Why the Stanford Teacher Performance Assessment (TPA) Is a Step in the Right Direction

Marcy Singer-Gabella

Jennifer Wallace

What (and why) is the TPA?

Marcy Singer-Gabella

We who have a stake in teacher preparation need assessments that not only can help us predict teacher candidates’ future teaching effectiveness, but equally importantly, improve it, through the improvement of teacher preparation itself. The current political climate makes the need for credible and valid assessments particularly clear.

The Teacher Performance Assessment (TPA) is intended to address precisely this need. Designed to capture evidence of prospective teachers’ abilities to promote student learning of important subject matter, the TPA has demonstrated the potential to be both predictive and formative.

The TPA architecture has its roots in the National Board for Professional Teaching Standards portfolio assessment, which in turn informed the Connecticut BEST assessment system for teachers at the end of the induction phase, as well as the Performance Assessment for California Teachers (PACT).¹ Like its predecessors, the TPA is

- discipline specific, or focused on the teaching of particular content;
- student centered, examining teaching practice in relationship to student learning; and

¹ The PACT is one of three options for performance-based licensure assessment in California.
• analytic, targeting evaluation and feedback along multiple dimensions
• integrative, representing the complexity of teaching by requiring candidates to demonstrate skills and knowledge that typically are assessed in isolation.

The TPA is not intended to be the assessment, but rather part of a system of measures used by programs and licensing agencies. Other measures may be used by programs to track candidates’ progress in coursework and fieldwork. Typically these program assessments target smaller chunks of teaching practice, for example in the form of child case studies, analyses of student learning or curriculum, or action research projects. The TPA serves as a capstone assessment that pulls together various facets of teaching practice, and also allows for programs to examine their candidates’ progress against an externally recognized benchmark.

The anchoring task of the TPA is referred to as the “teaching event,” and occurs during student teaching, residency, or internship. The teaching event engages candidates in deep documentation of a meaningful chunk of instruction designed around a big idea or essential question. Candidates plan a sequence of three to five lessons that build toward a central, cohesive purpose. These lessons should scaffold learners’ understanding of important content and skills while ensuring learners’ “academic language” development, i.e., the understanding of vocabulary, symbolic representations, and language structures needed to make sense of and apply subject matter. There is a premium placed on candidates’ abilities to tailor strategies, materials, and lesson plans to the specific needs of the learners in the class. Candidates should thus demonstrate understanding of who the learners are, what assets they bring, and how to continually adapt instruction to fit their needs.

Candidates’ documentation takes the form of artifacts and commentaries organized into four “buckets” or tasks: planning, instruction, assessment, and analysis of teaching. Artifacts provide a kind of independent evidence of what transpired in the classroom. Commentaries provide windows into a teacher’s thinking: candidates are prompted to explain what they intended to happen and why, what actually happened and why, and to identify and justify next steps for learning and instruction.

For the planning task, candidates provide information on the instructional and social context of the classroom; gather artifacts, including lesson plans, handouts, overheads, and student work; and write a lesson commentary that
explains the plans and the thinking behind them. The instruction task requires candidates to select video excerpts of their work with students during the lesson sequence, and to write a commentary explaining what happened in class, and why they made the decisions they did (e.g., why they responded to students in particular ways, adjusted their plans, etc.). The assessment task challenges candidates to analyze the evidence they have gathered of all students’ learning, and to represent that analysis on a chart or other display. Candidates also focus in on the learning of a subset of students, showing how the students’ learning is reflected in their work samples, documenting feedback given to those students, and indicating how that feedback would advance the students’ learning. Candidates also discuss implications of their analyses for instructional next steps. The analysis of teaching task asks candidates to assess the overall impact of their instructional decisions on student learning, and to describe what they would do differently given the opportunity. In each task, candidates are asked to track their work in preparing students for the language demands of the classroom.

Candidates’ documentation and analyses are assessed along a set of twelve rubrics targeting instructional competencies associated with effective teaching. The five-point rubrics are additive and analytic: lower levels describe a more limited repertoire of skills; the upper levels describe more extensive and strategic knowledge and skills. There is a movement from fairly thin rationale and reflection to much deeper rationale and reflection, and from a focus on what the teacher is doing to a focus on what the students are doing. Finally, the rubrics capture a shift in candidates’ abilities to differentiate among learning needs—from attention to students as a whole, to consideration of learners as individuals.

There are many players involved in the development and implementation of the TPA. Stanford University has led, and continues to lead, the instrument design, development, and research. The American Association of Colleges for Teacher Education (AACTE) is providing national leadership to the TPA Consortium (TPAC) through overall project management and support for communication among all stakeholders. The Council of Chief State School Officers (CCSSO) served as an official partner early on in the area of state education agency policy development and support; Peter McWalters (former consultant to CCSSO) and Deanna Hill of West Wind Education Policy have taken up this work, focusing attention on strategy and logistics around state adoption. And finally, Pearson has signed on as an operational
partner for implementation, scorer recruitment and training, and the qualification of scores.

The Consortium now includes twenty-six partnering states, many of which are engaged in piloting this year. Some states are taking up the TPA in response to policy mandate; higher education institutions in these states are moving more quickly toward full implementation with teacher candidates. Each state team includes representatives from not only the institutions of higher education, but also the central policy-making organizations in those states, such as departments of education and licensing boards. The TPAC leadership was prescient in requiring the involvement of key representatives from these stakeholder groups in order to build strong understanding and broad ownership. No doubt, such engagement will support the TPAC goal of establishing a scalable model.

Jennifer Wallace

Of the twenty-six states now in the TPA Consortium, there are six that are considered accelerated states because they are on a shorter timeline to full implementation. I believe it is still true that Washington State is on the shortest of those timelines, which is why we are often asked to share our story. At the same time, we’re grateful for, and clear on the benefits of, being part of a larger multi-state consortium. The real strength of the consortium approach is the feedback and involvement of all its members, and we encourage states to be involved in that dialogue early on, even if they are just considering whether the TPA is right for them. We’ve realized that with our different assumptions about its use in terms of policy and how it is going to be situated within our approach to continuous improvement of programs or our program accountability systems, it’s not going to realize its potential in terms of a consortium unless we are having a much more open dialogue across states about its use. So I’m going to share Washington’s story of how it’s going thus far, what we have encountered, and where we anticipate we will be going as we implement this.

In Washington State, we’ve had a performance-based pedagogy assessment in place since 2003. It was a home-grown instrument, and it became clear to us over time that it was lacking sufficient validity and reliability to be used either for consequential decisions about candidates or as part of our accountability system. As a result, the Washington Association of Colleges
for Teacher Education (WACTE) took the lead in exploring the options out there that could yield the validity and reliability both our institutions and the Standards Board were seeking. As we started shopping for instruments that could serve our purposes, it was very important to Washington to be able to do so in the context of our own state standards and the way that they have evolved over time. Our standards began as a list of topics that simply implied rather than described desired behaviors or outcomes: inquiry and research, classroom management, professional ethics. In 2000, the standards evolved to include those that were more performance based, with actual descriptions of desired teacher behaviors that we wanted to see prior to licensure: aligns goals, instruction, and outcomes; manages the classroom effectively; uses multiple instructional strategies. And since 2008, we have had in place evidence-based standards that describe student learning behaviors that we expect to see as a result of good teaching: students are cognitively engaged, students can articulate their understanding, students monitor their own progress. We wanted to make sure that both of those components were very much a part of the assessment we chose, and we were very happy to see that they were part of the TPA. One thing that was most encouraging to us in considering the TPA is that it provides the ability to modify the assessment in a way that can align with particular qualities of a state standard while still providing similarities in the instrument to allow for comparability across states.

At about the time we were leaning toward the advantages of the TPA, the TPA Consortium was forming so the WACTE brought its recommendation to the Professional Educator Standards Board to join the Consortium. We then brought this recommendation to the legislature, asking for its support in putting the TPA in place as a reliable statewide classroom-based means of evaluating the effectiveness of teacher candidates at the pre-service level, and also as part of our preparation program accountability system. While we received support from the legislature in terms of officially adopting this instrument, they provided no funding support. Only through the funds within our programs and our own scraping together of money have we been able to continue on this road.

So it was in May 2009 that the Standards Board voted to be part of the TPA Consortium, with the rationale being not only the quality of the instrument but also the potential research agenda that accompanies this. We would very much like to demonstrate that candidates who perform well on the TPA do in fact go on to be effective in their classrooms as measured by student
achievement gains. I know that a lot of states are under pressure for those kinds of measures, and we are looking forward to that kind of validity. And then there is the idea of reciprocity. Washington is a big importer. About one-third of our teachers come from other states, so we are very much looking forward to those states participating in the TPA and their candidates completing it prior to seeking certification in Washington. In terms of our successes with TPA—which have kept us on track in Washington and which are critical for any state going down this road—one critical component has been the high level of program participation and buy-in across the state. This has been pretty amazing, particularly in light of the fact that they have done so all on their own resources. The other thing that is becoming clear is that it’s prompting changes in our program design in terms of field experience far more than we ever anticipated. This is a real game changer, not only for our programs but also in terms of the relationship between programs and school districts. We regulate preparation programs, but not school districts. Field placement is more difficult than ever, as is getting districts to understand that the presence of pre-service interns can and should be an integral part of school improvement as well as a smart workforce development strategy.

In Washington, we are going to use the TPA as a high-stakes measure of accountability, both for candidates and for programs themselves, and it will go fully consequential in the 2012–13 school year, required for all completers in Washington programs. In addition, we have added a component that includes the measure of student voice. We are structuring a new framework of evidence that we will require for program review and approval, and this will be a key part of it.

There are many challenges and policy questions that we are facing, which every state will have to face, such as the type of remediation that this is going to require, how we can work with programs around that, and how we should deal with retakes in light of the limited time that candidates are out in schools. What about appeals? And will we use this for add-on endorsements, since this is a content-specific assessment, or just for certificates? And we’ve pretty much ruled this out for Washington, but there is the question of whether this would be a requirement for certified teachers coming from other states. A big question around accountability is how to weight different measures for review and approval. We must look at the surrounding data and whom we are weeding out prior to student teaching. One of the critiques about the TPA involves the fact that we actually anticipate pretty high pass rates, which is most likely going to earn some
kind of criticism about its being meaningless if a large number of people are passing it. We are already beginning to see programs think much more about recruitment and selection—who is going to be entering student teaching and what decisions need to be made in terms of individuals who are actually going to go on to assume this assessment and enter teaching. So though we actually hope to be seeing high pass rates, we will need to look at the surrounding data in order to make the results and the TPA meaningful beyond just the score. There are issues around balancing faculty scoring—which is valuable in terms of program improvement—with external scoring—which for our legislature is extremely important. There are operational issues, such as how we are communicating with the districts, as this will place greater demands on them in terms of their supervision and work with candidates. And then there are the nuts-and-bolts issues around such things as video permissions. These are just some of the issues that we are trying to resolve as we implement this plan in Washington.

Marcy Singer-Gabella

Here is the on-the-ground perspective. At Vanderbilt, our history with the assessment model goes back to the 2007–08 academic year. At that time, we were looking to establish a culminating assessment for our masters’ students that would be grounded in real teaching practice, would provide comparable data about our candidates across program areas, and would be sensitive to important subject-matter differences. In addition, it was vitally important to us that the assessment tool provide usable data for both our programs and our candidates.

During our 2008–09 pilot, we found the assessment to be challenging not only for candidates but also for faculty—but the challenge seemed to be the right kind of challenge and one worth repeating. With the encouragement of the AACTE and Stanford, we invited faculty from the University of Tennessee–Knoxville and the University of Memphis to join us for the following year, to see how this type of assessment would work in much larger, public institutions. In spring 2010, following more debriefing at Vanderbilt and with our new partners, site teams at each campus agreed that the TPA had the potential to provide the kind of data on program and candidate performance that would hold up, therefore serving our need for public accountability. So we “reupped” for the official spring 2011 TPAC pilot and were joined by five more Tennessee institutions; in total,
approximately four hundred candidates completed the assessment. This coming year, there will be twelve institutions participating at some level, for a total of somewhere between eight hundred and one thousand candidates. At Vanderbilt, the TPA is now a requirement for all candidates seeking Early Childhood, Elementary, and Secondary Education licensure.

Since we began working with the TPA at Vanderbilt, faculty conversations about our aims for candidates and how to achieve them have become more focused, coherent, and fruitful. The TPA spotlights knowledge and skills required to support the learning of all students. These include careful planning and design of learning experiences that take into account learner strengths, needs, and differences; systematic assessment and adjustment of instruction in response to students’ learning; and explicit attention to the language demands of learning new content. Although our faculty would have agreed on the importance of these elements of sound practice, the TPA has given us a shared language as well as images of what these elements look like and what it looks like to learn them. This has allowed us to more effectively communicate expectations to our candidates and to our partners in schools. For example, we have become more explicit about what kinds of classroom experiences candidates must have. If we want our graduates to be able to foster, assess, and respond to learners’ understandings, then candidates need to have opportunities to figure out and test what questions to ask, what students’ answers suggest about what they know, and what should be the right next step to push student learning to the next level.

The TPA has also given us a clearer view of candidates’ abilities to put their knowledge, skills, and understandings to work. Our first two years with the TPA showed us that we were not doing the job we needed to do in helping candidates systematically assess and interpret evidence of all students’ learning. Nor were we helping them get better at recognizing the language demands of learning math or science or social studies, and creating supports for students (beyond vocabulary lists). Thus, we went back and revised courses and field experiences, from foundations to student teaching, to provide explicit attention and practice in these areas. Students’ spring submissions suggest that we have gotten better at this in several programs, and we are planning further revisions for next year.

This is not a simple model—in fact, some parts feel pretty grueling. What allows us to make it work are the strong technical support, collaboration, and guidance from those at Stanford and increasingly from our peers at
other universities in Tennessee and across the country. What makes this model challenging is what makes it potent: it examines candidates’ abilities to do the central tasks of teaching. Because it gets at the heart of teaching practice (not just knowledge about practice), it is a credible and informative tool for continuous learning about our candidates, our programs, and how to teach.

At Vanderbilt, we’ve been lucky to have candidates who arrive pretty accomplished and usually succeed in teaching. But according to feedback from the field and from our supervisors’ observations, over the past three years they have gotten even better. We credit changes spurred by the TPA to a large degree. We think that the TPA model has a positive impact on candidates’ preparedness because of its emphasis from start to finish on P–12 learners.

First, the TPA’s structure and specific prompts require candidates to align standards, activities, and assessments in relation to diverse learners’ needs and development—and maintain this alignment. Second, the TPA requires candidates to continuously pay attention to students’ developing understandings as reflected in what students say and do—not just the class as a whole but students individually. Thus, candidates need to make sure they glean information about each student in the course of each lesson. And third, the TPA requires candidates to use the insight they are getting about students to determine the next move. And again, they have to think about that next move in terms of individuals and small groups, not just the group as a whole. Learning to keep all of these balls in the air is difficult and takes practice. Because we are now getting data in relation to these very specific expectations, we can give candidates not just more but also better opportunities to practice, as well as better coaching and feedback to inform their learning.

So what does it take for campuses to implement the TPA? There are a number of things to consider. One is figuring out how the TPA fits into the set of assessments already in place—and which of these it could replace. In Tennessee, this has looked different across campuses. At Vanderbilt, there was a similar bundle of planning and assessment tasks, but without video or attention to academic language. At UT Knoxville, where candidates take part in a yearlong internship and are given credit for one year of teaching when they graduate, candidates have historically been required to undergo
the state teacher evaluation process. The TPA has substituted for about one-third of that process.

Campuses also must determine who supports candidates, how to begin implementation and how many programs to involve in the piloting, what the consequences of failing are, what opportunities will be available for remediation, and how costs will be managed. The added expenses to campuses reside not just in the cost of scoring the assessments but in the amount of support required—particularly at start-up, when candidates have not seen the TPA as an integral part of their program. At Vanderbilt, we reduced the ratio of supervisors to student teachers and paid part-time people a bit extra to provide that extra support. Recently we have scaled back that support somewhat. In addition to time, personnel, and assessment fees, there are the costs of purchasing and maintaining equipment. As we move to a Pearson platform with an external fee, we are continuing to worry about cost. Candidates do not generally have money to spare, and there are other content assessments required by the State. Currently we are discussing the possibility of folding the TPA scoring cost into the university tuition structure, so that it is covered (in part or in whole) through financial aid.

One final but critical consideration is how programs will review and learn from candidates’ work. For campuses, the real power of the TPA lies in the fact that it makes candidates’ performance visible, offering real data about how well we are preparing candidates to support student learning and achievement. We need to make sure that we keep taking advantage of this, by sitting together and thinking carefully about what the TPA data tell us, and the implications for our next steps.