

TE² Community Brief ■ September 2011

Making it Meaningful: Building a Fair Evaluation System

Vivian Yuen Ting Liu
& Laura Steen Mulfinger, Ph.D.




COMMUNITIES
for **TEACHING**
EXCELLENCE

“Our system of evaluating teachers has never been adequate. For too long and too often, teacher evaluation—in both design and implementation—has failed to achieve what must be our goal: continuously improving and informing teaching so as to better educate all students.”

Randi Weingarten, President, American Federation of Teachers

Of the many reforms associated with the growing teaching effectiveness movement, one of the most essential involves how teachers are evaluated. Right now, most teacher evaluation systems do not help teachers improve their teaching nor students improve their learning. This has serious consequences, particularly for low-income and minority students, who tend to have fewer resources outside of school and thus depend on teachers almost exclusively to help them advance academically.

Research confirms that teachers have the single greatest impact on student achievement—more than any other factor in the school (Rivkin, et al.). That’s why an effective evaluation system is an essential tool for helping improve teaching and learning, particularly for the most vulnerable students who, upon experiencing academic failure early in their schooling, may never quite catch up.

Fortunately, in recent years, researchers and policymakers have examined how best to address the failings of current teacher evaluation systems. In an effort to help schools and districts determine which teachers are effective and which need

support, research has identified a number of essential elements for a fair and comprehensive evaluation process. These elements incorporate a variety of performance measures, provide useful feedback, and make meaningful distinctions in teaching effectiveness. Perhaps most importantly, they help identify which teachers do the best job with struggling or high-need students, so that these students can be placed with the teachers who are most capable of helping them succeed.

What follows are an examination of the limitations of current teacher evaluation systems, a review of promising research-based reform strategies, and some examples from states and school districts that have begun to do this critical and challenging work. Putting in place systems that rely on research and best practices will go a long way toward enhancing student success and closing the achievement gap.

THE TROUBLE WITH “DRIVE-BY” EVALUATIONS

Whatever your profession—doctor, accountant, salesperson—you undoubtedly receive regular evaluations of your job performance. Ideally, your evaluation is conducted by

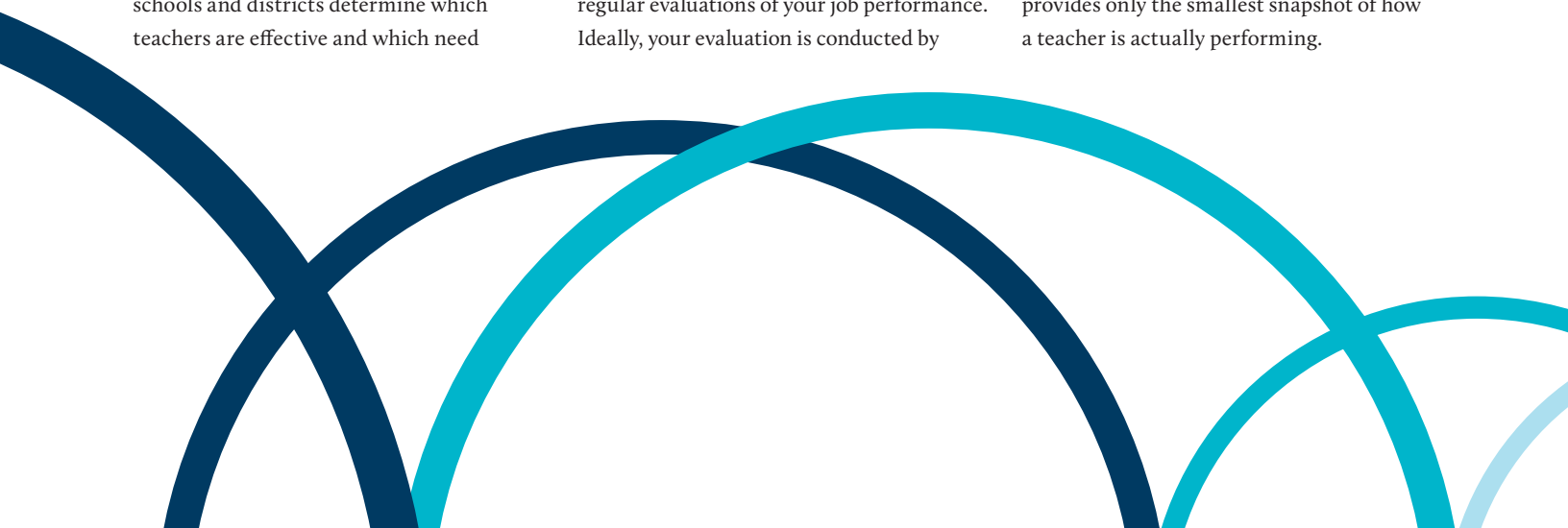
someone knowledgeable about your work, and who can provide meaningful feedback to help you progress. Depending on your field, your evaluation may also take specific performance targets into consideration: Did you meet your performance goals? Do your clients or patients express satisfaction with the service or care they’ve received?

The teaching profession should be no different.

However, in reality, it is very different. In a majority of school districts, teacher evaluations incorporate almost none of the criteria described above. Instead, they are usually based solely on brief, superficial classroom observations that offer little in the way of constructive feedback.

“Once, my evaluation consisted of an assistant principal walking into the computer lab where my students were working, giving me two thumbs up, and immediately walking out.”

This form of evaluation has been described by some as a “drive-by,” because it typically consists of a single, fleeting classroom visit by a principal or administrator, with limited training in evaluation, wielding a checklist of classroom conditions and teacher behaviors that often do not focus directly on the quality of instruction (Rothman and Toch). Not only does this offer little opportunity for meaningful feedback, it provides only the smallest snapshot of how a teacher is actually performing.



In fact, on average, most teachers are observed in their classrooms just once or twice per year, for fewer than 76 minutes total (Weisberg, et al.).

Teachers themselves have expressed deep dissatisfaction with this process. As one teacher described it, “Once, my evaluation consisted of an assistant principal walking into the computer lab where my students were working, giving me two thumbs up, and immediately walking out” (Ahlgren). In a 2008 survey of more than 1,000 teachers, only 26 percent reported that their most recent formal evaluation was “useful and effective.” A large number—41 percent—said it was “just a formality,” while another 32 percent said that at best it was “well-intentioned but not particularly helpful” to their teaching practices (Duffett).

After being observed in the classroom, teachers receive a performance rating. But many evaluation systems have only two ratings—satisfactory or unsatisfactory—which do nothing to distinguish excellent teachers from fair or poor ones. And even in districts that include more than two rating levels, the vast majority of teachers still receive a satisfactory rating – even in schools where fewer than half of the students are succeeding academically.

The New Teacher Project’s 2009 national study on teaching effectiveness, “The Widget Effect,” found that across the twelve districts they surveyed, 94 percent of teachers received one of the top two ratings.

Less than one percent were rated unsatisfactory. On average, only 10 percent of failing schools issued at least one unsatisfactory rating to a tenured teacher (Weisberg et al.).

In a 2008 survey of more than 1,000 teachers, only 26 percent reported that their most recent formal evaluation was “useful and effective.”

Virtually everyone – from teachers and union leaders to policymakers and community advocates – agrees that existing evaluation systems are largely one-dimensional and do little to help teachers grow and excel. Fortunately, there are research-based steps for creating a robust teacher evaluation system, one that can lead to improved teaching practices and increased student performance.

A FAIR SYSTEM BENEFITS EVERYONE

Before delving into the elements that comprise a successful evaluation system, it’s important to note that any new system is only as strong as the team that builds it. In other words, collaboration is essential.

As states and districts begin to reform their teacher evaluation systems, involving as many stakeholders as possible and encouraging feedback throughout the process is crucial.

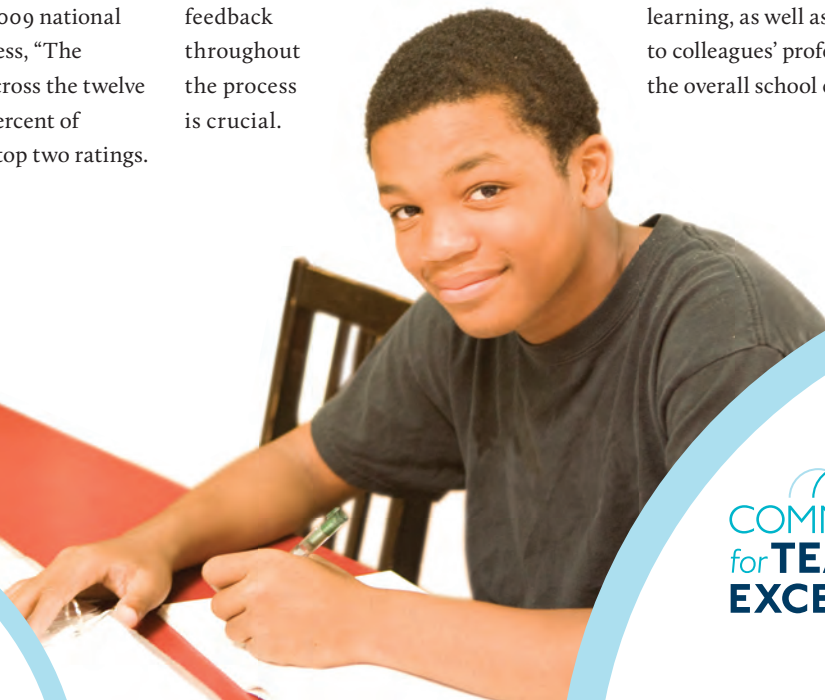
Gaining perspective through the lens of teachers, parents, district leaders, unions, administrators, and community members ensures that the process produces valuable information from all stakeholders, is fair and inclusive, and finally, that the evaluation system will be successful and sustainable.

In addition to being developed collaboratively, evaluation systems should include the following elements:

Multiple Measures. Given the complexity of teaching, observations alone cannot capture the entirety of a teacher’s impact on student learning. This is why a number of methods are needed—commonly referred to as “multiple measures”—to fairly evaluate how well a teacher is performing. Along with observations, multiple measures may include a variety of research-based tools and materials such as:

- Student achievement data
- Samples of student work
- Sample lesson plans and assignments
- Student and parent surveys
- Principal and peer evaluations
- Teacher’s self-assessments

By incorporating multiple measures, the evaluator can glean a far more accurate picture of a teacher’s impact on student learning, as well as his or her contribution to colleagues’ professional development and the overall school community.



Research conducted by the Measures of Effective Teaching (MET) project, a multi-year study of seven urban school districts including New York City, Charlotte-Mecklenburg, Hillsborough County, Memphis, Dallas, Denver, and Pittsburgh confirms that evaluations using multiple measures can indeed draw a more complete picture of teaching ability, and help school leaders offer specific diagnostic feedback for improvement (Atkinson et al.).

It is also necessary to ensure that the measures used are valid and reliable—that they measure what they are designed to measure (one wouldn't use a ruler to determine how much someone weighs),

produce consistently accurate results, and are aligned with clearly defined professional teaching standards. It's essential that the principals, administrators, and teachers using these measures are thoroughly trained to perform them objectively and consistently.

Differentiation. Rating teachers as either satisfactory or unsatisfactory does not meaningfully distinguish one from the next, so schools are unable to determine which teachers are extraordinary, average, or in need of improvement. A fair evaluation system should include a teacher performance rating scale with no fewer than three levels, but preferably more. For example, a system with four rating levels could rate teachers as “highly effective,” “effective,” “needs improvement,” or “ineffective.”

Increasing the number of categories gives teachers a clear picture of their current performance, while also allowing for meaningful distinctions between each level.

“The Widget Effect” emphasizes this point: “The differentiation of teacher effectiveness should be a priority for school administrators and one for which they are held accountable. Administrators must receive rigorous training and ongoing support so that they can make fair and consistent assessments of performance against established standards and provide constructive feedback and differentiated support to teachers” (Weisberg, et al.).

Feedback. The point of differentiating between teachers is not to punish some and reward others; the goal is to help teachers improve. Feedback is fundamental to this process. Utilizing a multiple-level rating system informed by multiple measures of performance provides a more accurate depiction of a teacher's effectiveness.

The point of differentiating between teachers is not to punish some and reward others; the goal is to help teachers improve.

This serves as an important diagnostic tool for identifying practices that produce positive results and determining areas that need attention and assistance. In other words, more accurate assessments of teachers allow for more targeted and useful recommendations on how to improve their practices.



ABOUT THE TE² COMMUNITY BRIEFS

Community: A group of people with a common characteristic or interest living together within a larger society (Merriam-Webster Dictionary).

Communities for Teaching Excellence believes that every community, and all of its members, including teachers, parents, community-based organizations, and interested citizens, benefit from inclusive and meaningful engagement on education issues.

Guided by this fundamental principle, we have created a series of documents called TE² Community Briefs. The series consists of several brief, informative, and research-based pieces on a variety of teaching effectiveness and equity (TE²) topics, such as fair evaluation, seniority, strategic compensation, and professional development, among others.

After an extensive review of the work being done around TE² we determined that a majority of the work tends to be either lengthy academic research reports or short op-eds or blogs. Few, if any documents are intended to educate and engage the community. The TE² Community Briefs fill this void, by providing the reader with a comprehensive, research-supported summary of select TE² topics, including examples of states and school districts that are doing this work.

This Community Brief is the second in the series to inform the community about the emerging TE² movement. Specifically, this brief examines the state of current evaluation systems and how they fail to adequately differentiate between teaching practices that produce positive results or those that need attention and assistance. It further reviews promising reform strategies, and includes examples from states and school districts that have begun to do this critical and challenging work.

Teachers themselves are eager for feedback that helps them understand how they're doing and what they can do to further develop and strengthen their skills. A 2011 study showed that less experienced teachers in particular want more feedback on their instruction and classroom impact. In fact, 75 percent of teachers in the study said they prefer having a principal who "frequently observes my classroom and gives me detailed feedback on how I'm doing" as opposed to having a principal who "conducts formal observations of my teaching only once a year and gives me only general feedback" (Cogshall, et al.).

Student Performance. It is impossible to truly evaluate a teacher without examining the academic performance of his or her students. That's why examining data from state assessments is an essential piece of the teacher evaluation puzzle. These assessments help states, districts, schools, teachers, and parents know whether students are mastering the standards and

curriculum, and if they are performing at the appropriate grade level.

When student performance measures are part of the evaluation, it is essential that the data are considered appropriately in order to ensure fairness. Rather than looking only at how high students score on a given test, evaluations should also consider the amount of progress—or growth—that students make from one year to the next.

Most states and districts rely on research-based "student growth measures" (also called "value-added modeling" or "student growth percentiles") to help determine teaching effectiveness. These measures hold a teacher accountable only for the growth in learning that takes place while the student is enrolled in his or her classroom. Student growth measures, and value-added models in particular, can also account for factors over which teachers and schools have little control, such as poverty or language ability.

Student growth measures hold a teacher accountable only for the growth in learning that takes place while the student is enrolled in his or her classroom.

Consequences. Information from evaluations should be used to help guide important decisions such as tenure and placement. As outlined in "Evaluation 2.0," a blueprint for evaluation reform by The New Teacher Project, the evaluation process must have meaningful implications—both positive and negative—in order for it to support the teaching workforce and improve outcomes for students. Evaluations should inform decisions about an individual teacher's professional development or opportunities for promotion, as well as decisions involving tenure, compensation, dismissal and the like.



TURNING RESEARCH INTO REALITY

Despite overwhelming evidence that most evaluation systems need to be overhauled, there are still concerns about the feasibility of replacing old systems with new ones. But several states and school districts have recently demonstrated that creating evaluation systems based on research and collaboration is entirely within reach.

Two examples are the state of Delaware and the Houston Independent School District. Both have redesigned their evaluation systems based on many of the recommendations outlined above—involving a variety of stakeholders, using multiple measures including student performance data, instituting broader rating scales, and providing feedback to teachers. Their efforts are instructive and can be used to encourage and inform the work of other states and districts.

Every teacher is evaluated at least four times throughout the year and receives feedback on student progress, classroom observations, performance, and professional growth.

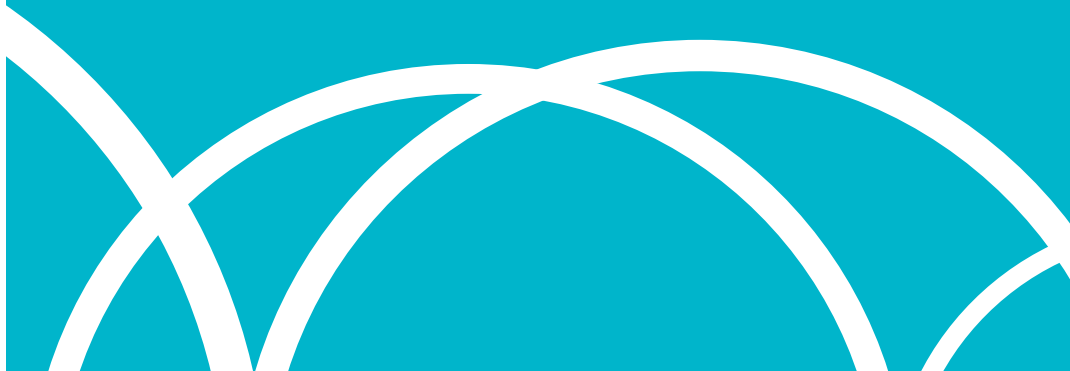
**DELAWARE:
COMPREHENSIVE & COLLABORATIVE**

Delaware’s effort to reform its evaluation process provides an example of transparency, collaboration, and a commitment to ongoing improvement. The Delaware public school system is relatively small, with about 130,000 students (approximately half of whom are students of color), and 8,000 teachers in 200 schools. Close to 50% of all students come from low-income families. In 2007, the state implemented the Delaware Performance Appraisal System II (DPAS)—a comprehensive teacher evaluation system with the goals of “improving quality assurance and professional growth.” As part of this new process, DPAS identified five standards of effective teaching, each of which is given equal weight:

- Planning and preparation
- Classroom environment
- Instruction
- Professional responsibilities
- Student assessment

Every teacher is evaluated at least four times throughout the year and receives feedback on student progress, classroom observations, performance, and professional growth. The system also utilizes multiple measures of performance assessment including reviews of lesson plans, parent surveys, student work portfolios, and communication with families. In addition, teachers must demonstrate that they use student assessment and accountability data to measure student progress.

Since initiating DPAS, the State Department of Education has partnered with teachers and administrators to help improve the system. The system is evaluated annually and refinements are put in place. In the 2011-2012 school year, DPAS II is expanding its use of student growth measures and will incorporate recommendations developed by a “Stakeholder Working Group,” made up of teachers, parents, administrators, community members, and researchers.



HOUSTON INDEPENDENT SCHOOL DISTRICT: BALANCED & THOUGHTFUL

In contrast to Delaware, the Houston Independent School District (HISD) is vast. As the nation's 7th largest public-school district, HISD serves more than 200,000 students in 300 schools and employs over 13,000 teachers. A significant percentage of the district's students are students of color (Latinos are 62%, 27% are Black); and 80% come from low-income families.

HISD demonstrates how teacher evaluations can use multiple measures and incorporate student achievement data in a balanced and thoughtful way. Following a six-month collaborative and open process during which HISD actively sought input from more than 2,600 teachers, 500 school administrators, 500 parents, and 600 community members, the HISD Board of Education approved a new assessment system for teachers in May 2011 to be implemented in the 2011-2012 school year.

This new evaluation system includes a four-level rating scale and gives equal weight to three criteria: student performance, instructional practice, and professional expectations.

The system will use multiple measures to assess performance within these three criteria including student achievement data, classroom observations, reviews of lesson plans, teacher attendance, communication with parents, collaboration with colleagues, and efforts to seek input for improving performance.

Evaluators will hold annual conferences to provide specific feedback to all teachers, and will conduct classroom observations and walk-throughs—both of which will be followed by written feedback and in-person meetings. Under the old system, about half the district's teachers were exempt from annual observations based on their experience and performance.

Ensuring an effective teacher in every classroom is one of HISD's key strategies for increasing achievement. As one teacher who helped design the new system noted, "I need to be ready to be valued by what my children are learning and not simply by how much I love them; I need to be ready to realize that although teaching is my calling in life, it is not my right" (Mellon).

Overhauling current teacher evaluation systems and designing fair evaluation processes are challenging, costly, and time consuming efforts. Yet both Delaware and Houston have taken advantage of current research and best practices to improve their systems.

Research concludes that, "teacher evaluations pay much larger dividends when they also play a role in improving teaching" (Toch and Rothman). Teachers are entitled to an evaluation process that reflects their performance and contribution to students' success and the greater school community, and supports their continuous development. All students deserve effective instruction and teachers who help them strive for the best. Developing and implementing fair evaluation systems is key to making this happen.

All students deserve effective instruction and teachers who help them strive for the best. Developing and implementing fair evaluation systems is key to making this happen.

WHAT CAN YOU DO?

Find out about the teacher evaluation system in your state or district—does it consider student performance or use multiple measures? Become informed about local efforts to improve teaching effectiveness.

For more information on the research or to get involved in the TE² movement contact Communities for Teaching Excellence at www.4teachingexcellence.org. We're working to ensure effective teaching for every student, in every classroom, every year.



SOURCES & CREDITS

Ahlgren, T. (2011). *Testimony Before the Indiana Senate*. Retrieved August 25, 2011 from <http://www.teachplus.org/page/news-6/news/may-2011-65.html>.

Atkinson, M., Caldwell, N., Ferguson, R., Gitomer, D., Hirsch, E., McCaffrey, G., Pea, R., Phelps, G., Ramsdell, R., and Staiger, D. (2010). *Learning About Teaching: Initial Findings from the Measures of Effective Teaching Project*. (Seattle, WA: Bill & Melinda Gates Foundation).

Cogshall, J. G., Behrstock-Sherratt, E., and Frill, K. (2011). *Workplaces That Support High-Performing Teaching and Learning: Insights From Generation Y Teachers*. (Washington, D.C.: American Institute for Research and American Federation of Teachers), p. 7.

Duffett, A., Farkas, S., Rotherham, A. J. and Silva, E. (2008). *Waiting to be Won Over: Teachers Speak on the Profession, Unions and Reform*. (Washington, D.C.: Education Sector Reports), p. 3.

Goe, L., Bell, C., and Little, O. (2008). *Approaches to Evaluating Teacher Effectiveness: A Research Synthesis*. (Washington, D.C.: National Comprehensive Center for Teacher Quality), p. 8.

Hamill, S. (2011). *Forging a New Partnership: The Story of Teacher Union and School District Collaboration in Pittsburgh*. (Washington, D.C.: The Aspen Institute Education & Society Program), p. 9.

Houston Independent School District (2011). *New HISD Teacher Appraisal System Wins Board Approval*, May 12. Retrieved August 7, 2011 from http://www.houstonisd.org/HISDConnectEnglish/Images/PDF/BOEM051211_PR.pdf.

Jerald, C. and Van Hook, K. (2011). *More than Measurement: The TAP System's Lessons Learned for Designing Better Teacher Evaluation Systems*. (Washington, D.C.: National Institute for Excellence in Teaching).

Measuring Effective Teaching Project (2010). *Learning about Teaching: Initial Findings from the Measuring Effective Teaching Program*. (Seattle, WA: Bill and Melinda Gates Foundation).

Mellon, E. (2011). "HISD Approves Tougher Evaluations for Teachers." *The Houston Chronicle*, May 12. Retrieved August 2, 2011 from <http://www.chron.com/dispatch/story.mpl/metropolitan/7563030.html>.

Rivkin, S. G., Hanushek, E. A., and Kain, J. F. (2005). "Teachers, Schools, and Academic Achievement." *Econometrica*, Vol. 73, No. 2, 417– 458.

Rothman, R. and Toch, T. *Rush to Judgment: Teacher Evaluation in Public Education*. (Washington, D.C.: Education Sector.), p. 2.

Schacter, J., Thum, Y. M., & Zifkin, D. (2006). How Much Does Creative Teaching Enhance Elementary School Students' Achievement? *Journal of Creative Behavior*, 40(1), 47–72.

The New Teacher Project (2010, October). *Teacher Evaluation 2.0*. (Brooklyn, NY).

Weingarten, R. (2010). *A New Path Forward: Four Approaches to Quality Teaching and Better Schools*. Retrieved August 7, 2011 from http://www.aft.org/pdfs/press/sp_weingarten011210.pdf.

Weisberg, D., Sexton, S., Mulhern, J., and Keeling, D. (2009). *The Widget Effect: Our National Failure to Acknowledge and Act on Differences in Teacher Effectiveness*. (Brooklyn, NY: The New Teacher Project).

PHOTO CREDITS: Front Cover © Istockphoto.com | Page 3 © Devin Ford, devinford.com | Page 4 and Back cover © Istockphoto.com

OUR MISSION

To improve students' academic achievement and their future outcomes by empowering communities to advocate for an effective teacher for every student, in every classroom, every year.

ACKNOWLEDGEMENTS

Many thanks to Maria Casillas and Laura Schifter, for their comments, edits, and feedback, and overall willingness to improve the quality of the publication.

Communities for Teaching Excellence,

A project of Rockefeller Philanthropy Advisors
448 South Hill St., Suite 408
Los Angeles, CA 90013 | 213.489.3002
[4teachingexcellence.org](http://www.4teachingexcellence.org)