. Never mind that under the initial standards, all schools would eventually fail, according to Douglas Staiger, a Dartmouth economist who studies performance measurement in both education and health care.

Like physicians, Staiger notes, teachers have a strong culture of independence that can translate into resentment of oversight. A more fundamental problem was that teachers — also like doctors — know that their work requires customization that is based on essential human variables and expert judgment; standardization can be counterproductive.

Critiques of performance measurement in education include familiar themes: measures are focused too narrowly and sometimes on the wrong aspects of process or outcomes; they don't ac-

Table 1. Top-Down and Bottom-Up Performance Measurement in Health Care and Education.

Variable	Health Care	Education
Desired product	Healthy, long-lived patients	Well-adjusted, knowledgeable, skilled, intellectually agile students
External contributors to outcomes	Genetic makeup, environmental exposures, adherence to medication, diet and exercise habits, income, educational level	Family situation, neighborhood safety, cultural values, preschool preparation, early language exposure, family income, parents' educational level, learning disabilities, class size and dynamics
Sample top-down performance measures		
Process measures	Percentage of patients who receive recommended screening tests	Teacher's classroom management, as assessed by the principal by means of classroom observation
Outcome measures	Rehospitalization rates and mortality	Students' achievement gains from one year to the next on standardized tests, usually in mathematics or reading comprehension
Surrogate measures	Percentage of patients in whom cholesterol levels are within a designated range	Teacher's certification status, number of years of teaching experience, possession of a graduate degree
Sample bottom-up performance measures		
Measures of competence and effectiveness	Utilization and quality metrics as reviewed annually by a colleague; variation on measures with no right or wrong rate (e.g., number of radiology tests per 1000 patients), shared in a blinded or unblinded fashion	Discussion and questioning skills, ability to engage students in learning, management of student behavior, as shown in the classroom and assessed in person or with the use of video by more than one trained observer
Measures of patient or student experience	Effectiveness of physician communication about diagnosis and treatment, coordination of care between physicians and nurses, adequacy of pain control	Usefulness of feedback on homework, teacher's ability to manage time in class, the expectation that students will correct their mistakes

count for the whole student, nor do they assess everything that makes teachers effective. Although there is some adjustment of scores for baseline characteristics of the student population, it's often inadequate and does nothing to improve things for the least fortunate. Moreover, assessors' expectations may be impossible to meet.

The effect of education reform in narrowing the focus of schools is widely decried. Even students who struggle with math or writing may be remarkably creative thinkers, designers, or planners, points out Mary Gallagher, a fifth-grade teacher in Cambridge, Massachusetts. Making these students feel like failures because they don't fit the mandated mold often turns them off to education. Just as required screenings now monopolize physician visits, crowding out individualized interactions, after education reform, says Gallagher, teachers no longer had time to discover and foster students' passions so that they could learn to enjoy learning. Syllabi were packed with drills, recesses eliminated, arts cut from the budget, interdisciplinary proj-

ects discontinued . . . and personnel decisions driven by test scores.

Those scores were generally the only performance measures available, but many educators argued that given students' varied developmental timelines and the myriad uncontrollable factors in their lives, it was absurd to hold teachers accountable for such outcomes (akin to physicians losing performance points when patients miss health-improvement targets because they were dealt bad genetic and environmental hands). Education experts confirmed that the law's goal of 100% proficiency by 2014 was unachievable.² And nobody was analyzing how to obtain good outcomes.

To fill that need, the Gates Foundation Measures of Effective Teaching (MET) project is studying process measures of teacher effectiveness, having videotapes of class sessions scored by observers trained to identify specific competencies.³ According to Thomas Kane, MET project director and a Harvard education professor who has done research with Staiger, teachers have been clamoring for assessment of classroom

techniques rather than outcomes. “I actually wonder what’s going to happen once they get what they ask for,” says Kane. “They might say, ‘Hey, wait a minute! I may not be using conventional methods and so don’t score well on your rubric, but I’m getting the results.’”

Physicians do make such arguments. Finding themselves mired in process measures that are based on best-practice guidelines and that sometimes seem inappropriately restrictive and not necessarily related to outcomes, many physicians balk at the resulting diminished flexibility and individualization. But when physicians demand outcome measures instead, Susan Edgman-Levitan from Massachusetts General Hospital — like Kane on the flip side — warns, “Be careful what you ask for.” Edgman-Levitan, an expert in measuring what matters to patients, cautions that doctors may have no control over outcomes: it may be “God and the patient” who determine, for instance, whether a chronic condition worsens over time.

Furthermore, people “have wildly unrealistic expectations about what can be improved how quickly,” says Edgman-Levitan. In trying to meet those expectations, institutions often change their practice in any way that makes them look better on the mandated measure, sometimes with negative consequences for students or patients. “Teaching to the test” is not the only result with health care parallels; most disturbingly, because it’s difficult to help the neediest “customers,” some hospitals turn away high-risk patients, just as charter schools fare better if they exclude the neediest children. In both fields, the most significant risk factor may be low income, but neither physicians nor teachers can eradicate poverty.

COUNTERMOVEMENT AND COLLABORATION

Well-documented effects of the accountability movement are demoralization and gaming of the system, including outright cheating, and pay-for-performance systems provide incentives for such manipulation. Adverse reactions have caused some school districts to backpedal on tying teacher pay to student test scores. When Tom Tomberlin, of the Center for Human Capital Strategies at the Charlotte–Mecklenburg (North Carolina) school district, unveiled plans to develop a pay-for-performance program, the

pushback persuaded him to drop the “pay for” and focus only on improving performance, asking teachers to help answer three questions: “What do you do in your classroom that matters and that you would *want* to be measured on?”; “If it matters, how do we measure it?”; and “Who measures it?”

Tomberlin’s work-in-progress has much in common with some other past and current local efforts: it’s a grassroots development rather than a government fiat; it uses multiple measures of teachers’ effectiveness; it aims to change school culture, partly by encouraging teacher collaboration; and it embraces the Theory of Continuous Improvement. Various school districts seem to be independently discovering similar ways of counteracting the ill effects of the accountability movement while retaining its laudable ambitions. And many critics of federal mandates see promise in this countermovement.

These initiatives have good reason to engage teachers from conception through implementation: collaboration is key to both process and product. “We’re making educated, well-reasoned guesses about what the . . . instructional practices are that lead to better student outcomes,” explains Tomberlin, but “no one knows whether there’s a causal relationship” between the two. With a top-down mandate, when “we find out that it doesn’t have any impact, teachers feel sour that they’re being experimented on, as well as the students,” he says. By tapping into teachers’ expertise and inventiveness, leaders hope to foster a “willingness to adapt and evolve and look at the data to make sure that what we’re doing makes sense, and if it doesn’t, we tweak, but now we’ve got partners in that process rather than guinea pigs.”

The inclusive process and multimeasure approach have won over some skeptics; veteran teacher Lysa Craig believes that Tomberlin’s teacher-effectiveness project will offer “a global perspective on who [teachers] are” and, equally important, will emphasize true professional development, in part through a peer-mentoring system that Craig is helping to design. Indeed, many educators see this countermovement as a professionalization process. “Hopefully,” says Craig, “this will be a way, down the road, that people will see a teacher as a professional and not just a little bit higher-class babysitter.” Self-regulation, after all, is considered one of the hallmarks of a profession.

Twenty years ago, physicians probably didn't require any such boost in public esteem, but some doctors complain that mandated guidelines have deprofessionalized their field, turning some medicine into rote, cookbook practice requiring no creative or critical thought. Although the natural reaction may be to defend one's turf, the performance-improvement countermovement is big on welcoming "consumers" into the tent. Like the new federal Patient-Centered Outcomes Research Institute and patient-experience surveys seeking input from intended beneficiaries of health care improvement, new initiatives in education engage students in assessing teachers and schools. Research from the MET project shows that student-survey results about teaching practices are excellent predictors of teacher effectiveness and student achievement.

In health care, says Edgman-Levitan, physicians find information on patients' experience of care "much more actionable" than other performance data. If, for instance, patients say that their doctors haven't explained their diagnosis, doctors don't question "what kind of information the person's referring to; they know exactly where to go." In education, "the magic in the student evaluation," says Kane, "is that it's measured on a scale that teachers just naturally care about. Especially if it's not just a popularity contest," but "more specific things, like 'When I turn in homework, I get useful feedback that helps me improve. Do you agree, disagree?' 'In this class, we learn to correct our mistakes. Do you agree, disagree?'. . . . Those kinds of questions are emotional dynamite," eliciting big behavioral responses.

Tomberlin and colleagues plan to link each chosen performance measure to professional-development activities that can help teachers improve their rating, shifting from a reward-and-punishment to a learning-and-improvement mentality. "In both education and health," says Kane, "our goal eventually is to get to a set of measures that practitioners are willing to embrace and internalize and use for trying to improve their practice."

SCHOOL LESSONS?

So does K-12 education hold lessons for health care? Tomberlin, considering the inverse question, invited the chief executive officer of a local health system to tell teachers engaged in his

project about a similar process behind a new nursing-effectiveness evaluation system. Recognizing that much of their nurses' time was devoted to "compliance activities, charting and paperwork and inventory," recalls Tomberlin, "and very little time was being spent on direct patient care . . . they worked with the nurses to define what effective nursing practices were and how nurses would be evaluated." The teachers in Tomberlin's district then generated ideas for applying health care practices to education — such as the notion of professional consultation, which would require organizational changes to facilitate teachers' consulting with peers, and "instructional rounds," which would get "lots of eyes in the room looking at lots of different interactions between the teacher and students."

But educators, in turn, have probably learned some lessons applicable to health care. Some of those Kane mentions, such as the need to seek outcomes with inherent value for practitioners, are already being learned by some health care reformers. It's also important to use multiple measures, some related to processes and some related to outcomes, to get the whole picture. Kane stresses that tying measures to compensation and expecting immediate improvement is targeting the wrong problem: such incentives might work if professionals simply weren't working hard enough, but not if they lacked the requisite talent, skills, or experience — so "maybe you don't look for short-term changes with a fixed population" but rather "gradual changes in practice."

Ultimately, success may require a broad cultural shift. Staiger suggests that isolated strategies such as pay for performance may be doomed to failure, but if an entire organization genuinely focuses on improving performance, linking pay-to improvement underscores that commitment. Some newly successful school districts, he notes, have "built everything around improving kids' academic performance," beginning by retaining only teachers who embraced that mission. Such collective commitment pays dividends in myriad arenas: Staiger says that the states that his research showed "were out ahead on the use of beta-blockers for heart attack patients in the '90s" were also "the first adopters of hybrid corn in the 1950s and were ahead in the high-school movement in the 1910s and '20s and had the highest adoption of PCs in the '90s. . . . Whatever is making things work cuts across

health care and education.” Interestingly, those states also score high on measures of social capital: people tend to belong to clubs and to know and trust their neighbors, deriving value from social networks and cooperation.

Some health care institutions have started down a road like Tomberlin’s,⁴ and the Institute of Medicine recently delineated a path to a learning health care system.⁵ A new project at Partners HealthCare in Boston resembles the teacher-effectiveness initiative in the Charlotte–Mecklenburg district: professionals focused on a particular health condition have formed teams (including patient representatives) to redesign care processes for that condition and to choose multiple measures of success that can be used, studied, and refined in a continuous-improvement program.

Bottom-up approaches won’t solve all quality problems or eradicate top-down mandates. Local performance measures may not be comparable across institutions. If pay is linked to metrics, the gaming will continue. Although in theory multiple measures make assessments fair, in practice, qualitative evaluations may be unduly influenced by quantitative ones. Successful local assessment systems can’t merely be cloned, since the collaborative development process is essential. And reacting to generic external mandates consumes resources that could be devoted to creating better systems. But despite government pressure, Tomberlin, for one, is encouraging teachers to build their evaluation system carefully, to avert more ineffective or harmful policies.

Many observers might agree with Rob Weil, deputy director of the Educational Issues Department at the American Federation of Teachers, that the accountability movement resembles “a huge boat in the middle of the ocean” going in the wrong direction and that “there will be a whole lot of collateral damage before the course is turned.” Fortunately, “there are people pushing on the rudder to change it.” The effort of the countermovement to improve practice from the inside out may someday provide a long-term response to demands for uniformly high quality in essential human services.

Disclosure forms provided by the author are available with the full text of this article at NEJM.org.

I thank Thomas Lee, Jerome Groopman, Pamela Hartzband, Meredith Rosenthal, Andrew Eisenhauer, Greeshma Shetty, and Janet Corrigan, in addition to everyone quoted in this article, for sharing their insights on performance measurement and accountability.

1. Berwick DM. Continuous improvement as an ideal in health care. *N Engl J Med* 1989;320:53-6.
2. Ravitch D. The death and life of the great American school system: how testing and choice are undermining education. New York: Basic Books, 2010.
3. Project MET. Working with teachers to develop fair and reliable measures of effective teaching. Bill and Melinda Gates Foundation, June 2010 (www.metproject.org/downloads/met-framing-paper.pdf).
4. Nichols LM. Be not afraid. *N Engl J Med* 2010;362:e30.
5. Best care at lower cost: the path to continuously learning health care in America — consensus report. Institute of Medicine, September 6, 2012 (<http://www.iom.edu/Reports/2012/Best-Care-at-Lower-Cost-The-Path-to-Continuously-Learning-Health-Care-in-America/Report-Brief.aspx>).

DOI: 10.1056/NEJMms1306048

Copyright © 2013 Massachusetts Medical Society.

NEJM CLINICAL PRACTICE CENTER

Explore a new page designed specifically for practicing clinicians, the NEJM Clinical Practice Center, at www.NEJM.org/clinical-practice-center. Find practice-changing research, reviews from our Clinical Practice series, a curated collection of clinical cases, and interactive features designed to hone your diagnostic skills.