

## Taking Stock of Performance Assessments in Teaching

**Jon Snyder**  
*Bank Street College*

For the past twenty years, approximately a human generation, educators, policy makers, and psychometricians have been developing performance assessments of teaching. From rudimentary portfolios to legally defensible assessment systems, the field has traversed a fair distance for their efforts. This generational moment provides an opportunity to take stock of what may still stretch ahead.

Certainly critics and critiques remain. Some argue that the “it” of teaching cannot be defined with sufficient clarity to be assessed. As in golf, the critique goes, some highly successful golfers use a swing that no one would ever teach. If a golf swing, so much less complicated and involving only a single golfer (a golf score does not depend upon how the gallery responds to the swing) cannot be thus defined, how can teaching? The pressures to improve one’s golf score, however, do not compare to the pressures to increase the life options of children in the care of teachers, so the demands for an assessment are not likely to abate.

Others argue that the reliability of performance assessments cannot meet measurement standards. This is a bit like arguing that we should give multiple-choice tests of basic arithmetic rather than driving tests to grant a driver’s license because the arithmetic tests have higher reliability numbers than do the assessors of driving tests. Still others, often from within the field of teacher education, argue that, nearly by

---

*Jon Snyder is the dean of the Graduate School of Education at Bank Street College in New York City.*

definition, assessments limit the richness of their programs, require resources that could be more usefully spent in other ways, and/or harm the nature of the relationships essential for learning.

Looking ahead, however, it is unlikely that demands for assessments of teaching are going to disappear. It is also highly unlikely that other approaches to assessing teaching will come anywhere near measuring what matters about teaching. Looking back, it is also clear from the articles in this journal as well as in an array of professional literature, that performance assessments have demonstrated an increasingly enriched potential for serving multiple important functions in the professional preparation of teachers and thus the opportunity for children to leave their K-12 schooling with the options to pursue a future of their own choosing.

If performance assessments are unlikely to go away, or if they did go away, would likely usher in much less beneficial alternatives, what issues and challenges should be focused upon into the near future of the art, craft, and science of performance assessments of teaching?

### Multiple Functions

One challenge that brings with it a host of issues is the multiple functions performance assessments are being asked to serve. They are supposed to meet:

- The professional responsibility of the field to assure the public that teachers are “safe to practice” with children (i.e., individual evaluation);
- The institutional accountability of preparation programs (i.e., continual program improvement); and
- The strengths, interests, and needs of teacher candidates (i.e., support learning from teaching).

The core issue when attempting to meet multiple functions with a single approach is to find a way to balance the (sometimes) competing demands so that no single function dominates to the detriment of the other functions. For instance, the history of high stakes assessment clearly demonstrates that it is distinctly possible that better meeting the demand for a psychometrically sound and legally defensible evaluation of individual teacher candidates would lessen the value of performance assessments to support program improvement or candidate learning. Similarly, this outcome is also distinctly possible should either of the other two functions assume dominance. At the same time, however,

developing different assessments for all three functions is quite likely to sink the entire professional education boat in a sea of assessments. One loses weight through diet and exercise, not by weighing oneself.

Within each of the three functions, there are core issues to address while attempting to maintain the necessary balance. For the purpose of individual evaluation, there is the need to build a better mousetrap. If the assessments neither distinguish between candidates nor assure the public that a teacher is ready to begin teaching, then both the profession, and more importantly the children, suffer. One challenge in this regard is that context matters. It is fundamentally impossible to make a judgment about teaching without considering the multiple contexts within which it occurs. On the other hand, it is extraordinarily difficult to meld multiple subjective perspectives (make an evaluative judgment) that considering contexts requires. In addition, it will be necessary to become clear about comparison groups. Any single data point makes more sense when it can be compared to another data point. Information becomes “real” when it can be compared to something else. So the question is, make it real, compared to what?

To meet the demands for institutional accountability, a core challenge is how to move from data to action. There is a widespread notion that data magically change the world; that knowing that a program is not as good as one would like it to be in some area automatically improves the program in that area. Or that the shame of being towards the tail end in a horse race automatically means one will finish in the money the next race. Human endeavors do not work that way. The challenge is how to move from data (assuming the collection of good data), through interpretation, and into programmatic improvements. A second significant challenge for programs is understanding and using the tendency for field experience placements to influence scores on performance assessments. Teacher candidates clearly need to understand the realities of the schools, classrooms, and communities in which they will be working. If those realities, however, are not matched with the values and goals of the program or the “it” of teaching measured in the performance assessments, then both the candidate and the program are left between a rock and a hard place. The challenges of coordination, communication, and congruence among and between expectations from the field, the college, and the performance assessments will need to be addressed.

To use performance assessment data to support the learning of candidates, teacher educators require the kind of data, in a time frame, that is of use for that purpose. This is a particularly difficult balance to locate as the demands of institutional accountability for aggregated data, by definition, create different data needs and time frames than

either individual decision-making or individual support. Meeting the demands for program improvement and support of individual candidates also requires teacher educators to use distinctly different sets of analytic skills. All individual decisions and individual support are thinking with an “N of 1.” Programmatic decisions require thinking with aggregated data where the individual is purposefully hidden, at least temporarily, from view.

### Resources

Quality performance assessments require resources, especially when being asked to fulfill multiple functions. A quality performance assessment just for the function of individual assessment, such as those required for National Board for Professional Teaching Standards certification, requires several thousands of dollars per candidate for the scoring and processing of assessments. That figure does not include the time of the candidate or, if the candidates are supported, the time of the mentors who work with the candidates. If performance assessment results are also expected to support programmatic development, then the time and expertise required to compile, analyze, and display the aggregated data as well as to move from data to constructive action must also be taken into account. If the performance assessment process is expected to improve candidates’ development in the learning from teaching process, then the time, expense, and teacher educator expertise required cannot be ignored.

Historically all of these costs have been borne by the candidates—though when the assessments are embedded in public teacher education programs, the costs are subsidized by state funds. While it is possible for candidates to continue to bear some of these costs and also possible for teacher education programs to reallocate existing resources (time and expertise of teacher educators), quality cannot be sustained in the long run on the backs of prospective teachers and teacher educators. While not universally true, in this instance, children are most likely to receive what the public is willing to pay for. Institutions of higher education and districts, who share the responsibility for the education of teachers, will need to re-allocate resources, but additional funds will also need to be located if performance assessments are to continue towards meeting the demands being placed upon them.

### The Human Factor

Just as the purpose of teaching children is to intentionally shape and support their growth and development, so the purpose of teaching

teachers is to intentionally shape and support their professional growth and development. In either instance, therefore, it is essential that the strengths, interests, and needs of the learner be taken into account in the construction and use of assessments. Whether assessing teachers or children, the fundamental goal of the enterprise, the prime directive, must be remembered. Often as the field seeks to define the outcomes of an enterprise, assessments become more focused on the defined outcomes than on the nature of the learner and the processes of development. This inevitably limits the capacity of the assessments to support those very desired outcomes. Lucy Sprague Mitchell, the founder of what eventually became Bank Street College, often told the following relevant story when describing her initial work in the field of measuring the growth and development of children. Bank Street researchers were working with researchers with a more positivistic and mechanical approach to measurement than was Bank Street. They wrote that they were having trouble measuring the height and physical dimensions of children because the children would not stay still long enough to be measured. They kept wiggling. Ms. Mitchell responded that it was the wiggle that was of the most interest and value to understanding children and how to create environments that supported their development. The same holds true of teachers. As humans they do not hold still for long. They, in their own ways, squirm just as children do. The wiggles of children and their teachers may prove the most difficult part of performance to assess, but the wiggles remain the name of the game.