

Portfolio Assessment as a Tool for Teacher Evaluation: Outline and Key Points of Presentation to the Maryland Council for Educator Effectiveness (5/16/11)

Gail Lynn Goldberg, Ph.D.

Introduction:

Establish that a portfolio-based system is one plausible way to assess teacher performance through evidence of student growth, but that there are various considerations that must be addressed.

Background:

Portfolios are not new--began to appear in educational contexts in the 1980s as collections of work that showcased the student as a learner, and allowed for students and teachers alike to have a better sense of growth over time.

Various types of portfolios are customarily used for different purposes:

- **Presentation or “showcase” portfolio:** to catalogue best work/final accomplishments
- **Growth portfolio:** to show student progress over time; primary purpose is *usually* instructional (to highlight strengths and weaknesses to drive instructional decisions)
- **Evaluation portfolio:** to demonstrate how well a student/teacher has met specified goals and standards over a period of time

The portfolio that the committee envisions would be a **hybrid**—a portfolio that provides evidence of student growth *for the purposes of teacher evaluation*. This is certainly doable, bearing in mind that a successful portfolio assessment system:

- Is ongoing rather than representative of a single point-in-time
- Allows a window into process as well as products
- Provides opportunity to revisit and revise, guided by evaluative criteria
- Allows for diverse means of demonstrating competency (at least some choice)
- Serves as a demonstration of strengths rather than weaknesses/deficits
- Encourages reflection and decision-making regarding future

Entries/Portfolio Contents:

Will likely need to include evidence of both student learning and the teacher’s role in that learning. Other entry options include:

- evidence of more gradual evolution of piece of work over time and/or skills progression along with evidence of teacher feedback supporting that evolution/progression
- student process journal that captures student’s understanding and use of teacher feedback
- identification of goals and reflections on advancement towards those goals (by both student and teacher)
- evidence of student’s strengths and weaknesses at two or more points in time (with explanatory account by teacher)

Constraints for high-stakes use of a portfolio assessment system:

- To ensure against teachers “stacking the deck” or “cherry-picking” by selecting uncharacteristically weak work samples to illustrate baseline, the determination of entry(ies) should be made prior to review by the teacher.
- Any support (peer, teacher, tutor, parent, etc.) must be documented (e.g., with context map)
- Most critically, perhaps, a portfolio “menu” must be created as a blueprint for the contents of all of the “measure of growth” portfolios. Only by specifying the contents to a *considerable* degree can judgments about teacher effectiveness be made in a fair and valid way.

Possible approaches to gathering entries:

- Use the “vertical slice” approach to examining student work
- Collect several “paired” entries (same CA, same standard, same objective) for each of a set number of students
- Pay attention to the adequacy of the sample, the comparability of entries, the degree to which the entries are representative, and the “yardsticks” that are appropriate and sufficient to determine growth, such as:
 - becoming better at the same skill
 - demonstrating the same degree of competence but with more complex/demanding resources/stimuli
 - demonstrating less need for support/scaffolding and more independence with same skill

Scoring:

Given the intended high-stakes nature of this portfolio, a scoring system must be devised with care to ensure valid and reliable scores. Among the many challenges suggested by the past experience of large-scale portfolio systems):

- Scoring the portfolio as a whole, rather than assigning scores first to component entries, often results in weaker inter-rater reliability. Too many dimensions can also weaken reliability.
- The more “uniform” entries are (that is, the fewer standards, domains, genres, or other “categories” of entry are included), the higher is the agreement rate among those scoring the portfolio.
- The more unconstrained/varied the samples collected, the more thorough and specialized must be the background and expertise of raters.
- Typically, one or more independent readers reviews the portfolio contents and assigns scores according to one or more scoring tools (rubrics and/or checklists) for one or more dimensions
- The degree of “remove” varies (other teachers in the same school, same district, other district, etc.), but the closer to home (i.e., within LEA), the more opportunity for value-added in the form of professional development associated with scoring.
- A well-designed system of auditing scores (double-scored, scored by external “expert” rater) is adequate to ensure accuracy of scores within an acceptable degree of disagreement (adjacent rather than discrepant scores).
- You may also wish to consider a mechanism for moderation (adjustment of scores, based on trends in scoring that show up in inter-rater judgments--done in various international assessments of teacher competency)

- Regarding *E-Portfolios*: Facilitate “distributed scoring” (portfolios are randomly assigned for scoring to a trained reader who can access those portfolios electronically and record their scores; supports remote scoring, which in turn increases interest among qualified individuals who you might consider as raters, and all training, monitoring, recalibration, and adjudication can occur online.

Feasibility:

- To reduce “housekeeping,” teachers could be randomly assigned from among a given year’s roster a subset of students for whom portfolios would be maintained (with the option to maintain them for all students)
- The sampling methodology should be determined by TAC
- To accommodate a mobile student population, include some “overage”

Cost:

- Not inexpensive
- Some costs are “hidden” by the use of teachers’ and administrators’ time, or institutional sharing of costs to convene a trained panel of expert reviewers
- Any estimates must account for handling, training, scoring, monitoring, and auditing
- See CSE (Center for the Study of Evaluation) Report #441, titled “Alternative Assessment Programs: What Are the True Costs? An Analysis of the Total Costs of Assessment in Kentucky and Vermont” by Lawrence O. Picus and Alisha Tralli (CRESST, University of Southern California) 1998
- Some guidelines
 - BCR—1 minute
 - Writing sample 2-5 minutes
 - Complex performance task (50 minutes-1 hr)
- Professional raters (contractor’s staff) typically earn \$12-\$15/hr; teachers paid \$18 or \$19 by ETS

Other thoughts:

Although it is my understanding that you are considering a portfolio as the mechanism for ascertaining student growth, that platform could allow the local option of including for consideration other entries/artifacts that serve to illustrate teacher competence:

- Products of the classroom (include any earlier versions to show genesis of this artifact)
- Lesson plans (*include any earlier versions to show genesis of this artifact)
- Student work with teacher comments
- Interpretive case studies
- Videotape of classroom interaction
- Videotaped interview
- A portfolio probe (question designed to project a work product into another setting or application)