

Measuring Outcomes: Lessons From the World of Public Education

Andrew Saultz, PhD¹

John W. Saultz, MD²

¹Department of Educational Leadership,
Miami University, Oxford, Ohio

²Department of Family Medicine, Oregon
Health & Science University, Portland,
Oregon

ABSTRACT

The quality and efficiency of American health care are increasingly measured using clinical and financial data with a goal of improving clinical practice. Proponents believe such efforts can improve outcomes, motivate clinicians, and inform the public about quality. Detractors point to problems with the accuracy of these measures and the risk of creating perverse incentives for both physicians and patients. Drawing on lessons from similar performance management policies in public education, we provide guidance about this trend for primary care physicians and health care policy makers. We argue that public school teacher evaluations that use value-added modeling foretell specific pitfalls for the use of similar models to evaluate physician effectiveness, and that unintended consequences of performance management in both education and health care can include the narrowing of purpose, deprofessionalization, and a loss of local/community control.

Ann Fam Med 2017;15:71-76. <https://doi.org/10.1370/afm.1995>.

INTRODUCTION

For much of the 20th century, the quality of medical care was considered to be an attribute of the person providing the care. Individuals seeking the best care looked for physicians who were board certified, trained in prestigious institutions, and recommended by friends or family members, or some combination thereof. Over the last 2 decades, this concept has eroded in the face of growing evidence of systemic problems with poor quality and uncontrollable costs in American health care.¹ Today, as new information technology facilitates electronic data collection, the paradigm of quality as a clinician attribute is being replaced by systems of objective quality measurement and reporting. The notion of determining health care quality by measuring outcomes is based on 3 assumptions: that quality can be accurately measured, that quality can be improved, and that reporting quality measurement data will influence how people use health care and how clinicians practice.²⁻⁶ Physicians and hospitals have challenged these assumptions while also expressing concerns about the high cost of implementing quality measurement.⁷⁻¹³ Foremost among their concerns is the inability of most outcomes studies to prove causal relationships between the care a patient receives and the outcome of that care, as quality studies rarely use random assignment of patients. Attempts to adjust quality measurements based on case mix complexity and social confounders have not resolved this controversy. In addition, data used to measure quality have generally come from electronic claims and patient satisfaction surveys, and, more recently, from data abstracted from electronic health records. All 3 sources of data have their limitations.

Meanwhile, the policy community including health purchasers and health plans has begun to focus on measurement of value.¹⁴⁻¹⁶ In this context, value is defined as the health outcomes achieved per dollar spent and potentially represents a way to prioritize health care services.¹⁷ Value measurement is a tempting policy target as the health care system struggles to achieve the triple aim of enhancing the experience of care for

AJC Annals Journal Club selection;
see inside back cover or <http://www.annfammed.org/AJC/>.

Conflicts of interest: authors report none.

CORRESPONDING AUTHOR

John W. Saultz, MD
Department of Family Medicine
Oregon Health & Science University
3181 SW Sam Jackson Park Rd, FM
Portland, OR 97239
saultz@ohsu.edu

individuals, improving population health, and lowering cost.¹⁸ But value measurement cannot work unless accurate and actionable quality measurement can be achieved, and this goal has not yet been attained.¹¹⁻¹³ So health policy leaders and health care professionals continue to struggle with a widening gap between the promise of quality improvement and value measurement, and the reality of their limitations. Of more concern to physicians is growing interest in measuring outcomes as an attribute of individual physician performance because of the effect this might have on their professional reputations.¹⁶

Health care is not the only sector of American life to experience attempts to measure outcomes and value. For more than a quarter century, public education has undergone a similar transformation. Although schools were once judged by the satisfaction of their communities, educational policy leaders are now focused on measuring and improving educational outcomes such as graduation rates and standardized test scores. Just as many physicians and hospitals resist health outcomes measurement, teachers and local school leaders have criticized educational outcomes measurement for failing to account for differences in the populations of students being served. It is our contention that health care professionals and health policy leaders can learn much from teachers and education reformers. The purpose of this article is to outline how outcomes measurement and public reporting of "quality" data in public education can serve as both a positive example and a cautionary tale for similar changes in health care.

PERFORMANCE MANAGEMENT IN EDUCATION

Performance management is a term developed to depict a movement that began in the 1980s to reinvent government using ideas from the private sector, with a goal of improving public institutions and restoring public trust through efficiency and effectiveness.^{19,20} Performance management links information to decision making by defining performance standards, collecting data, establishing sanctions and rewards, and reporting performance.²¹

The use of performance management in educational policy began in the 1980s and 1990s at the state level. Most states developed educational standards including ones outlining what courses schools should teach and specific high school graduation requirements, which were previously determined by local school boards.²² Over time, states also required more information about student performance on standardized tests, and some states established sanctions for schools that were consistently deemed to be low-performing.²³

The No Child Left Behind Act (NCLB) of 2002 was a historic shift by the federal government to codify the main components of educational performance management into federal law. The law mandated that states establish statewide standards, test all students in grades 3 through 8 in reading and mathematics, and report the results to the public.²⁴ NCLB allowed the US Department of Education to set annual goals for academic progress, called adequate yearly progress, and stipulated that schools that failed to achieve these goals for 3 consecutive years must notify parents about alternative educational options. If a school fails adequate yearly progress for 5 straight years, it is mandated to restructure itself as a charter school, replace most of the school staff, or close.^{19,25} Through public reporting, the authors of NCLB hoped that the public would pressure schools to meet performance targets.^{26,27}

NCLB introduced federal accountability at the local school level. Data were collected and reported for each school, as well as each school district. Over time, policy makers began to shift the focus from school level accountability to accountability at the individual teacher level. The quintessential policy change in this movement was reforming teacher evaluation. Teacher evaluation had been historically based on direct observation by the principal. The new evaluation model uses value-added modeling to isolate the individual teacher's impact on student performance. The use of value-added models (VAMs) for teacher evaluation is premised on the assumptions that student learning is measured by standardized tests, and that a more effective teacher leads to larger improvement in student examination scores.²⁸ Each teacher receives a VAM score, usually calculated by estimating a predicted test score of a student based on his/her results from previous standardized tests. A VAM then measures year-to-year learning growth for each student within a teacher's class, and compares the actual student performance with the predicted performance. The difference between these 2 scores is attributed to the teacher effect and is used to create the teacher value-added score.²⁹ Since 2009, more than two-thirds of states have changed teacher evaluation policy to include some measure of student performance on standardized examinations.³⁰ The overwhelming majority of these states have included some measurement of value added by each teacher even though empiric evidence suggests that the use of VAM in teachers' evaluations is not statistically reliable.³¹

Using VAM to evaluate teachers represents 2 major shifts in educational policy. First, VAMs shifted the accountability focus from the school to the individual teacher.³² Second, they shifted the goal from a minimum level of proficiency for all students to a system that measures growth in learning for every student.

Older accountability systems measured only the percentage of students achieving minimum proficiency on standardized tests.³³ This older model has been criticized for incenting teachers and schools to ignore high-performing students who would easily meet the target levels of proficiency and very low performing students who had a small probability of meeting the standard.³⁴ VAM evaluations are designed to get teachers to focus on all students, as the growth in learning is calculated for the total student population, regardless of previous test scores.²⁹

EFFECTS OF PERFORMANCE MANAGEMENT IN THE EDUCATION SYSTEM

How have these performance management reforms affected teachers, schools, and the public's perception of public education? One positive result is that the public has more detailed information about schools, teachers, and student achievement.³⁵ There have also been negative consequences of these changes. Performance management policies, and VAM specifically, were developed outside of the school system to improve performance and accountability.²⁹ The policies have been imposed on teachers, not undertaken with or by teachers, and this practice has reduced the ability of teachers to self-regulate their profession. VAM also places exclusive focus on end results, rather than the process of how to improve practice.²⁹ This focus on end results disregards any specialized knowledge that teachers might have that could inform others about how to improve and, in some cases, has seriously harmed teacher morale. Teachers have less autonomy over determining what quality teaching is, and professional excellence has been defined solely by who can produce the highest student learning growth through VAM.

VAM also leads to more individualized teaching. Under NCLB, accountability at the school level incentivized teachers to collaborate to improve all students' learning. Conversely, VAM incentivizes teachers to care about only the students in their own classes.³² For example, it is easier for a teacher to maximize his or her own VAM if the previous year's teacher performed badly because it is easier to improve low test scores. At its worst, VAM incentivizes teachers to think about only the students they interact with, and to not assist other students or teachers. Further, some scholars link the increased use of performance management, and subsequent pressure on teachers, with cheating scandals in Georgia, Ohio, and Texas.^{36,37}

Performance management has led to a narrowing of school curricula.³⁸ Given that schools and teachers are evaluated only on subjects that are tested (ie, math and reading), they have reallocated resources to

focus more exclusively on these areas.³⁴ Schools therefore have decreased the amount of time dedicated to recess, art, and music.³⁸ Conversely, a large study of teacher evaluation policies, funded by the Bill and Melinda Gates Foundation, found that teacher evaluations are more accurate when they include multiple measures of effectiveness.³⁹ Despite this evidence, VAM scores that primarily rely on test scores are still the central component in many teacher evaluation policies.⁴⁰ This situation highlights a compromise between trying to implement measures of performance without investing the time and money needed to make them more accurate and useful.

The educational system has also experienced more criticism of teacher preparation programs following the implementation of VAM evaluations.⁴¹ As VAM creates individual teacher-level data on effectiveness, it is easier than ever to link a teacher's performance with his/her preparatory institution. Over time, program ranking could shift from one that emphasizes the characteristics and aptitude of entering students to one that focuses on the effectiveness of their graduates in the field. This paradigm shift is made possible because individual teachers have VAM rankings.

Performance management created a system in which all schools are expected to have the same end goals (notably, high math and reading test scores). As a result, public dialog around normative questions has been silenced.⁴² What educational system does the public want? How does this preference vary from one community to the next? Research shows that citizens have different priorities for what public education should accomplish, including academic rigor, citizenship, and the arts.⁴³ When the system moved to evaluating schools, and ultimately teachers, by standardized test scores in just a few subjects, it assumed a uniform definition of what successful schools should produce.

Performance management has also changed the public's perception of the teaching field, and teacher quality. For example, in 2010, the *Los Angeles Times* began publicizing the VAM scores of every teacher in the LA Unified School District online. Teacher evaluations historically were confidential documents shared only between administrators and individual teachers. Today, parents and students can gather information on a teacher's effectiveness before they even meet him/her. This ability has a profound impact on the relationship between teachers and families, and may influence how the public views the teaching profession and public schools generally.

Lastly, VAM relies on a public knowledge of numeracy that does not exist. Although the use of advanced statistics to measure the teacher effect on student learning helps address concerns about differ-

ent student demographics across classrooms, it also condenses a teacher's ability to complete his/her job into a single number. There is no evidence that the public (including most teachers and physicians) has the statistical sophistication to understand how VAM functions, the advantages and disadvantages of the model, or statistical error. The public tends to believe that numbers are precise and accurately reflect reality,⁴⁴ but it is unclear whether they can understand the statistical complexity of modern teacher evaluation models.

LESSONS FOR PHYSICIANS

Performance management in education has progressively reduced local control of public education, with control moving first to the state and then to the federal government. This shift has taken place because of frustration about a perceived lack of accountability and a belief that management tools that work in business will improve educational outcomes. In general, teachers (particularly teachers unions) are perceived as resisting these changes, and their arguments about the importance of professional standards defined by experienced teachers and the limitations of standardized data have been discounted. The focus of education has narrowed, and the morale of teachers has suffered. Performance management is newer in health care, and although physicians may have a longer and more deeply established tradition of professional self-regulation, there is every reason to believe the same trends are at work. Studying the history of performance management in public education suggests that resisting the use of data to measure and improve outcomes is not a wise reaction to performance management. Physicians are better advised to embrace continuous quality improvement using data from their own electronic health records to demonstrate quality care to their communities. Performance management has been "done to" teachers⁴⁵; physicians should develop their own sources of data and communicate directly to the public about them.

The growth of performance management has increased the availability of data about schools for researchers, policy makers, and the general public, but there has been no corresponding increase in the public's ability to understand what these data mean. Furthermore, most measures have been defined without sufficient input from the community about which outcomes matter most to them. In both health care and education, measurement has limitations and must be complemented by the unique priorities of patients, families, and local communities. These priorities are best communicated by patients and communities themselves. Physicians need to empower communities to do this more effectively than the public education system has.

Shifting from school-based to teacher-based accountability has led to a decline in teacher collaboration and an increase in teacher isolation, and may prompt teachers to avoid working in schools with large populations of high-need students.⁴⁶ Physicians should resist the temptation to provide care simply to improve measured outcomes. Both education and health care differ in important ways from the world of industrial production, where performance management was devised. Just as teachers' VAM scores cannot capture their capacity for mentorship or their impact on individual student lives, process variables such as physician compassion, patient advocacy, and trustworthiness may matter more than measurable outcomes for many patients.

Finally, education level is a major social determinant of health. Poor educational achievement is an impediment to community health, so physicians should actively participate in debates about the future of public education in their own communities.

LESSONS FOR HEALTH POLICY LEADERS

The goal of our health care system should be to improve health, not just to improve health care. Heavy-handed management approaches that focus on a narrow list of outcomes will have unintended and unpredictable consequences. Such effects are clear in the experience of public education.

Efficiency and quality in both health care and education are different from business efficiency in important ways. Schools do not simply produce graduates. Outstanding teachers also inspire individual students to make a difference in the world. Similarly, physicians do more than treat disease and prolong life. They care for the sick and their families with compassion and have the capacity to change how people live within their communities. Using outcomes measurement and public reporting to create financial incentives and penalties for public education assumes that schools will respond to such risk with positive changes. It is not clear that this approach has worked as planned in education. In addition, performance management in education assumes a level of statistical sophistication that does not exist among teachers, school administrators, or the general public. Publicly reported outcomes data are useful only if people can understand them.

Outcomes measurement and quality improvement need to be an integral part of professional self-regulation by physicians and hospitals if we want to avoid diminishing physician professionalism. Although one might argue that physician self-regulation has not prevented unsustainable cost and poor population health, it has also produced remarkable achievements in American health care.

Direct input from patients, families, and communities, derived in partnership with their own physicians and hospitals is essential. Employers and government might buy health care, but patients consume it. Bypassing or ignoring their input about what matters most to them in the guise of improving outcomes may not improve the very outcomes that matter most to those we have promised to serve. Data about health outcomes need to be provided in the context of public priorities in a way that can be understood by community members.

Like the education of a student, the health of a patient cannot be isolated from broader societal issues of homelessness, hunger, and economic inequities. Outcomes measurement tells only part of the story, so we must search for a broader definition of quality than can be achieved with quantitative measurement.

CONCLUSION

The use of centrally mandated outcomes measurement in American public education is 2 decades ahead of similar efforts in health care. Current efforts by Medicare, Medicaid, private health plans, and health systems to apply performance management techniques to health care are accelerating. These tools could be useful if properly applied, so physicians should learn from teachers that opposition will not cause such efforts to go away. But the risk of applying these techniques simplistically is substantial. Health policy leaders should not discount the expertise of physicians in the heavy-handed way that educational policy leaders have in the case of teachers. Both educational policy and health policy need to recognize the limitations of business models in service sectors, reconnect with local communities, and more effectively listen to their voices.

Because the relationship between the health of a community and its educational level is well established, it is time to consider these 2 essential domains of community life together rather than separately.⁴⁷ Parents depend on a strong school system if they want their children to live long and happy lives, but students cannot learn if they are not healthy. Physicians seek to engage patients and communities in their own health, but without a sound educational foundation, the knowledge gap is often too wide. Communities cannot succeed without both education and health care, so spending more and more money on health care while cutting public education will not produce better health.

In her 2013 commentary in the *New England Journal of Medicine* comparing education and health care, Debra Malina has called for a countermovement to improve practices from the inside as a way to balance the weaknesses of top-down performance management in health care.⁴⁸ We agree that health care can learn

from the experience of educational reform. Perhaps health care can also provide positive lessons for the education system in how to enhance professionalism and improve outcomes "from the inside." We have emphasized the similarities between health care and education in this article, but this emphasis should not suggest that the differences are unimportant. The available data about health care quality and cost are much more extensive and detailed than are data sources about educational achievement. Health care accounts for a substantial portion of the US economy and is much more empowered to resist change than the schools have proven to be. Nevertheless, just and equitable systems for health care and public education are essential for our democracy.

To read or post commentaries in response to this article, see it online at <http://www.annfamned.org/content/15/1/71>.

Submitted March 29, 2016; submitted, revised, July 14, 2016; accepted July 28, 2016.

Key words: outcomes measurement; performance management; health policy; metrics; quality of health care; reference standards

References

1. Institute of Medicine. *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, DC: National Academy Press; 2001.
2. Fung CH, Lim YW, Mattke S, Damberg C, Shekelle PG. Systematic review: the evidence that publishing patient care performance data improves quality of care. *Ann Intern Med*. 2008;148(2):111-123.
3. Berenson RA, Pronovost PJ, Krumholz HM. *Achieving the Potential of Health Care Performance Measures*. Princeton, NJ: Robert Wood Johnson Foundation Urban Institute; 2013.
4. Kerr EA, Hayward RA. Patient-centered performance management: enhancing value for patients and health care systems. *JAMA*. 2013; 310(2):137-138.
5. Doran T, Roland M. Lessons from major initiatives to improve primary care in the United Kingdom. *Health Aff (Millwood)*. 2010;29(5): 1023-1029.
6. Campbell SM, Reeves D, Kontopantelis E, Sibbald B, Roland M. Effects of pay for performance on the quality of primary care in England. *N Engl J Med*. 2009;361(4):368-378.
7. McGlynn EA, Schneider EC, Kerr EA. Reimagining quality measurement. *N Engl J Med*. 2014;371(23):2150-2153.
8. Halladay JR, Stearns SC, Wroth T, et al. Cost to primary care practices of responding to payer requests for quality and performance data. *Ann Fam Med*. 2009;7(6):495-503.
9. Cassel CK, Jain SH. Assessing individual physician performance: does measurement suppress motivation? *JAMA*. 2012;307(24): 2595-2596.
10. Hong CS, Atlas SJ, Chang Y, et al. Relationship between patient panel characteristics and primary care physician clinical performance rankings. *JAMA*. 2010;304(10):1107-1113.
11. Werner RM, Bradlow ET. Relationship between Medicare's hospital compare performance measures and mortality rates. *JAMA*. 2006; 296(22):2694-2702.
12. Werner RM, Asch DA. The unintended consequences of publicly reporting quality information. *JAMA*. 2005;293(10):1239-1244.

13. Berenson RA, Kaye DR. Grading a physician's value—the misapplication of performance measurement. *N Engl J Med*. 2013;369(22):2079-2081.
14. Berg M, Ikkersheim D. *Measuring the Value of Healthcare Delivery: Cutting Through Complexity*. New York, NY: KPMG International Cooperative; 2013. Publication no. 130731.
15. Kerr EA, Hayward RA. Patient-centered performance management: enhancing value for patients and health care systems. *JAMA*. 2013;310(2):137-138.
16. Fletcher JM, Horwitz LI, Bradley E. Estimating the value-added of attending physicians on patient outcomes. Cambridge, MA: National Bureau of Economic Research; 2014.
17. Porter ME. What is value in health care? *N Engl J Med*. 2010;363(26):2477-2481.
18. Berwick DM, Nolan TW, Whittington J. The triple aim: care, health, and cost. *Health Aff (Millwood)*. 2008;27(3):759-769.
19. McDermott KA. *High-Stakes Reform: The Politics of Educational Accountability*. Washington, DC: Georgetown University Press; 2011.
20. Moynihan D. *The Dynamics of Performance Management: Constructing Information and Reform*. Washington, DC: Georgetown University Press; 2008.
21. Jacobsen R, Saultz A. Will performance management restore citizens' faith in public education? *Publ Perform Manag Rev*. 2016;39(2):476-497.
22. Rothstein R, Jacobsen R, Wilder T. *Grading Education: Getting Accountability Right*. New York, NY: Teachers College Press; 2008.
23. Kirst M, Wirt F. *The Political Dynamics of American Education*. 4th ed. New York, NY: Teachers College Press; 2009.
24. Jacobsen R, Saultz A, Snyder JW. When accountability strategies collide: do policy changes that raise accountability standards also erode public satisfaction? *Educ Policy*. 2013;27(2):360-389.
25. Sunderman G, Orfield G. Domesticating a revolution: No Child Left Behind and state administrative response. *Harv Educ Rev*. 2006;76(4):526-556.
26. Finn R. Shaping the system that grades city schools. *The New York Times*. <http://www.nytimes.com/2007/11/16/education/16lives.html>. Published Nov 16, 2007. Accessed Oct 25, 2016.
27. Hastings JS, Weinstein JM. Information, school choice, and academic achievement: evidence from two experiments. *Q J Econ*. 2008;123(4):1373-1414.
28. Newton XA, Darling-Hammond L, Haertel E, Thomas E. Value added modeling of teacher effectiveness: an exploration of stability across models and contexts. *Educ Policy Anal Arch*. 2010;18(23):1-27.
29. Harris DN. *Value-Added Measures in Education*. Cambridge, MA: Harvard Education Press; 2011.
30. Hull J. Trends in teacher evaluation: how states are measuring teacher performance. Center for Public Education, 2013. <http://www.centerforpubliceducation.org/Main-Menu/Evaluating-performance/Trends-in-Teacher-Evaluation-At-A-Glance/Trends-in-Teacher-Evaluation-Full-Report-PDF.pdf>. Published Oct 2013. Accessed Oct 25, 2016.
31. Goldhaber D. Exploring the potential of value-added performance measures to affect the quality of the teacher workforce. *Educ Res*. 2015;44(Special Issue):87-95.
32. Harris DN. Would accountability based on teacher value added be smart policy? An examination of the statistical properties and policy alternatives. *Educ Finance Policy*. 2009;4(4):319-350.
33. Taylor ES, Tyler JH. Can teacher evaluation improve teaching? *Educ Next*. 2012;12(4).
34. Booher-Jennings J. Below the bubble: 'educational triage' and the Texas accountability system. *Am Educ Res J*. 2005;42(2):231-268.
35. Jacobsen R, Saultz A, Snyder JW. Grading school report cards: issues for future research and practice. *Phi Delta Kappan*. 2013;95(2):64-67.
36. Saultz A, Aronson B, Murphy KM. Under pressure in Atlanta: school accountability and special education practices during the cheating scandal. *Teach Coll Rec*. 2016;118(4).
37. Nichols SL, Berliner DC. *Collateral Damage: How High-Stakes Testing Corrupts America's Schools*. Cambridge, MA: Harvard Education Press; 2007.
38. Zhao Y. *Catching Up or Leading the Way*. Alexandria, VA: ASCD; 2005.
39. Working With Teachers to Develop Fair and Reliable Measures of Effective Teaching. MET Project White Paper. Seattle, WA: Bill & Melinda Gates Foundation; 2010.
40. Darling-Hammond L. Can value-added add value to teacher evaluation? *Educ Res*. 2015;44(2):132-137.
41. Noell GH, Burns JL. Value-added assessment of teacher preparation: an illustration of emerging technology. *J Teach Educ*. 2006;57(1):37-50.
42. Jacobsen R. The voice of the people in educational policy. In: Sykes G, Plank D, Schneider B, eds. *Handbook of Educational Policy Research*. New York, NY: Rutledge; 2009:307-318.
43. Jacobsen R, Snyder JW, Saultz A. Understanding satisfaction with schools: the role of expectations. *J Public Adm Res Theory*. 2015;25(3):831-848.
44. Stone DA. *Policy Paradox*. New York, NY: W.W. Norton & Company; 2001.
45. Amrein-Beardsley A, Collins C. The SAS Education Value-added Assessment System in the Houston Independent School District: intended and unintended consequences. *Educ Pol Anal Arch*. 2012;20(12). <http://epaa.asu.edu/ojs/article/view/1096>. Accessed Jul 12, 2016.
46. Shavelson RJ, Linn RL, Baker EL, et al. *Problems With the Use of Student Test Scores to Evaluate Teachers*. Washington, DC: Economic Policy Institute; 2010. Briefing paper no. 278. <http://www.epi.org/publication/bp278/>. Published Aug 27, 2010. Accessed Oct 25, 2016.
47. Egarter S, Braverman P, Sadegh-Nobari T, Grossman-Kahn R, Dekker M. *Education and Health*. New York, NY: Robert Wood Johnson Foundation; 2011. Policy brief no. 5.
48. Malina D. Performance anxiety—what can health care learn from K-12 education? *N Engl J Med*. 2013;369(13):1268-1272.