2013

LEAP YEAR

Assessing and Supporting Effective First-Year Teachers

TNTP reimagine teaching

Acknowledgements

This report was a collaborative effort involving individuals across TNTP. Jennifer Mulhern and Erin Grogan led writing and research, with the support of Debra Wexler. Three members of our leadership team—Timothy Daly, Karolyn Belcher and David Keeling—shared valuable insights and helped shape our paper. Kathleen Carroll and James Cho led writing and design. Ana Menezes, Greg Vincent and Layla Avila contributed their expertise throughout this project.

We are deeply indebted to the scores of TNTP staff, including our site directors and partners, who worked tirelessly to fairly and completely implement ACE in our 15 participating programs. They took on a daunting challenge with persistence, creativity and incredible energy, and their work helped to push our thinking and provide our teachers with rich feedback and unwavering support.

Finally, we thank the school districts, principals and teachers who participated in ACE, who brought their very best efforts and most heartfelt commitment to this work.

LEAP YEAR

Assessing and Supporting Effective First-Year Teachers

SUMMARY 1

BACKGROUND: THE NEED FOR ACE 3

ACE: AN OVERVIEW 4

FINDINGS 9

RECOMMENDATIONS 16

APPENDIX 18

"UNTIL YOU SEE A PERSON IN A CLASSROOM WITH KIDS,

you can't know how they'll really do."

—ASSISTANT DISTRICT SUPERINTENDENT

A TEACHER'S MOST IMPORTANT YEAR IS THE FIRST YEAR.

The most important year of a teacher's career is the first year. It's the time when teachers are making the daunting leap from preparation to practice and, for many, a period of rollercoaster professional growth. Extensive research has also shown that a teacher's initial performance is a meaningful predictor of future performance—far more so than commonly used proxies like academic credentials or pathways into the profession.[1] Teachers who make a strong start are much more likely to become and remain strong educators over time.

However, rather than focusing intensively on whether the habits and skills developed during this critical period are the right ones and carefully monitoring performance for early indicators of long-term success, it has become common practice to disregard a teacher's first year—to treat it like a warm-up lap. The policies and practices of most schools and districts assume that new teachers uniformly struggle, and their performance is rarely assessed with an eye to how they are likely to fare in the future, with more experience.

The first year of teaching must be reconsidered. It is not a warm-up, but an opportunity to provide focused critical feedback, cultivate emerging strengths and make careful assessments about whether new teachers should be developed into career educators or encouraged to pursue another career.

This paper summarizes TNTP's efforts to treat the first year of teaching with the care it deserves, through the Assessment of Classroom Effectiveness (ACE).

ACE is a multiple-measures evaluation system designed specifically for new teachers. It was introduced in the 2011–12 school year in a subset of TNTP Academy programs, through which we train and recommend new teachers for state certification.[2]

ACE uses classroom observations, student surveys, student academic growth data and principal ratings to track participants' progress toward a clear performance standard: classroom performance and development that show a teacher has mastered essential skills and is on the path to becoming effective. Through ACE, we have begun to make teacher certification decisions based on classroom performance and growth, instead of paper qualifications and coursework.

The first year of teaching must be reconsidered. It is not a warm-up, but an opportunity to cultivate strengths and make careful assessments.

Approximately 1,000 teachers in 15 programs participated in ACE in its first year. While ACE continues to evolve, it has already given us a wealth of information about the performance of first-year teachers and yielded significant insights, including:

- **New teachers perform at different levels and grow at different rates.** Not all first-year teachers struggle; in fact, some of them are high performers on all ACE measures, including student surveys and academic growth.
- **A new teacher's initial performance predicts his or her future performance.** Teachers who are performing poorly in their first year rarely show dramatic improvement in their second year.
- **Multiple measures tend to point to the same conclusion about a teacher's potential.** Teachers who do well on one ACE measure earn generally high scores overall. However, principals tend to rate first-year teachers positively, even when other measures are far less positive.
- **First-year teachers who are purposeful, responsive and able to focus on student understanding develop more quickly.** These core skills are common among new teachers who develop rapidly.

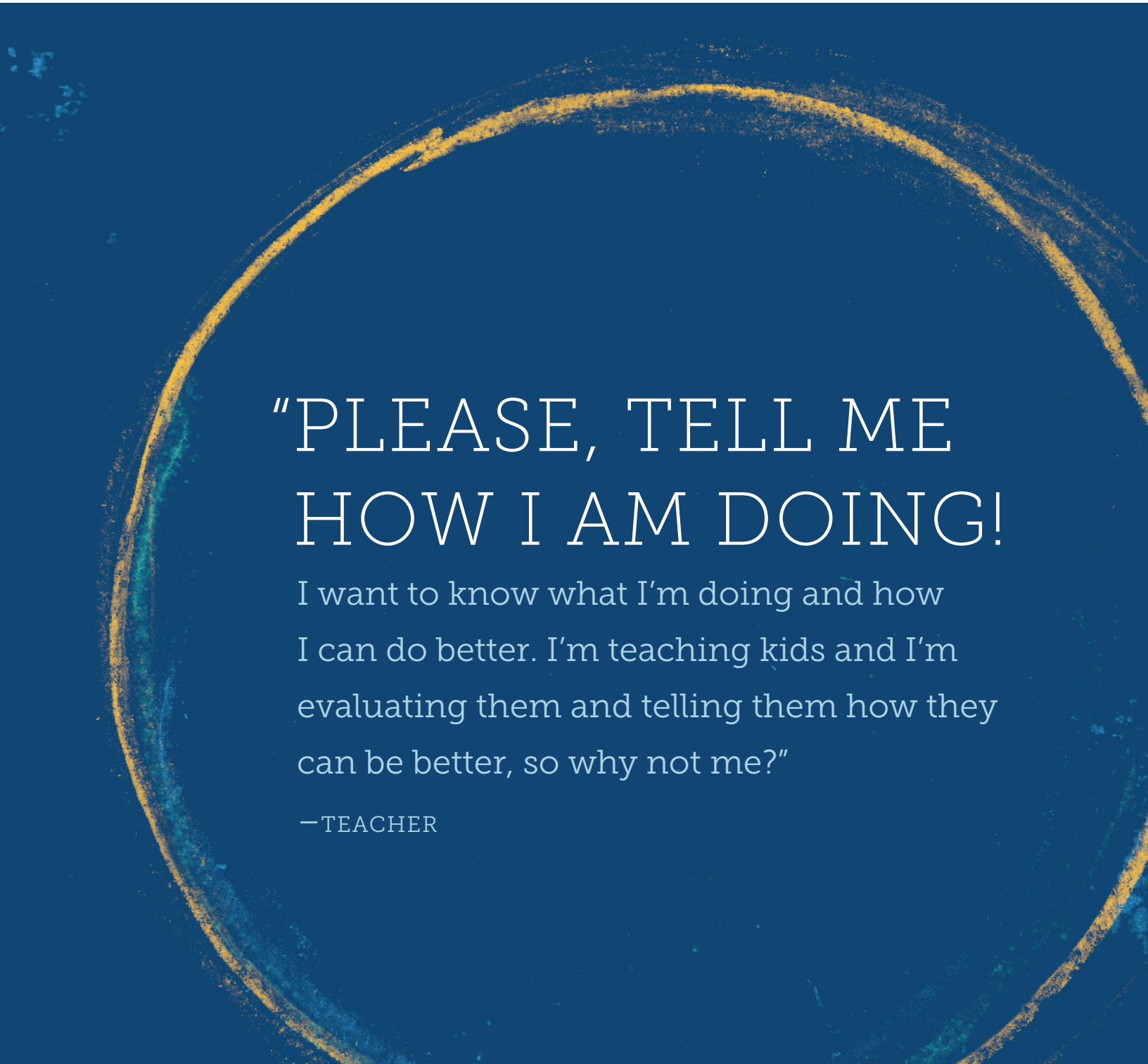
[1] See Atteberry, A., Loeb, S., & Wyckoff, J. (2013). Do first impressions matter? Improvement in early career teacher effectiveness (Working Paper 90). Washington, DC: CALDER; Clotfelter, C., Ladd, C., & Vigdor, J. (2007). Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. Economics of Education Review, 26 (6), 673–682; Kane, T., Rockoff, J., & Staiger, D. (2008). What does certification tell us about teacher effectiveness? Evidence from New York City. Economics of Education Review, 27 (6), 615–631; Gordon, R., Kane, T. & Staiger, D. (2006). Identifying effective teachers using performance on the job (Discussion Paper 2006-01). Washington, DC: Brookings Institution; and Staiger, D., & Rockoff, J. (2010). Searching for effective teachers with imperfect information. Journal of Economic Perspectives, 24 (3), 97–118.

[2] TNTP Academy programs train and certify teachers recruited by TNTP and other alternate-route certification programs.

This paper describes ACE and these findings in greater detail. In addition, we offer three recommendations for policymakers and education leaders based on our experiences with ACE to date:

- Connect teacher certification and on-the-job performance.
- Use classroom observations and student surveys when value-added data are unavailable.
- Focus new teachers on core skills, and look for rapid growth.

Not only is it possible to make meaningful determinations about the work of new teachers, it is essential. The alternative is to continue to certify struggling teachers who are unlikely to improve, and overlook powerful development opportunities for teachers who are most able to advance quickly. Both do a disservice to teachers and students.



BACKGROUND: THE NEED FOR ACE

Teacher recruitment has been TNTP's core work since our organization was founded in 1997. We operate a range of programs that bring high-quality teachers to high-need schools. We have recruited or trained approximately 50,000 teachers for urban classrooms, mainly in hard-to-staff subject areas, such as special education, math, science and bilingual education. The majority of these teachers have been prepared by our highly selective Teaching Fellows programs, which attract accomplished career changers and college graduates to teach in cities such as Baltimore, Chicago, Fort Worth, Memphis, Nashville, New Orleans, New York and Washington, D.C.[3]

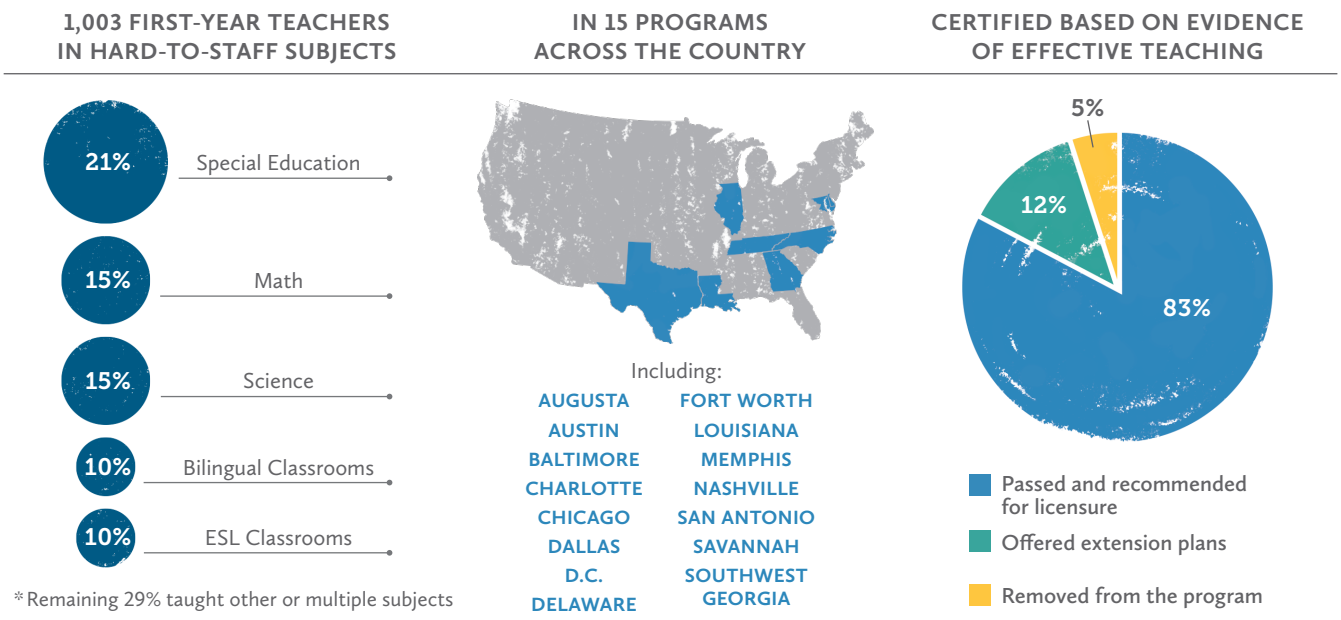
In recent years, as we sought new information about our teachers' performance, we found that, on average, they were roughly equal in effectiveness to teachers trained through other pathways. While we were reassured that our programs could produce comparably effective teachers despite their accelerated training schedules and focus on high-need schools and subjects, we were dissatisfied with the outcomes. It was clear that we were struggling with the same problem facing others in the field: our recruitment, selection and training models were not able to ensure consistently effective teaching. To achieve our mission, we had to do better.

This realization led us to refocus our efforts. We built and implemented ACE to hold ourselves and our teachers accountable for their classroom performance. We required teachers to master a core group of skills that appear to be critical to future success, and to show they could quickly learn and improve their classroom practice after receiving rigorous feedback and coaching.

We built and implemented ACE to hold ourselves and our teachers accountable for their classroom performance.

Teachers who progressed to our standard were recommended for certification. Those who struggled were given time and support to improve, but in the end were denied certification and removed from the classroom if they did not meet performance standards. In the 2011–12 school year, 83 percent of teachers who participated in ACE earned certification, 12 percent were put on extension plans and given another year to improve, and 5 percent were removed from the program.

FIGURE 1 ACE PARTICIPANTS, 2011–12



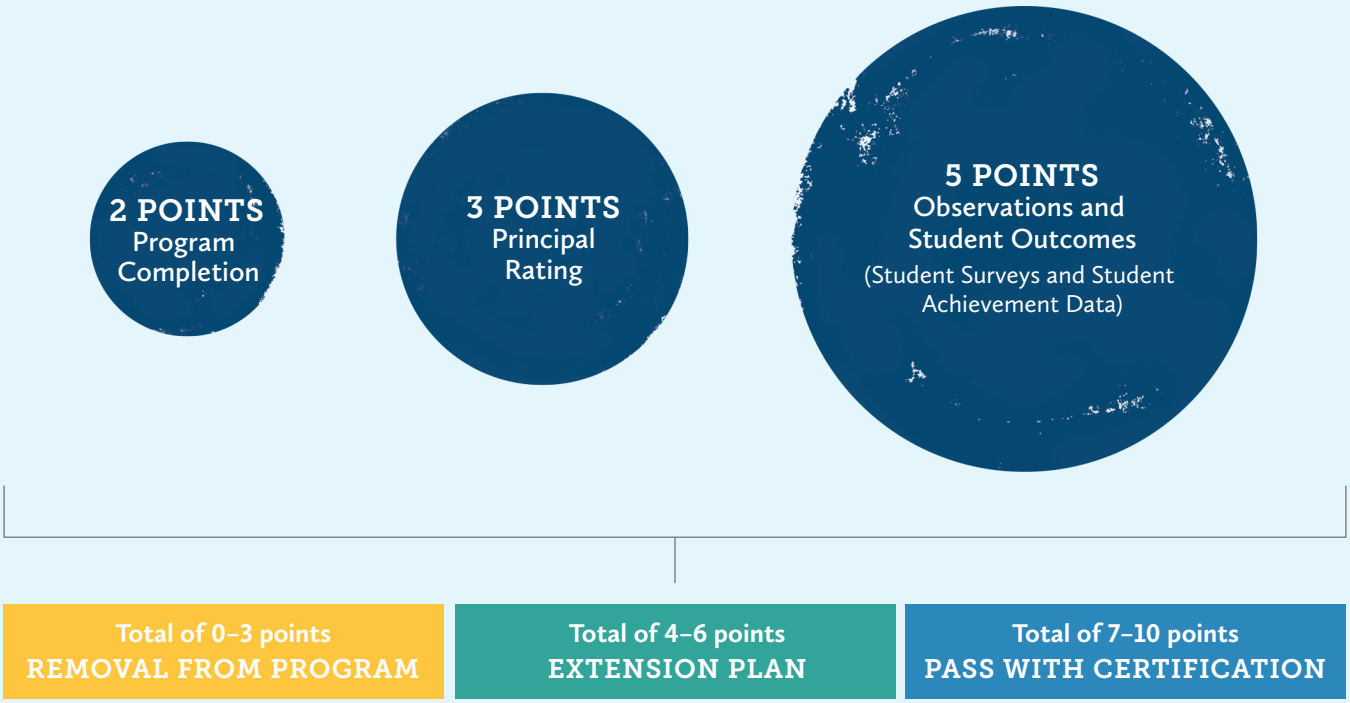
[3] For more information, see http://tntp.org/what-we-do/training

ACE: AN OVERVIEW

ACE is a multiple-measures evaluation system designed to monitor and support a teacher's effectiveness during his or her first year in the classroom. In the first year of implementation, classroom observations and student learning outcomes—based on student surveys and achievement data—made up 50 percent of a teacher's ACE score. Principal ratings made up 30 percent. The remaining 20 percent came from satisfying program requirements, including completing coursework, passing state licensure tests and demonstrating TNTP's professional values, which include setting high expectations, demonstrating flexibility and modeling a culture of achievement.

Altogether, teachers could earn up to 10 points. Teachers who earned seven points or more were those we considered ready for certification: they were showing an appropriate rate of development toward leading focused, purposeful classrooms, and were well on track to becoming effective. Participants' value-added data, where available, showed that teachers who scored in this range on ACE tended to meet or exceed academic growth standards in their state or district, and were on track to outperform the average teacher.[4]

FIGURE 2 ACE SCORING, 2011–12



Program Completion: satisfactory completion of coursework, passing licensure tests and adhering to TNTP's professional values. Principal Rating: candidates rated as "better than," "about the same as" or "worse than" the average first-year teacher. Student outcomes: scores on up to three classroom observations, student survey results and value-added data, where available. For more information visit http://tntp.org/what-we-do/training/tntp-academy. Source: TNTP.

[4] Value-added data from state assessment systems were available for some ACE participants in Tennessee and North Carolina. In Tennessee, of the 42 participants who passed ACE and received Tennessee Value-Added Assessment System (TVAAS) scores, 77% of ratings met or exceeded the state growth standard, including 43% exceeding expectations. In Charlotte, of the 12 participants who passed ACE and had value-added scores, 92% met or exceeded the state growth standard, including 34% exceeding expectations.

COMPONENTS OF ACE

CLASSROOM OBSERVATIONS

Participating teachers received three 45- to 60-minute observations during the year from 89 trained external observers. Observers assessed a set of core teaching competencies using a standard rubric and gave teachers ratings on each competency on a 5-point scale, leading to an overall rating on the same scale.[5] Our rubric was based on nine key skills teachers must master in order to become effective, such as:

- **Facilitating organized, student-centered, objective-driven lessons**
- **Promoting active participation and high academic expectations**
- **Building a positive and respectful classroom environment**

Our observations allowed us to track individual teachers' progress on critical skills throughout the year, and were positively correlated with other measures.[6]

Lesson learned: Build a strong bench of well-trained observers.

To ensure fairness and reliable overall scores, we gave each observer 30–40 hours of training and assigned multiple observers to visit teachers whenever possible. We also regularly asked observers to rate sample videos of lessons as a norming exercise, to ensure that observers rated teacher performance consistently and that scores were in line with our standards. About 5 percent of observers were unable to meet our standards after being given additional support and training and were excused, and their ratings were not accepted for use in ACE.

When assessing tradeoffs between adding observers and adding observations, the evidence is fairly clear—adding observers gives the greater boost to reliability. Giving teachers three different observers, instead of the same observer for each round, significantly increases the reliability of observations.[7]

PRINCIPAL RATINGS

Near the end of the school year, principals were asked to rate each teacher participating in ACE as "better than," "about the same as" or "worse than" the typical first-year teacher in their district.

Lesson learned: Beware of inflation in principal ratings.

We gave principals' ratings considerable weight, because principals have an important perspective on each teacher's performance: ideally, they see their classroom practice regularly throughout the year. However, we found that principal ratings were generally higher than other evidence would suggest, and tended to reflect low expectations for first-year performance. This was in line with our past experience with principal ratings, which we have often found to be inflated.[8] We concluded that a change was necessary. In the second year of ACE, we modified our scoring model to give principal ratings less weight.[9]

[5] The SY2011-12 observation rubric is available for download at: http://tntp.org/assets/documents/TNTP_ACE_Observation_Framework_2012.pdf

[6] Observations were positively and significantly correlated to the other ACE measures: principal ratings ($r=0.32$, $p<0.001$); student surveys ($r=0.32$, $p<0.001$); and value-added estimates ($r=0.22$, $p<0.01$). For intercorrelations of all measures, see Figure 6.

[7] The reliability of ACE observations was similar to that found in previous studies using different observation rubrics. Similar to the Measures of Effective Teaching study (Ho & Kane, 2012), we found that about 36% of the variation in performance was attributable to persistent differences between teachers. Reliability was highest for teachers with three unique observers ($\alpha=0.63$), but still reasonably high even when a teacher was observed by the same person for all observations ($\alpha=0.44$).

[8] Principal ratings track teacher evaluations generally, which historically have rated virtually all teachers as satisfactory or better, and rated less than 1 percent of teachers as ineffective. For more, see Weisberg, Sexton, Mulhern & Keeling (2009). *The Widget Effect: Our National Failure to Acknowledge and Act on Differences in Teacher Effectiveness.*

[9] See Appendix, "ACE in its Second Year."

STUDENT OUTCOMES

Student Surveys

We partnered with YouthTruth to administer surveys of student perceptions of the teachers participating in ACE and factored the results into their final ACE scores.[10] The survey included questions regarding the teacher's ability to challenge students and create a focused classroom environment (both of which are especially predictive of student learning[11]) as well as questions about whether students felt cared for and respected.

Surveys were available in multiple languages and were administered in late March and April, and teachers received their results and a formative report in July. During the 2011–12 school year, students completed surveys for 485 teachers, or about 48 percent of all ACE candidates. Students in grades K–2, in very small classes, or with severe disabilities did not take part, and some school districts declined to administer the surveys.

We included student surveys for a few reasons. Student achievement data were not readily available for many of our teachers, and we felt it was critical to include a measure of their impact on students in ACE. Student surveys were a way for us to access that information, by asking the people who knew their teaching best—students. Students see teachers throughout the year, whereas observations are based on snapshots of teaching. In addition, sharing student feedback with teachers was another way to provide significant insights into development for some of our teachers.

Lesson learned: A relatively small group of students can provide reliable data.

In the 2011–12 school year, a substantial number of teachers (between 150 and 200) were required to administer student surveys to eligible classes but did not collect sufficient data to meet program standards. Because surveys were to be administered in late spring, around the same time as many state assessments, some teachers encountered scheduling difficulties.

Our analysis indicated that survey data are reliable with as few as two classes participating, and we revamped our program accordingly.[12] We now administer surveys to only two classes, down from four, with a minimum of 15 participants total for each teacher, minimizing the amount of instructional time set aside for surveys. We also stepped up communications efforts around surveys and shifted their dates from late spring to February–March for the 2012–13 school year. We hope to improve survey completion rates, while also reducing the data collection burden on teachers and students.

[10] We used an adapted version of survey items used in the MET Project. To the best of our knowledge, TNTP is the first organization in the country to use student surveys as a substantial input into teacher certification decisions.

[11] Bill & Melinda Gates Foundation (2010). *Learning about Teaching: Initial Findings from the Measures of Effective Teaching Project.*

[12] We assessed the reliability coefficients under a variety of scenarios, including varying student participation rates from 5 through 30 students in one, two, or four class sections. Reliability coefficients in almost all scenarios were high, ranging from α=0.53 for 5 students in one elementary classroom to α=0.81 for 30 students in two secondary classrooms. There were diminishing returns to reliability with the addition of extra students and classrooms; for SY2012-13, we settled on a minimum of 15 students, in one classroom for elementary teachers (α=0.68) and in two classrooms for secondary teachers (α=0.76).

Student Achievement Data

The most important element of a teacher's performance is student academic growth. For this reason, TNTP is committed to including measures of academic growth in ACE wherever possible. However, most participating teachers had placements in grades or subjects where standardized testing data were not available, or where data did not lend themselves to the calculation of student growth or value-added models.

Approximately 14 percent of teachers had value-added scores factored into their ACE decisions. For those without value-added scores, student outcomes scores were based on classroom observations and student survey results.

Lesson learned: Assemble the best set of multiple measures possible and use those in the short term, while working on better assessments in the long term.

We initially tried to fill in the gaps in available data by developing our own assessments; however, we also needed data that could tell us how ACE teachers compared to other first-year teachers across the country. We were unable to find enough comparative data, and our program involved relatively small numbers of teachers across multiple subjects, grade levels and school districts, all using different assessments. Therefore, we determined that ACE involved too few teachers in its first year to make statistically reliable comparisons.

However, before we recognized the flaw in our approach, many teachers had invested time collecting baseline assessment data. We wish we could give them that time back.

We shifted to a more flexible approach, using value-added where comparable data were readily available, and weighting other measures more heavily where such data were not. Because we had multiple measures in place, we feel we were able to hold all of our teachers to the same high standard, even though every participant was not evaluated using identical measures. At the end of the year, we saw a similar distribution of outcomes among teachers who had value-added data and those who did not. We also found positive correlations in teacher performance across all of the individual measures (Figure 6).

"ACE IS A CONTINUOUS IMPROVEMENT MODEL, and I don't see evidence of that in colleges of education. Their admissions screen is not rigorous and they are measured on the success of people moving through the gate. But they are not held accountable for how well people do on the other side of the gate. We need to scale the ACE approach."

—ASSISTANT DISTRICT SUPERINTENDENT

FINDINGS

FINDING 1

New teachers perform at different levels and improve at different rates.

Results from ACE confirm that even carefully selected first-year teachers perform at different levels and grow on different development trajectories. Recent research looking solely at teacher value-added scores came to similar conclusions.[13]

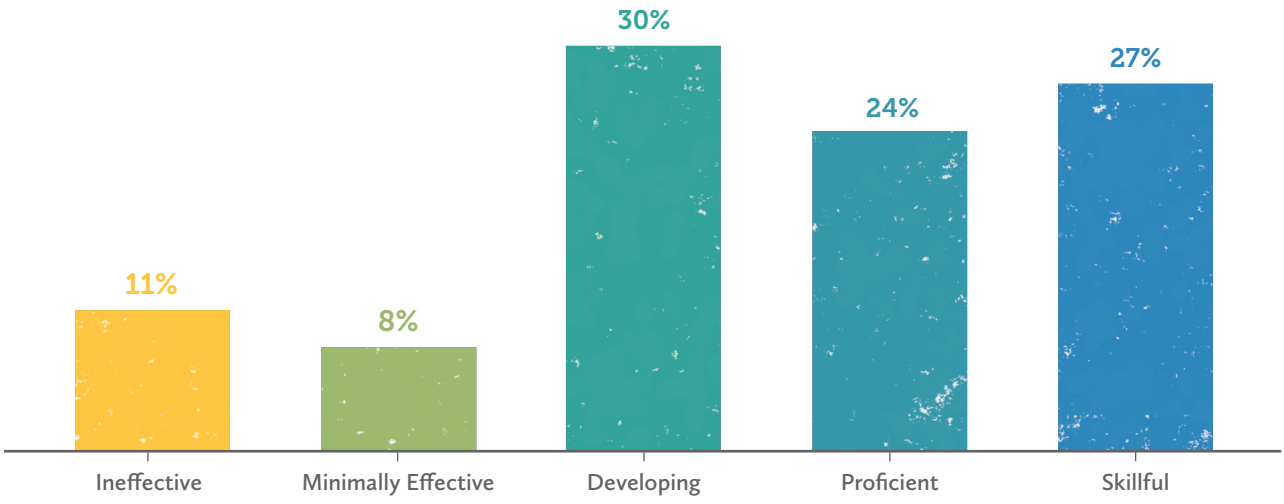
Our value-added results, where available, showed a wide range in first-year teacher performance. We translated teachers' value-added scores into a five-category performance scale ranging from "Ineffective," meaning they were performing much worse than the estimated performance of other first-year teachers in their district or state, to "Skillful," meaning they were outperforming estimates for other first-year teachers (Figure 3). About 19 percent of eligible ACE participants received ratings of "Ineffective" or "Minimally Effective" based on their value-added results, while 30 percent were rated as "Developing" and 51 percent received top ratings of "Proficient" or "Skillful" on our scale.

We found that patterns are much the same on classroom observations and student surveys. Some of our teachers started strong and developed quickly. Others struggled at the beginning of the year, but steadily mastered the critical skills they needed to be effective. Some barely improved at all—and a few teachers actually became less effective over the course of the school year.

Student survey data showed diverse outcomes, though teachers were more likely to receive positive ratings on our 5-category scale. About 56 percent of teachers received a "Developing" rating, while 37 percent were rated "Proficient." About 6 percent were rated "Ineffective" or "Minimally Effective," while just 1 percent was rated "Skillful."

Data from classroom observations showed that individual teachers grew at very different rates. Most teachers improved fairly quickly throughout the year, based on a series of three observations starting in January and concluding in the spring. On average, teachers gained about 0.20 points on a 5-point scale on each observation—a statistically significant increase.

FIGURE 3 VALUE-ADDED SCORE DISTRIBUTION, ACE TEACHERS

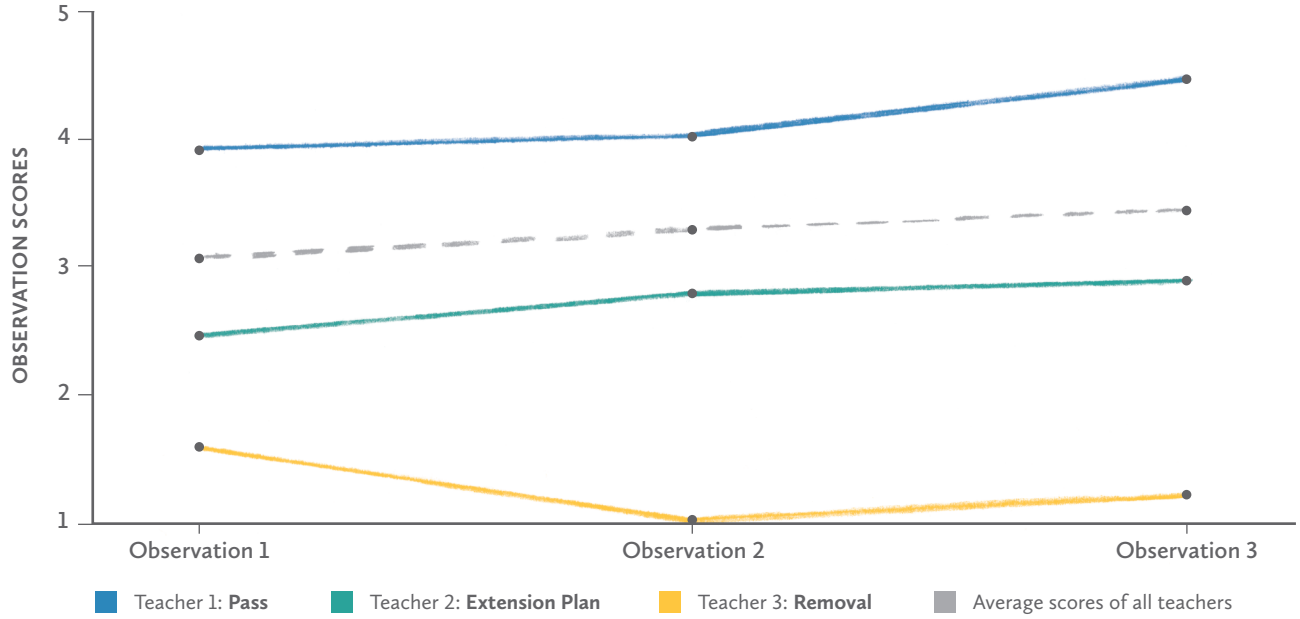


Value-added results showed a wide range in first-year performance.

Includes 142 teachers with value-added data, from Memphis, Nashville and Louisiana. SY2011–12. Source: TNTP.

[13] We found similar variation as in Atteberry, A., Loeb, S., & Wyckoff, J. (2013). Do first impressions matter? Improvement in early career teacher effectiveness (Working Paper 90). Washington, DC: CALDER. In that study, while many new teachers grew significantly in their early years, new teachers who began weakly often did not improve enough to recover from a poor start. Teachers in the bottom quintile in the first two years were still likely to be in the lowest quintile five years later.

FIGURE 4 AVERAGE VERSUS THREE INDIVIDUAL GROWTH TRAJECTORIES, ACE OBSERVATIONS



On average, teachers improve over the course of their first year—but not all teachers, and not all at the same rate.

Average observation scores for all ACE participants and scores from three individual ACE participants in San Antonio, Louisiana and Baltimore, SY 2011–12. Source: TNTP.

They did so by responding to detailed feedback and mastering new skills. For example, a teacher who struggled to ask her students follow-up questions during class discussions in January, after receiving her first observation report noting that issue, would work to master that skill by her next observation in March.

However, feedback did not always lead to improvement. Some teachers grew very slowly, if at all. About 12 percent began the year with very weak skills and improved, but not enough to earn observation scores higher than "Minimally Effective." Three out of four of those teachers failed to pass ACE.

In addition, among the 54 teachers denied certification at the end of the year, one in four actually showed a negative growth trajectory—and that was after getting off to an already weak start.

It is true that, on average, new teachers improve as they gain experience. But some do not, and some improve much faster than others. Experience does not guarantee a certain level of effectiveness, which makes measuring first-year teacher performance so critical.

FINDING 2

A teacher's initial performance predicts his or her future performance.

As we monitored our teachers throughout the year, we found that their initial performance was a reliable signal of their growth trajectory and overall outcomes.

Teachers who passed ACE started the program with significantly higher initial observation scores than those who were eventually extended or removed without certification—not only because they had higher scores at the start, but because they continued to outscore their peers in each successive round.[14] The passing teachers had an average score of 3.14 on the first observation, while both the extension and removal groups scored around 2.50 on average.

[14] Teachers who are in the top quintile on initial observation scores score, on average, about 0.6 points higher in Round 2, and about 0.5 points higher in Round 3. An analysis of variance showed these significant differences: teachers who passed scored 0.59 points higher than those who were extended and 0.64 points higher than those who were removed, $F(2, 999)=80.62$, $p<0.001$.

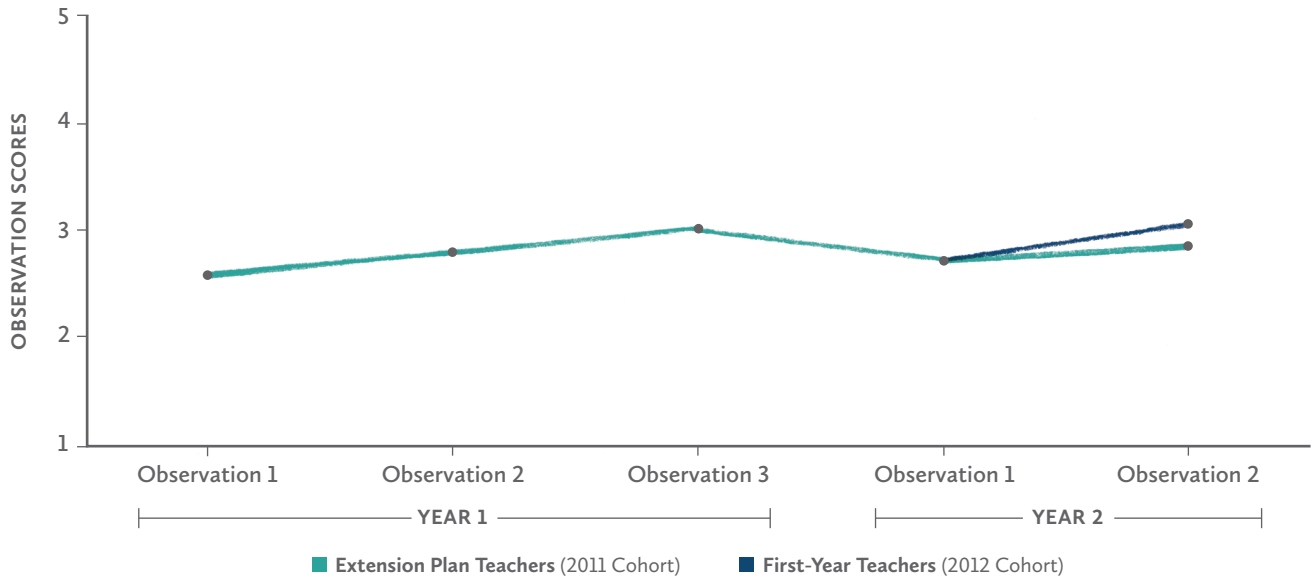
In other words, teachers who eventually passed ACE entered the observation phase of their year already scoring as "Developing" on average, and about 23 percent of those teachers scored as "Proficient" or "Skillful" on their very first observation. Meanwhile, teachers who did not pass ACE were considered "Minimally Effective," on average, during their initial observation.

Our extension plan program provided another opportunity to see how well ACE predicted future teacher performance. We granted 120 teachers who could not pass ACE outright but showed some potential the option to return for another year, and 88 teachers, or 73 percent, did. These teachers have so far continued to struggle in their second year; in fact, on average, their performance deteriorated. Extension plan teachers had a mean observation score of 2.81 at the end of their first year. As of January of their second year, based on multiple observations by observers who had not visited their classrooms the preceding year, the mean observation score for extension plan teachers had dropped to 2.72—not only lower than their earlier average, but also lower than the average score of 2.80 for current ACE teachers in their first year.[15] After more than a year in the classroom, not a single extension plan teacher earned an observation score in the "Proficient" or "Skillful" categories.

Most teachers on extension plans improved less in nearly two years than most first-year teachers had after just part of a single school year.

On average, extension plan teachers improved at about one-third the rate of the average ACE participant: by just 0.07 points between each observation round, compared to 0.20 points. In other words, when we compared observation score data for both groups after their second observation in January, most teachers on extension plans had improved less in nearly two years than most first-year teachers had after just part of a single school year.

FIGURE 5 OBSERVATION SCORES FOR EXTENSION PLAN TEACHERS, 2011–13



Many extension plan teachers did not improve significantly in their second year, earning lower average observation scores than first-year teachers.

Average scores from all ACE observations of extension plan teachers completed during SY2011–12, and scores from extension plan and first-year ACE teachers from September 2012–February 2013. Source: TNTP.

[15] Teachers on an extension plan received two observations, which were conducted between September and February using video. Videographers visited classrooms, taped lessons and submitted them for scoring. Two trained observers scored each video, and teachers received the average of the two observers' scores.

FINDING 3

Multiple measures tend to point to the same conclusion about a teacher's potential.

Results from classroom observations, student surveys, principal ratings and value-added data were positively correlated in ACE, just as they were in the recent Measures of Effective Teaching Project (MET) study.[16] A teacher who earned strong marks on one measure tended to earn strong marks on all.

For example, teachers whose students described purposeful, busy and demanding classrooms earned higher value-added scores and stronger observation scores. Those whose observations revealed poorly planned or insufficiently engaging lessons earned low marks in student surveys and had weak student achievement results.

At the same time, however, a teacher's performance on each measure was not uniform. Teachers earned slightly different scores on each measure, allowing us to get a nuanced picture of individual strengths, weaknesses and overall effectiveness. And, just as important, by having their performance evaluated along multiple measures throughout the year, including regular and rigorous classroom observations, they received feedback that could help them improve.

While all measures were positively correlated, the relationship between student survey results and value-added data was relatively weak. We believe this is because the number of teachers with both student survey results and value-added data was relatively small during our first year of ACE implementation: 43 teachers. However, this is a relationship that we plan to track closely in the coming years, since it is weaker than the relationships found in the MET study.

[16] Bill & Melinda Gates Foundation. (2010). *Learning about Teaching: Initial Findings from the Measures of Effective Teaching Project.*

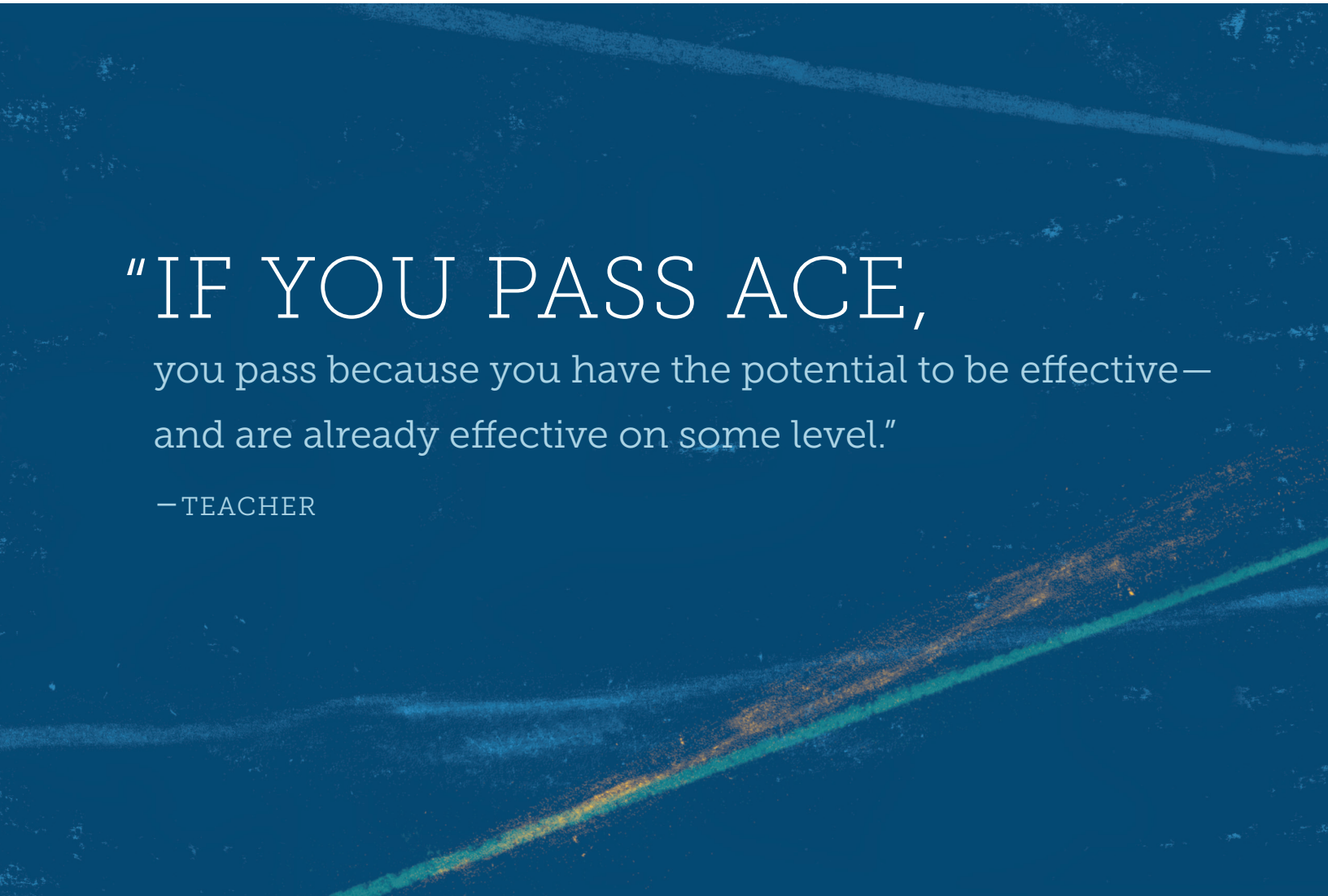


FIGURE 6 CORRELATION OF ACE MEASURES

	OBSERVATION SCORES	PRINCIPAL RATING	STUDENT SURVEY	VALUE-ADDED
Observation Scores[a]	1.00	0.32***	0.32***	0.22**
Principal Rating		1.00	0.30***	0.17*
Student Survey[b]			1.00	0.03
Value-added Score[c]				1.00

ACE components tell similar, but not identical, stories about teacher performance.

*p<0.05; **p<0.01; ***p<0.001. a: based on mean observation score; b: based on standardized student survey score from 483 teachers; c: based on the standardized value-added score for 142 teachers with value-added data, from Memphis, Nashville and Louisiana. SY2011–12. Source: TNTP.

FINDING 4

First-year teachers who are purposeful, responsive and focused on student understanding develop more quickly.

We looked for trends in classroom observation data to determine common qualities among teachers who advanced quickly.

One indicator on the ACE observation rubric was an especially powerful predictor of a teacher's growth. Teachers who earned scores of 4 or 5 on "Facilitates organized, student-centered, objective-driven lessons" improved more quickly than those who earned a 3. Those who earned especially low scores of 1 or 2 on this indicator had negative growth trajectories, losing 0.10 points per observation throughout the year.

We believe this measure was especially revealing because it encompasses a host of skills that undergird effective teaching: responsiveness, organization and content expertise. In our experience, teachers who can forecast and respond to student needs, present structured lessons and draw on a rich understanding of content to engage higher-order thinking skills are more likely to be successful.

In addition, we collected anecdotal evidence from our staff at various ACE sites to identify other factors that seemed to distinguish developing teachers from those who showed little growth. Teachers who could apply feedback from their observations tended to be more successful, while those who attributed weaknesses in their performance to school- or student-based factors tended to struggle longer. A teacher's willingness to take ownership of lackluster performance and respond with clear strategies to improve seems to signal future success in the classroom.

INSIGHTS INTO TEACHER DEVELOPMENT

Based on our first year of implementing ACE, we revised our approach to teacher preparation to fully integrate its evaluative components with the training and support we provide. Specifically, we used our data to assess which skills were most clearly associated with a strong foundation for growth among new teachers, and we revised our training based on that analysis.

Instead of trying to train our new teachers on everything, we now begin by focusing on four key "launch" skills that we believe are essential to a fast start in the classroom. These skills speak to teachers' ability to use time well, set clear expectations, implement routines and deliver instruction clearly. Under our new "Fast Start" model, we adapted our rubric to reflect their importance, and now spend much of our early training with new teachers practicing these fundamentals.

We organized these skills into four competencies:

- Delivers academic content clearly
- Maintains high academic expectations
- Maintains high behavioral expectations
- Maximizes instructional time

So far this year, we have found that teachers who master these foundational skills quickly move on to master all competencies evaluated through ACE. Meanwhile, teachers who struggle with these skills are much less likely to improve overall. As a result, we are thinking about how to tailor and segment our training to accelerate excellence early for those new teachers who start strong and make rapid progress.

At the same time, we are able to use our frequent observations as a targeted development opportunity. Each observation is an opportunity to clarify expectations and let teachers track their progress as they receive real-time feedback on their performance. Instructional coaches also pull teachers into small groups to practice teaching techniques that address their high-priority development needs. Through these processes, teachers are getting consistent feedback during their early months in the classroom, along with concrete strategies they can apply to improve specific skills.

In the future, we plan to make our professional development efforts even more strategic, by selecting a few critical priorities for each individual teacher based on his or her unique strengths. We are using the ACE observation data to better test and refine this more customized approach and track individual progress throughout the year. We also have enhanced the structured feedback we offer ACE participants following observations, to help them identify, understand and address gaps in their performance.

"THROUGH ACE OBSERVATIONS, I GOT ACTIONABLE FEEDBACK.

Now I'm in my second year and I don't have ACE, and I'm not being assessed as a teacher at all. I get observed twice a year by my principal... but it's really hard to get feedback."

—TEACHER

RECOMMENDATIONS

RECOMMENDATION 1

Connect teacher certification and on-the-job performance.

Every year, hundreds of thousands of new teachers earn their state license or are granted tenure with little regard to their ability to teach effectively in real classrooms. Some of them are not effective and never become effective. Course work and seat time remain the primary measures of adequate teacher preparation, and teachers' first-year performance is not given the care and scrutiny it deserves. Even when compelling evidence of weak performance is available, education leaders have hesitated to deny certification or remove new teachers from the classroom—though failing to do so puts students at risk and teachers on a path to an unsuccessful career.

Our experience with ACE suggests a better way. Policymakers should refocus certification on a teacher's actual performance in the classroom, based on a high standard for first-year excellence. Teacher preparation programs should be responsible for their teachers reaching an acceptable level of effectiveness. Our findings show that it is possible to meaningfully assess and differentiate first-year teacher performance and make sound certification and retention decisions based on the evidence.

Early-career teaching should be considered a learning period, but one with high standards based on rapid growth. Teachers should not be awarded licensure until they have spent sufficient time in the classroom to demonstrate their ability to become effective.

By the end of their first year, teachers should be able to create a positive classroom culture, manage student behavior and lead lessons in which the learning objective is clear. In addition, they should be responsive to feedback, and able to show that they are mastering and building on the crucial skills that set them up for long-term success: clearly delivering content, maintaining high expectations for students and maximizing instructional time.

First-year teachers who are unable to master these basic skills and show little or no growth should be denied certification; our evidence suggests that such teachers are almost certain to continue to struggle. Among teachers on an extension plan, 84 percent scored lower, on average, in their second year than the average first-year teacher.

RECOMMENDATION 2

Use classroom observations and student surveys when value-added data are unavailable.

Value-added scores are a powerful predictor of a teacher's future performance. We believe they should be at the center of evaluations whenever possible. But in their present form, they do not apply to the majority of teachers.

Other student-focused measures like frequent, high-quality observations from multiple external observers and student surveys are correlated with student achievement and can be collected for the vast majority of teachers. These measures can reliably capture real differences in a teacher's performance and professional growth.

Schools and preparation programs should base certification decisions on performance, even if value-added data are available for some, though not all, teachers. We need better information for more teachers, both from value-added data but also from the measures like student surveys and observations that are readily available today. Our experience shows that it is possible to move forward with rigorous evaluations even when student achievement data are not immediately available.

RECOMMENDATION 3

Focus new teachers on core skills, and look for rapid growth.

New teachers should concentrate on the skills that matter most for their future success. They need support around the basics: establishing a positive classroom culture and creating a sustainable classroom management style. Those who master these competencies quickly are soon able to develop more advanced skills, such as facilitating engaging lessons and calling on students' higher-order thinking skills.

Narrowing the focus to no more than 10 essential competencies allows first-year teachers to practice and improve quickly in those areas. Providing frequent, specific feedback on these targeted skills is essential, not only to help first-year teachers improve, but also to determine whether they can continually learn and grow—a hallmark of effective teachers. First-year teachers should show evidence that they are hearing and responding to feedback throughout the year.

Such focus can also make for stronger professional development opportunities at the outset of a teacher's career. We need to move past our instinct to sprinkle new teachers with wisdom and instead focus on the concrete skills they need be effective. To do this, we need to be disciplined about the feedback that we give to new teachers and focus on targeted, specific and immediate interventions. We must link evaluation to development, so that teachers no longer receive a one-size-fits-all development program but instead drive their own development, using information from observations and other sources of data to target specific practices that would lead to greater student learning.

"I think being evaluated was helpful. I know it made the whole year more stressful than a typical first year of teaching—which is already stressful enough—but I think it made me a better teacher in the long run. It kept me on my toes and made me really throw myself into it. I was trying hard for my school and for my kids, because I had to pass, I had to get better, I had to get a license."

—TEACHER

APPENDIX

ACE IN ITS SECOND YEAR

We adjusted ACE in several ways in 2012–13, the program's second year.

Scoring

The ACE score is now calculated using a simple weighted formula. If any of the measures are not available for an individual teacher, it is not included in the formula, with weights adjusted accordingly.

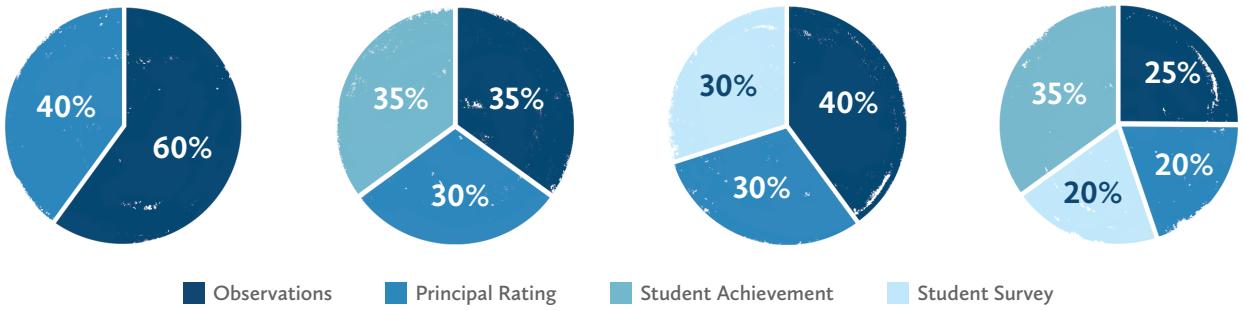
Total ACE 2013 score =
(VAM weighting × VAM points) +
(observation weighting × observation points) +
(principal weighting × principal points) +
(student survey weighting × student survey points)

Teachers no longer earn points for completing program requirements; instead, they are a stand-alone requirement, and teachers who fail to satisfy them do not pass ACE. In addition, principal ratings are no longer privileged in the new model. Under the old model, it was virtually impossible to deny a teacher certification unless her principal indicated she was much worse than other new teachers. Given principals' generally inflated responses, we felt that this allowed some low-performing teachers to continue teaching.

We also have moved total ACE scores from a 10-point scale to a simpler, 5-point scale. Teachers scoring 2.75 points or higher will pass ACE outright; those scoring between 2.50 and 2.74 points will be placed on an extension plan; and those scoring fewer than 2.50 points will be removed from the program without earning certification.

There are four possible scoring scenarios, depending on the measures available:

FIGURE 7 FOUR ACE SCORING SCENARIOS, 2012–13



ACE scores are calculated under a simple weighted formula, including the available information for each participating teacher.

Observations

Teachers now receive more frequent observations and more frequent feedback, starting earlier in the year. We also narrowed our rubric from nine to seven key skills, which are tightly aligned to our pre-service training curriculum and reinforced by a variety of program activities during the school year.

Following each observation, teachers receive supporting evidence showing how observers arrived at the score for each of these competencies, so they can identify things they have done well, while also gaining specific feedback on ways to improve.

Instructional competencies

- Delivers lessons
- Checks for student understanding of content
- Responds to student learning needs
- Builds higher-order thinking skills

Classroom culture competencies

- Maintains high academic expectations
- Maintains high behavioral expectations
- Maximizes instructional time

Surveys

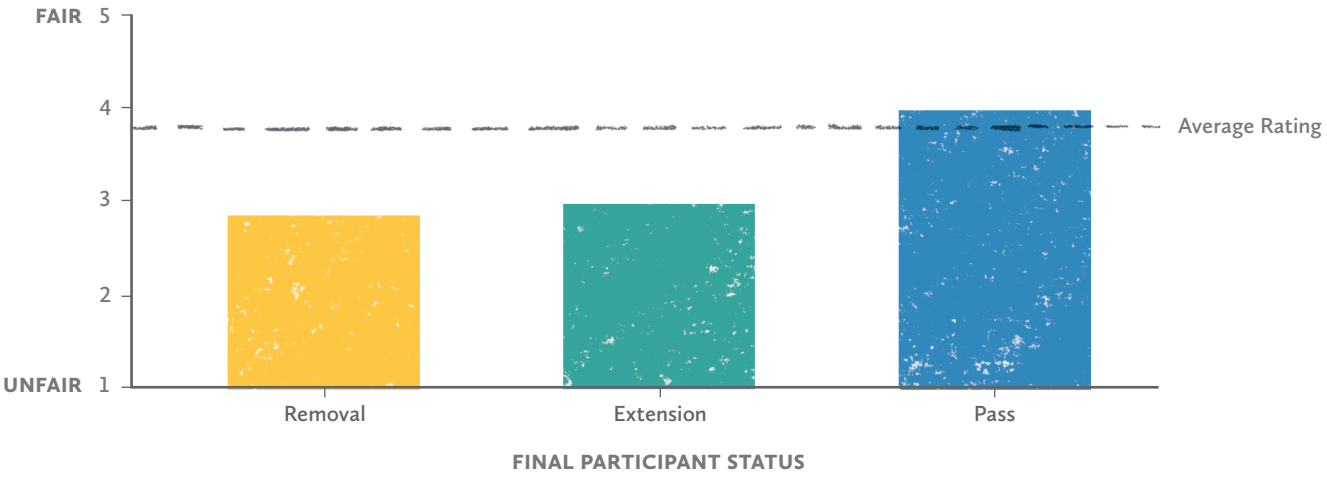
We have adjusted the timing of student surveys: they are administered in February and March, and teachers receive the results and a formative report in April. This allows teachers to hear and respond to feedback from their students before the close of the school year and follows an administration timeline that does not conflict with many testing schedules. Student feedback is also a powerful development tool for coaches working to develop skills with new teachers.

TEACHERS' PERCEPTIONS OF ACE

At the end of the first year of ACE, we studied our participants' perceptions to gauge whether the evaluation process was seen as fair. Through a series of survey questions covering the fairness of both the evaluation procedures and evaluation outcomes, we were able to create an overall "fairness perception" score between 1 (unfair) and 5 (fair).

Overall, teachers perceived ACE as somewhat fair, with a mean score of 3.80. While teachers who did not pass ACE did not rate the program as fair, their mean rating fell right at the midpoint of the fairness scale, suggesting a fairly neutral view on this issue. These results are consistent with a broad body of research, across multiple professions, indicating that people who ultimately do not meet the bar for performance are most likely to believe both the process and outcomes were unfair.[18]

FIGURE 8 AVERAGE FAIRNESS PERCEPTION RATING OF ACE TEACHERS, BY OUTCOME



Overall, teachers considered ACE to be somewhat fair. Teachers who passed ACE were more likely to consider the program fair.

Total teachers surveyed: 1,003. Total survey respondents: Removal 24; Extension 90; Pass 554. SY2011–12. Source: TNTP.

[18] For example, Colquitt, J., Conlon, D., Wesson, M., Porter, C. & Ng, K. (2001). Justice at the Millennium: A meta-analytic review of 25 years of organizational justice research. *Journal of Applied Psychology*, 86(3), 425–445.

About TNTP

TNTP is a national non-profit organization working to end educational inequality by ensuring that all students get excellent teachers. Founded by teachers and inspired by the power of great teaching to change lives, we help schools, districts and states grow great teachers, manage their teaching talent strategically, and build systems that prioritize effective teaching in every classroom. Since 1997, we have recruited or trained nearly 50,000 teachers for high-need schools, catalyzed large-scale reform through acclaimed studies such as *The Widget Effect* (2009) and *The Irreplaceables* (2012), pioneered next-generation teacher evaluation and development systems, and launched one of the nation's premiere awards for excellent teaching, the Fishman Prize for Superlative Classroom Practice. Today TNTP is active in more than 25 cities. For more information, visit **www.tntp.org**.