

#### COMMENTARY

# Lessons from Health Care

by Karla Brooks Baehr

Originally published in District Management Journal, v.15, Spring 2014

Like doctors, teachers are beginning to confront the implications of the bell curve of teacher performance.

# Lessons from Health Care

| Karla Brooks Baehr

wenty-five years ago, during an earlier recession, I was a young superintendent facing the twin challenges of cutting budgets and improving school quality. My principals and I felt put-upon and singled out. Fortunately, leaders of the local education foundation in my suburban community decided to offer us some much-needed perspective. They invited several CEOs to meet with us to describe the leadership challenges they were facing in their respective organizations: a hospital, an insurance company, a bank, and a law firm. Over dinner, they each described the rapid changes necessitated by technological advances, sweeping regulatory change, higher expectations from consumers, and new competitors. In the ensuing discussion about leadership in an environment where "doing more with less" had become the norm, I was both humbled and comforted: humbled by the vastly more sweeping changes my CEO colleagues were facing at the time, and comforted by the fact that I now knew I was not alone.

Ever since then, I have been more curious about the challenges leaders in other fields face and what their responses might teach me about tackling challenges in education. Some powerful lessons have come from surgeon, public

health researcher, and writer Atul Gawande. In his New Yorker article "The Bell Curve," Gawande explores how and why the treatment of cystic fibrosis (CF) has changed so dramatically over the past 50 years. A diagnosis of CF was once a death sentence: in 1964, the average life expectancy for a child with CF was only three years.<sup>2</sup> Today, the average life expectancy is well over 40 years! Gawande centers his essay "The Bell Curve" on Cincinnati Children's Hospital as it confronts a shocking discovery: despite its stellar reputation as a preeminent children's hospital, Cincinnati Children's performance treating children with cystic fibrosis was, at best, mediocre—it was low on the left tail of the bell curve. How Cincinnati Children's chose to address its mediocrity has much to teach those of us who lead in schools. There are compelling lessons about the power of data, the momentum created by focusing on leading versus lagging indicators, and the difference that unflaggingly high expectations can make in performance. The story resonated with me as I reflected on my experiences as superintendent and later as an education policymaker in Massachusetts; I think many others in the education field will find these lessons relevant as well.



# The Bell Curve of Performance

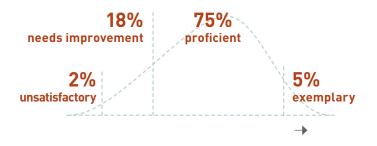
Differences among hospitals and doctors in a particular specialty were once thought to be insignificant. Gawande explains that it was generally assumed that "most places were clustered around the very best outcomes. But the evidence has begun to indicate otherwise. What you tend to find is a bell curve: a handful of teams with disturbingly poor outcomes for their patients, a handful with remarkably good results, and a great undistinguished middle."

Significant disparities in results have led to vigorous debates in medical circles about what constitutes fair measures and what to do with information that reveals differences—make it public? to whom? how? Disparities have led to soul-searching, as well. Gawande asserts that these disparities "contradict the belief nearly all of us have that we are doing our job as well as it can be done." He concludes: "If the bell curve is a fact, then so is the reality that most doctors are going to be average. There is no shame in being one of them, right? Except, of course, there is. Somehow what troubles people isn't so much being average as settling for it."

Like doctors, teachers are beginning to confront the implications of the bell curve of teacher performance. In Massachusetts, for example, until 2012, most districts rated teachers' performance as either satisfactory or unsatisfactory, and very few districts used measures of student learning to make the determination. The new evaluation system Massachusetts developed to help us win federal

Race to the Top grant funding is changing all that: there are now four rating levels instead of two; student achievement of learning goals plays a role in determining the rating; and each teacher (and principal) will receive a rating of high, moderate, or low impact on student learning based on their students' growth on common measures of student learning.<sup>6</sup>

In its first year of implementation, we were already feeling its impact: a bell curve of teacher performance was beginning to take shape. For example, before implementing the new system, the Springfield (MA) district's distribution of teacher performance mirrored the national pattern, in which 99% of teachers are rated satisfactory in binary systems, 94% of teachers are rated in the top two categories in systems using three or more rating categories, and fewer than 1% of teachers are being rated unsatisfactory, regardless of the number of rating categories. At the end of its first year under the new system, Springfield's distribution began to look more like a bell curve as shown below:



When student growth scores are fully utilized to assign ratings of high, moderate, or low impact on student learning, Massachusetts' educators expect the performance-rating curve to approximate even more closely a bell curve, with a small proportion of low- and highimpact ratings and a large proportion of moderate ratings.

While exposing the wide range of performance is unsettling to many educators—and fear of public exposure is palpable—there are signs that moving from the illusion that we are all delivering good outcomes to the reality of a bell curve is prompting improvement efforts. Teachers, like doctors, do not want to settle for "average." Already, districts are identifying educators whose students consistently over time and across measures are demonstrating high levels of growth, and these districts are inviting teacher teams to collaborate to identify, emulate, and adapt the practices of the high-performing teachers to their own contexts. Like their counterparts in CF centers, they are beginning to accept the reality of a bell curve of teaching performance and are working to improve their position on the curve. In time, we hope to see the bell curve tighten, and eventually to shift the entire curve to the right: we hope that what is today's moderate impact will become tomorrow's low impact (Exhibit 1).

**EXHIBIT 1: THE BELL CURVE OF PERFORMANCE** 



Our Goal: Tighten the Curve and Shift It to the Right Learning from one another, we seek to reduce the gap between low and high performance (TIGHTEN) and make yesterday's high performance tomorrow's merely average (SHIFT).

#### Data as a Call to Action

In his article, Gawande explains that the amount of data available about CF treatment and outcomes is unusual. The CF Foundation first started collecting data 50 years ago because it had to: a pediatrician in Cleveland named LeRoy Matthews was claiming to have an annual mortality rate of less than 2% at a time when the national mortality rate was estimated to be higher than 20% a year. Practitioners thought the claim preposterous, and so the CF Foundation commissioned a study that collected data from all 31 CF centers, with the promise that no individual center's data would be made public. The results? The median estimated age at death in Cleveland was, in fact, 21 years—seven times the age of patients treated elsewhere. Cleveland's secret was that its doctors saw CF as a cumulative disease, and therefore provided aggressive treatment long before patients became sick. The Cleveland treatment quickly became the standard in CF centers; within six years of the study's publication, average life expectancy nationwide reached 18 years. The database on treatment centers proved to be so useful that the CF Foundation has continued its collection to this day.8

> Gawande notes that such credible data does not exist for most fields of medicine. He is dismissive of the usual public measures of physician and hospital "performance," which fail to provide meaningful or actionable information. As an example, he cites "a company on the Web called HealthGrades, which for \$7.95 will give you a report card on any physician you choose," but that reveals little more than the doctor's certification, whether the doctor has a criminal conviction, has had his or her license revoked, or has been disciplined. While this is useful information, Gawande points out that it sets the bar low and gives patients little meaningful information about the doctor's performance.

His concerns echo mine and those of many of my colleagues about most conventional public measures of schools. My suburban district had SAT scores, rates of students going on to four-year private colleges, and results on the statewide math and ELA assessments that were among the highest in the state. I left to lead an urban district where these measures told a very different story of impact. I longed for a more credible measure of performance. In 2008, I got one! The state had developed a measure of annual student growth based on a comparison of each student with a cohort of students statewide who had had similar scores in prior years on each state assessment. When the median growth score in mathematics for our urban middle school students was higher than the scores in the surrounding suburban middle schools, I had a measure of our teachers' performance that I found credible. Having access to more detailed, meaningful, credible and comparative data can help us better understand how we are performing, and can help drive improvement in the education field in much the same way as was done in the field of CF.

## "Going Naked" Gets Results

Partly as a consequence of the availability of credible data for cystic fibrosis, CF care is now far more consistent across doctors than it is for most diseases. The 117 CF centers all undergo a rigorous certification process, follow the same detailed guidelines for CF treatment, and participate in research trials to develop new and better treatments. Given the consistency in approach and availability of credible data, one would expect that performance among the 117 centers would not vary greatly. That is not the case, however; in 2003, for example, life expectancy with CF was 33 years nationally, but at the best center it was more than 47 years—over 40% higher! 10

Gawande recounts that in 1997, the doctors at Cincinnati Children's Hospital's CF Center (Cincinnati CF) knew that their center was underperforming: lung function in patients under 12 years old was in the bottom 25% of all CF patients nationwide. But no one else knew how the center was performing because the CF Foundation had guaranteed anonymity to ensure participation in the data study. That changed in 2001 when Cincinnati CF agreed to "go naked" with its comparative data in exchange for a

multimillion-dollar grant. They shared the performance data with families of their CF patients, described commitments they intended to make to raise performance, and invited family members to have at least one seat on each of the committees being set up to chart the course forward. Not a single family left Cincinnati CF. Gawande described one family that was tempted to leave but decided to stay because they "were impressed that the team had told them the truth, ... had made no excuses, and everyone appeared desperate to do better." But doing better was not going to be easy: because of the guarantee of anonymity, Cincinnati staff did not know which centers they might learn from. They had a very difficult time convincing the CF Foundation to give them the names of the top-performing centers, but once they got the names, they were able to study their results and practices. Knowing who the top performers were and what they did was a key to improvement.

I likewise saw the power of credible and public comparative measures of performance in 2010 when the first schools in Massachusetts were identified as in need of turnaround. An elementary school in my former district was on the list. For years, many on its staff had dismissed comparisons with other schools, arguing that their school had an unusually high percentage of the district's "neediest" students. But armed now with student growth percentile scores, district leaders were able to point to a school nearby with even more students with disabilities, a similar number of English Language Learners, and nearly as many students eligible for free lunch. The nearby school had had three years of growth scores exceeding the 60th percentile, while the turnaround school had yet to see growth scores exceeding the 25th percentile. The staff got angry, but this time at themselves. Their competitive juices flowed; they resolved to understand the practices in place at their more successful neighbor school and adapt them to their context. Their efforts were supported by a new resource made available by the state: an online tool that enabled anyone-parent, teacher, or student-to compare their own school's performance with that of any other school in the state on a range of selected input and output measures. This benchmarking tool also allowed users to identify and display data for the nine schools in the state most like their own based on grade configuration, size, and basic demographics.<sup>12</sup> Using this tool, district leaders discovered a high-poverty school in a nearby city that →

was substantially outperforming even their highest-performing elementary school; they set about to learn and adapt that school's approach with English Language Learners.

In education, we haven't had a choice about whether key performance data are public or not. Policymakers have made us "go naked." And our families have more choices than CF patients for whom another center may be 100 miles or more away. Yet, we do have a choice about how

Like doctors, teachers are beginning to confront the implications of the bell curve of teacher performance.

we respond to the data. Cincinnati's approach of sharing the brutal truth non-defensively, conveying its determination to achieve the results the top performers were achieving, and inviting families to play a role in shaping the plan for improvement is proving relevant in our field. The leadership team charged with turning around a failing middle school in Boston, for example, succeeded in retaining a very large proportion of families using just such a strategy.<sup>13</sup>

# Leading-Not Lagging-Indicators Drive Change

In medicine, the life expectancy rate, like the median student growth score in education, is a "lagging" indicator. Lagging indicators don't give doctors enough guidance about what kind of changes they can make to improve performance, other than sending their sickest patients to other hospitals. This "solution" is similar to one a few of my colleagues used in the early days of school accountability when we lacked credible measures like student growth scores: they sought to "turn around" failing schools primarily by reassigning low-scoring students or

relocating programs that served low-scoring students! Having access to credible lagging indicators can spark an urgent desire to change, but it is access to actionable leading indicators that will help us know how to change. Practitioners in both fields need more than lagging indicators if we are to adopt a continuous improvement approach to our work and develop the standards of practice and innovation that move a field forward over the long haul.

For both CF practitioners and patients, lung function is one of the two leading indicators that matter most. (Body Mass Index, a measure of nutrition, is the other.) Lung function makes a marvelous focus for CF treatment for three reasons: it is the single best predictor of survival rate for CF patients; it can be easily measured as often as daily by doctors and patients alike; and its measurement can be quite precise. Organizing medical intervention around a leading indicator with these characteristics has been key to improving CF treatment and outcomes.

What are the leading indicators around which educators can focus our actions? What is education's equivalent of lung function? It might be reading proficiency. Because third-grade reading proficiency is a strong predictor of future school success, elementary schools typically see dramatically improved performance when they take these steps: (1) organize their schedules and staffing around literacy starting in preschool, (2) develop interim assessments of reading and use the results regularly to identify interventions and mid-course corrections, and (3) strengthen their professional expertise in teaching reading. But reading isn't as readily or precisely measured as lung function. Nor is it entirely clear what some other leading indicators should be: fifth-grade attendance? eighth-grade writing proficiency? ninth-grade promotion rate? success in high school college-prep math courses? Yet, a lesson that is clear from "The Bell Curve" is that we might achieve breakthrough results by organizing our "treatment"—our "education" treatment—around a few, highly predictive, readily measurable leading indicators. What these indicators should be is something to which we should be giving very careful thought. One promising candidate for high schools is a relentless focus on literacy skills as assessed regularly against a school-wide rubric. This strategy has proven critical to the sustained turnaround of a 4,200-student, high-poverty urban high school in Massachusetts. 14

# Common Attributes of Highly Effective Teachers / Physicians





- Have clear learning / health outcome targets and their students / patients understand what it takes to get better and own their learning / health. They focus on ultimate learning / health outcomes more than on compliance with required assignments / protocols.
- Create a culture of redemption. They assess frequently and see students' / patients' mistakes as a road map to improvement.
- Constantly and frequently tweak their lessons / treatment plans in response to how students / patients are doing.
  Students' / patients' learning / health needs are more important than lesson / treatment plans.
- Ask questions that go to the heart of the subject / medical condition and teach students / patients to pose their own questions. They track misunderstandings and then clarify them for their students / patients. As students / patients learn to ask the right questions—those related to their learning targets—they begin to own the goals and maximize their learning / health.
- They create a culture of high expectations coupled with good relationships. These are not friendships but partnerships ["You and me, in this together"] focused on academic achievement / improved health. This tenacity, concern, and love for each student / patient are obvious, yet are linked directly to unyielding aspirations for each student / patient.

### Data and Focus Spark Innovation

Many innovations in CF treatment developed from the medical community's tight, shared focus on improving lung function. Those innovations consistently shown to improve lung function quickly spread, becoming standard procedure at CF centers. For example, a mechanized vest invented at the Minneapolis CF Center to replace the daily manual "chest thumping" that had been an essential component of treatment for every CF patient led to more uniformly effective decongestion, greater convenience for patients (and their families), and therefore more consistent use of the treatment.

A similarly tight, shared focus on reading proficiency is leading to some promising innovations, including systematic support for caregivers to use enriched vocabularies with young children and adaptive software that customizes literacy challenges to children's stage of language development. As with CF treatment, innovation is being spurred by a focus on a key indicator and data about what is working and what is not. Already, the longer school days and intentional culture-building efforts of Massachusetts' highest-performing charter schools are becoming centerpieces of the state's efforts to turn around failing traditional schools. Careful examination of the impact of innovations made possible by

new technologies—Khan Academy, Teach to One, and "flipped classrooms"—may yield the education equivalent of the CF mechanized yest.

#### The Power of High Expectations

While conducting research for his article, Gawande visited Cincinnati CF and then went to Minneapolis CF, one of the top five centers, to discern why one's performance was mediocre and the other's was consistently among the very highest. What he found at Cincinnati CF, he wrote, "impressed me and, given its [low] ranking, surprised me"; he described the staff as "skilled, energetic, and dedicated." He witnessed "real medicine: untidy, human, but practiced carefully and conscientiously—as well as anyone could ask for." Gawande then went to Minneapolis, where he found patients "got the same things that patients everywhere did. ... Yet, somehow, everything ... was different."

The differences were exemplified in how a doctor at each center treated an adolescent girl whose lung function had declined since her last visit. The sharpest difference was in expectations and each doctor's response. In Minneapolis, Dr. Warwick would not settle for lung function less than 100%, meaning lung function at least as good as that in someone without CF. →

In Cincinnati, Dr. Daines settled for less. Dr. Warwick demanded the patient take an active role in her treatment. He asked probing questions to get at the patient's behavior and underlying assumptions so that he could collaborate with her to invent ways for her to contribute to her own care. Dr. Warwick paid attention to seemingly trivial differences—and made sure his patient did, too: he explained to her the difference between a 99.95% chance each day of staying well with CF treatment and a 99.5% chance of staying well without treatment that day. He added up the difference of that .05% over 365 days to show that the seemingly tiny daily difference meant the difference between an 83% chance of making it through the year without getting sick and only a 16% chance. Dr. Warwick insisted that she enter the hospital for two to three days to "make up lost ground." Dr. Daines made no such demands. Faced with slippage from 80% to 67%, Dr. Daines settled for having her patient return for a follow-up visit in three months instead of six.

Dr. Warwick's relentless, aggressive, and collaborative focus on achieving a goal that might seem beyond reasonable felt to me like the approach taken by some of the strongest reading and math teachers I've observed over the years. Recently, two researchers studied elementary and middle school teachers whose students showed significant gains in learning for three consecutive years to identify common attributes of their approach. Simple word substitutions in each of the five attributes uncovered through the research make clear the almost uncanny similarity between highly effective patient care in Minneapolis and highly effective teaching practices: both depend on ambitious and clear targets, responsiveness to data, a strong relationship and partnership between teacher/physician and student/patient, and insightful questions (Exhibit 2).

Findings from two Massachusetts studies of school turnaround efforts reinforce how relentless, aggressive focus can make the difference between exemplary and mediocre performance of schools. The studies explored the differences between rapidly improving and "stuck" schools. Rapidly improving schools implemented three strategies effectively and with a degree of sophistication that was "substantively different than those schools with little or no immediate achievement gains":

- The school has an instruction- and results-oriented principal who has galvanized both individual and collective responsibility for the improved achievement of all students through a variety of deliberate improvement structures, expectations, practices, and continuous feedback.
- The school has created instruction-specific teaming and teacher-specific coaching for pursuing ongoing instructional improvement.
- The school has developed a well-orchestrated system of ongoing data collection and analysis that informs a continuously responsive and adaptive system of tiered instruction directly attentive to students' specific academic needs.<sup>19</sup>

Just as at the Minneapolis and Cincinnati CF centers, the telling difference between the high- and low-performing schools came not from the strategy they used, but from the difference in focus, intensity, intentionality, and sense of urgency each school brought to the challenge of implementing that strategy.

#### A Hopeful Outcome

So how has Cincinnati CF fared since "going naked" with its data in 2002? Cincinnati CF now ranks among the best CF centers,<sup>20</sup> and a recent advertisement boasts an average lung function of over 100%, making CF patients almost indistinguishable from people without CF! (See Exhibit 3) The bell curve has shifted significantly to the right over time. And, performance differences among the CF centers have narrowed, although a substantial gap still remains between the highest and lowest performers: while the mean lung function among the top CF centers is very close to 100%, the mean for all centers is only 75%. Perhaps most impressively, the Minneapolis center has sustained its position as a top performer for over 40 years. The systems, processes, and procedures that the leaders there have established and sustained built the ongoing capacity to learn, adapt quickly, innovate, and remain leaders in the field. Gawande believes this capacity distinguishes great performers from those whose technical skill and knowledge are comparable, but whose results are only good-or worse.

When we were debating the draft of educator evaluation regulations in Massachusetts, one critic decried the consequences of building a system that assumed a bell curve of performance. She observed that student growth percentile (SGP) was a normative measure: by definition, 49% of the students would make above average growth and 49% below average. She asked, "What if absolute performance of students rises so that the typical student of tomorrow is performing at dramatically higher levels than the typical student today? Wouldn't it be unfair to have the same percentage of teachers rated as having a low impact on student achievement?" I responded by saying that if that happens, that would be a really nice problem to have!

As educators, we can accept that a bell curve of performance does exist—and always will-for our districts, our schools, our educators—and ourselves. We can seek to move the performance of our individual districts to the right—from low to average, from average to high, or from high to higher. We can learn with and from one another how to tighten the bell curve, and we hope to shift it to the right: yesterday's high performance in education should become tomorrow's merely average. To do so, we need to take seriously the lessons from health care that Gawande detailed in "The Bell Curve." Educators, too, need to seek out credible, actionable data. and respond non-defensively with high expectations and a sense of urgency; we can emulate the "focus, aggressiveness, and inventiveness"21 that characterize Minneapolis' approach to treatment—and now Cincinnati's. And we can cultivate our districts' capacities to learn and adapt quickly by developing and sustaining processes, structures, and procedures that support those capacities. In the end, we owe our students nothing less.

#### **EXHIBIT 3:**

In a recent ad, Cincinnati Children's acknowledges its poor performance ten years ago and now celebrates that its outcomes are among the best in the country.



#### NOTES

- Atul Gawande, "The Bell Curve," The New Yorker, December 6, 2004. Accessed March 15, 2014. http:// www.newyorker.com/printables/fact/041206fa\_fact.
- 2. Cystic fibrosis (CF) is a genetic disease that affects cells' ability to manage chloride. The chloride thickens secretions throughout the body. Lung capacity shrinks, making breathing increasingly difficult; blocked enzymes make it harder to absorb food, so the danger of malnutrition is high.
- 3. Gawande, "The Bell Curve," paragraph 12.
- 4. Ibid. paragraph 14.
- 5. Ibid., paragraphs 83-84.
- 6. For details on the Massachusetts requirements for educator evaluation, see www.doe.mass.edu/edeval/.
- See "The Widget Effect: Our National Failure to Acknowledge and Act on Differences in Teacher Effectiveness," The New Teacher Project, 2009, available at widgeteffect.org/download/TheWidgetEffect.pdf.
- 8. Gawande, "The Bell Curve," paragraphs 20-24.

- For a description of Massachusetts Student Growth Percentile (SGP), see www.doe.mass.edu/mcas/ growth/.
- 10. Gawande, "The Bell Curve," paragraph 24.
- 11. Ibid., paragraph 32.
- To see the Massachusetts District and School Analysis and Review Tools (DART), see www.doe.mass.edu/ apa/dart/.
- 13. For a brief summary of Orchard Gardens' turnaround and sustainability strategies, see http://www. edvestors.org/wp-content/uploads/2013/07/05-Orchard-Gardens-showcase-briefing.04.04.13.pdf. For a longer case study on its turnaround strategy and partners, see www.timeandlearning.org/files/OG CaseStudy.pdf.
- 14. For more on the turnaround of Brockton (MA) High School, see Susan Szachowicz, "Transformed by Literacy," in Principal Leadership, November 2010, available at www.betterhighschools.org/webinar/ documents/Transformed\_by\_Literacy\_by\_Dr.\_ Szachowicz.pdf.

- 15. Gawande, "The Bell Curve," paragraph 39.
- 16. Ibid., paragraph 41.
- 17. Ibid., paragraph 42.
- Jodi Newton and Betty Winches, "How to Maximize Learning for All Students," in Reading Improvement 50 (Summer 2013): 71-74; as summarized in Marshall Memo #525, February 24, 2014.
- 19. See "Emerging Trends in School Turnaround 2012 and 2013," at www.doe.mass.edu/apa/sss/turnaround/.
- See Avery Comarow, "A 'Best' Hospital for Cystic Fibrosis Kids," US News & World Report, July 19, 2010, at www.health.usnews.com/health-news/besthospitals/articles/2010/07/19/a-best-hospital-forcystic-fibrosis-kids.
- 21. Gawande, "The Bell Curve," paragraph 66.