Assessing and Evaluating Teacher Education Programs

Frank C. Worrell
School Psychology Program
Graduate School of Education
University of California, Berkeley
Overview

• Broader Educational Context

• Political Context

• Task Force Report on Evaluating Teacher Education Programs
Educational Context I: Achievement Gap
Nat. Assess of Educational Progress (NAEP) Reading Scores - 2009

4th Grade

African Am | Amer Indian | Asian Am | Hispanic | White

Scores:
- African Am: 205
- Amer Indian: 200
- Asian Am: 235
- Hispanic: 205
- White: 230
NAEP Mathematics Scores

4th Grade

8th Grade

African Am
Amer Indian
Asian Am
Hispanic
White
NAEP Civics Scores

[Bar chart showing NAEP Civics Scores for African American, Hispanic American, Native American, Asian American, and European American students across 4th, 8th, and 12th grades.]
Performance Levels
National Assessment of Educational Progress (NAEP)

- **Advanced:** Superior performance
- **Proficient:** Solid academic performance
- **Basic:** Partial mastery of skills
- **Below Basic:** Less than partial mastery
NAEP Reading Performance

Bar chart showing reading performance by different racial groups:
- Below Basic
- At Basic
- Above Basic

Groups:
- African American
- Hispanic American
- Native American
- Asian American
- European American
NAEP Mathematics Levels

2009 Mathematics 4th Grade

- Below Basic
- At Basic
- Above Basic

- African American
- Hispanic American
- Native American
- Asian American
- European American
NAEP Reading Scores 2009

Eligible Free/Reduced Lunch

2005: 205
2007: 205
2009: 205

Not Eligible

2005: 235
2007: 235
2009: 235
Nat. Assess of Educational Progress (NAEP) Reading Scores - 2009

4th Grade

<table>
<thead>
<tr>
<th>Group</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Af Am</td>
<td>200</td>
</tr>
<tr>
<td>Am Ind</td>
<td>195</td>
</tr>
<tr>
<td>As Am</td>
<td>240</td>
</tr>
<tr>
<td>Hispanic</td>
<td>205</td>
</tr>
<tr>
<td>Eur Am</td>
<td>230</td>
</tr>
<tr>
<td>Eligible Free/Red</td>
<td>210</td>
</tr>
<tr>
<td>Not Eligible</td>
<td>220</td>
</tr>
</tbody>
</table>
School Districts with Lowest Graduation Rates (2008)

- 41. Oakland
- 42. Los Angeles
- 43. New York
- 44. Dallas
- 45. Minneapolis
- 46. Columbus
- 47. Baltimore
- 48. Cleveland
- 49. Indianapolis
- 50. Detroit

- 45.6%
- 45.3%
- 45.2%
- 44.4%
- 43.7%
- 40.9%
- 34.6%
- 34.1%
- 30.5%
- 24.9%
Educational Context II: Excellence Gaps

Plucker, Burroughs, & Song (2010)
Plucker, Hardesty, & Burroughs (2013)
Performance Levels
National Assessment of Educational Progress (NAEP)

- **Advanced:** Superior performance
- **Proficient:** Solid academic performance
- **Basic:** Partial mastery of skills
- **Below Basic:** Less than partial mastery
Percent of Students Scoring Advanced on NAEP Reading in Grade 4 (2011)
Percent of Students Scoring Advanced on NAEP Reading in Grade 8 (2011)
<table>
<thead>
<tr>
<th>Year</th>
<th>Math Grade 4</th>
<th>Math Grade 8</th>
<th>Math Grade 12</th>
<th>Reading Grade 4</th>
<th>Reading Grade 8</th>
<th>Reading Grade 12</th>
</tr>
</thead>
<tbody>
<tr>
<td>1996</td>
<td>2.2</td>
<td>3.7</td>
<td>2.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1998</td>
<td></td>
<td></td>
<td>7.1</td>
<td>2.6</td>
<td>5.6</td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>2.5</td>
<td>4.7</td>
<td>2.4</td>
<td>6.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2002</td>
<td></td>
<td></td>
<td>7.1</td>
<td>2.8</td>
<td>4.5</td>
<td></td>
</tr>
<tr>
<td>2003</td>
<td>3.9</td>
<td>5.4</td>
<td></td>
<td>7.7</td>
<td>3.2</td>
<td></td>
</tr>
<tr>
<td>2005</td>
<td>5.0</td>
<td>6.0</td>
<td>2.2</td>
<td>7.5</td>
<td>3.0</td>
<td>4.6</td>
</tr>
<tr>
<td>2007</td>
<td>5.6</td>
<td>7.0</td>
<td></td>
<td>7.9</td>
<td>2.8</td>
<td></td>
</tr>
<tr>
<td>2009</td>
<td>5.9</td>
<td>7.9</td>
<td>2.7</td>
<td>7.7</td>
<td>2.8</td>
<td>5.3</td>
</tr>
<tr>
<td>2011</td>
<td>6.7</td>
<td>8.3</td>
<td></td>
<td>8.0</td>
<td>3.4</td>
<td></td>
</tr>
</tbody>
</table>
Educational Context III: Impact of Teachers

Teaching

1. Formative Evaluation
2. Microteaching
3. Teacher Clarity
4. Teacher-student relationships
5. Teaching meta-cognition
Curricula

1. Vocabulary programs
2. Repeated reading
3. Creativity
4. Phonics
5. Comprehension
1. Acceleration
2. Controlling classroom behavior
3. Classroom climate
4. Small group learning
Individual Student

1. Intelligence
2. Prior achievement
3. Persistence/engagement
4. Motivation
5. Preschool
Political Context

1. Accountability
2. Standards
3. Accreditation
Political Context 1

- Increased Focus on Accountability in Education
  - No Child Left Behind
  - Race to the Top

- K-12
  - Charter schools
  - Teacher pay incentives
Political Context II

- Increased focus on National Standards
  - Student performance on TIMSS, PISA
  - Next Generation Science Standards
  - Common Core

- Higher Education
  - University as vocational education
  - Focus on for-profit providers
  - CHEA (increased standardization)
Political Context III

• Changes in teacher education accreditation
  • Merging of NCATE and TEAC
  • Commission for the Accreditation of Education Programs (CAEP)

• Concerns about the “value” of teacher preparation programs
  • Very few programs are not re-accredited.
  • Teach for America
  • Is teacher education effective?
  • Is teacher education necessary?
Assessing and Evaluating Teacher Education Programs

A Report of a Task Force convened by the Board of Educational Affairs of the American Psychological Association
Goal of Task Force

- Not intended to advocate solely.
- What are the current methods in use today?
- What does the literature have to say about these methods?
- Under what conditions can and should they be used?
- Intended to provide guidance to programs, CAEP, and policy makers.
Task Force Members

- Mary Brabeck, New York University
- Carol Anne Dwyer, Educational Testing Service
- Kurt Geisinger, Buros Institute, University of Nebraska, Lincoln
- Ron Marx, University of Arizona
- George Noell, Louisiana State University
- Robert Pianta, University of Virginia
- Frank C. Worrell (Chair), University of California, Berkeley
Review Process: Sponsors

- APA’s Board of Educational Affairs
- CAEP Commission on Standards and Performance Reporting
- Coalition for Psychology in the Schools and Education
- Committee on Psychological Tests and Assessment
Review Process: Other Groups

- Data Quality Campaign
- National Council on Measurement in Education
- National Center for Education Statistics
- West Ed
- Southern Regional Education Board
- Carnegie Foundation
- New York State Department of Education
- Council of Chief State School Officers
- American Association for Colleges of Teacher Education
- American Association of Universities Education Deans
- U.S. Department of Education
Principles I

• Evaluating student learning is a critical element of effective teaching and therefore should be an ongoing part of preparation.

• Distinguishing more-effective from less-effective teaching validly, reliably, and fairly, although difficult, is possible.
Principles II

• Using multiple sources of data will result in better-quality data for making decisions.

• The design of explicit feedback loops from the data into program improvement activities is an important requirement of a good assessment process.

• Standardization may take different forms with different types of assessments, but the underlying principle is the same.
Principles III

- Thorough and effective training in analyzing and using data for decision-making will be necessary to create a valid, fair, and useful assessment system.

- Important decisions that could benefit from improved data are being made every day and will continue to be made whether or not high-quality data are available.
Technical Considerations I

- **Standards for Educational and Psychological Testing**

- Validity - meaningfulness that is associated with a set of values or scores.

- Reliability and precision of scores also important.

- Fairness - minimizing construct -irrelevant variance.
The use of measures in teacher education has two primary functions: (a) program improvement and (b) accountability.

Formative evaluations of programs should provide information to programs that predicts accurately how they will do in a formal summative evaluation.

Teacher education program improvement is a necessary but not sufficient condition of positive summative evaluations. All programs should be focused upon improvement constantly.
Technical Considerations III

- To determine the validity of the information collected for the program review, one must decide the extent to which the measures being used in the evaluation relate to student teacher learning and student learning outcomes and the extent to which they do not.

- All program decisions involve judgments that use collected data as well as informed judgments of the professionals involved. Standard-setting procedures are available that can help a program decide it is succeeding.
Sources of Data in Use

- Surveys
- Observations
- Learning outcomes of students
Using Surveys for Evaluation

- Surveys are among the most longstanding and commonly used methods of evaluating teacher performance.
- Surveys of Master and Supervising teachers are also used with some frequency to evaluate teacher education candidates.
- Until recently, surveys have not been used as part of a comprehensive means of evaluating teacher preparation programs.
Types of Surveys that Are/Can be Used to Evaluate Teacher Education Programs

- Surveys given to teacher education candidates.
- Surveys given to recent graduates of teacher education programs.
- Surveys of employers and supervisors of graduates of teacher education programs.
- Surveys of students being instructed by recent teacher education program graduates.
Surveys of Teachers/Teacher Education Candidates

- Satisfaction with teacher education program attended.
- Perceptions of preparedness for the teaching profession.
- Perceptions of perceived competence in teaching.
- Types of teaching practices/behaviors that they are using that they learned in the program.
- How well are they able to teach to the “Standards.”
- Perceptions of gaps in training - lessons that they wish they had been taught.
Strengths and Limitations of Surveys of Student Teachers/Teachers

• **Strengths:**
  • Can collect data on cohorts of teacher education program graduates at relatively low cost (especially with the advent of internet-based data collection).
  • Allow for easy comparisons across programs, cohorts, and individuals.

• **Limitations:**
  • Often developed locally: no evidence of psychometric rigor and predictive validity with student achievement.
  • Potential bias in self-reports.
  • Low response rates can limit generalizability.
Surveys of Employers/Supervisors

- Perceptions of general preparedness for the teaching profession.
- Perceptions of perceived competence in teaching.
- Types of teaching practices/behaviors that they are using.
- How well are they able to teach to the “Standards.”
- Perceptions of gaps in training.
Strengths of Surveys of Employers/Supervisors

- Potentially cost-effective.
- Allows for easy comparisons across programs and cohorts.
- May distinguish between most- and least-effective teachers (principals).
- Some evidence of predictive validity with student achievement (principals).
Limitations of Surveys of Employers/Supervisors

- Concerns about lack of standardization and psychometric rigor.
- Often completed without good sense of what teacher is actually doing.
- Low response rates can limit generalizability.
Surveys of Students

- Used more in higher education than K-12 settings.

- Teaching behaviors from the perspective of students
  - E.g., Clarity, interaction, organization, enthusiasm/interest, respect/care, challenge

- Perceptions of how much students are learning in class.

- Perceptions of expectations for student achievement and behavior and any differential treatment shown toward students.
Strengths of Surveys of Students

- Provides multiple raters who have seen teacher in action frequently.
- Low inference behavioral ratings are related to achievement outcomes.
- Considerable research base in higher education that is potentially transferable.
- Some research with K-12 populations.
Limitations of Surveys of Students

- May not yield reliable scores/valid inferences in primary grades.
- May be affected by course difficulty, grading leniency, lack of curricular knowledge.
Survey Use in Teacher Education

- Can be developed to match curricular and program standards.
- Can be checked for reliability and validity.
- With collaboration among programs and districts, can provide a common set of questions that allow for program, cohort, and individual comparisons.
- Can be useful in identifying areas of weakness or areas for remediation.
- Can be useful in monitoring growth of skill levels.
Using Standardized Observations to Evaluate Teacher Education Programs

- Observations of teachers’ interactions and classroom processes help identify effective practices and can be a valuable tool in building capacity for teaching and learning.

- Three components of standardization should be considered when evaluating an observation instrument:
  1. the training protocol
  2. the policies and procedures for carrying out the observations
  3. scoring directions.
The Training Protocol

- Are there directions for use?

- Is a training manual available?

- Are there guidelines for how to prepare to become an observer, and what measures show an observer has met a “gold standard”?
Procedures

- Is there information provided about how long the observation should last, best time of day to conduct the observation, to what degree the person evaluated can select the conditions of the observation?

- Is scoring conducted during or after? Is a rubric provided with examples of practices that correspond to different scores?
Using Standardized Observations to Evaluate Teacher Education Programs

- The observation measures need to be demonstrably consistent across time.

- Scores on the observation instrument need to have established empirical associations with student achievement and other learning outcomes.

- If these criteria are met, standardized observation instruments can be useful at all stages of a teacher’s development from candidate, to novice, to experienced professional.
Value-Added Models

- Student learning outcomes have emerged as the pre-eminent concern of stakeholders in assessing teacher preparation.
  - Achievement gap
  - Excellence gap
  - NAEP, PISA, TIMSS
Using VAAs to Assess Preparation Programs

- In tested grades and subjects, research has emerged suggesting value added assessments may be useful.

- Current value added assessments can leverage large data systems across large districts or states to deal with the challenges of measures and geographic dispersion of program completers.

- Aggregating across many graduates, schools, and districts provides options that are not available when assessing individual teachers.
Value-Added Assessment Strengths

- In some areas (e.g., teaching reading decoding skills to students in the early grades), well-developed progress-monitoring measures are available with extensively documented criterion- and/or norm-referenced standards for acceptable performance or growth.

- In these cases, using student progress measures for students taught by teacher candidates is feasible. Additionally, these types of measures lend themselves relatively directly to examining aggregate and program results.
VAA Concerns in Evaluating Preparation Programs

- In most subject areas (e.g., high school biology, instrumental band, special education for severely disabled students), well-developed and technically adequate measures are not yet available.

- In these contexts, the only currently viable method is to devise explicit learning targets that are directly tied to immediate instructional goals and that can be directly and practically measured.

- One challenge for this type of student learning assessment is the establishment of standards for candidate performance and the aggregation of dissimilar data for program appraisal and improvement. Results will have to be converted to a common metric such as effect size or goal attainment scaling.
Concerns re Student Teachers

- Need to separate the efficacy of the teacher education candidates from other intertwined factors such as efficacy of supervising teacher.

- Can student teachers be responsible for teaching specific content?

- Can be useful in looking at efficacy of candidates in alternative certification programs where teacher candidate is teacher of record for the year.
Other Concerns re VAA

• Considerable controversy has emerged regarding the precision of value added estimates, their utility, and the degree to which they can accurately assess the contribution of programs to learning gains.

• Value added measures can contribute to narrowing the curriculum to what is tested.

• Value added assessments are complex and require a high level of technical expertise to implement well.

• Value added assessments do not sufficiently control other inputs into learning.
VAA Concerns Continued

- Data difficult to obtain
  - Geographic dispersion, finances, diversity of subject matter.

- Heterogeneity of classrooms
  - Starting points, ceiling effects, children with disabilities

- Need partnerships involving states education agencies, school districts, & programs.

- Need standardized data like student learning objectives (SLOs) that meet quality assurance thresholds.
A Question and A Comment

• Are surveys, observations, and value added assessments more or less reliable than the measures that are currently in use?

• We should not let the perfect be the enemy of the good. Decisions about program effectiveness need to be made using the most trustworthy data and methods currently available.
Recommendations

1. Require the use of strong empirical evidence of positive impact of program graduates on student learning.

2. Design statewide, longitudinal data systems that collect performance data with good technical quality that address the following stages of teacher preparation:
   - Selection
   - Progression
   - Program completion
   - Post-graduation
Recommendations

3. Track program elements and candidate attributes that predict positive contributions to PreK-12 student learning.

4. Develop valid measures of student learning outcomes for all school subjects and grades to assess student learning outcomes similar to those available in math, language arts, and science.

5. Dedicate appropriate resources for data collection and analysis
   - Assign time for faculty and professional staff to collect pupil and teacher data
   - Analyze and use data regularly for program improvement
Recommendations

6. Identify and retain staff with technical skills, time, and resources to analyze data.
   • Partner with school districts and state agencies on data access and analysis

7. Commit to a system of continuous improvement based on examination of program data.
   • Allocate sufficient time and resources for faculty to review and reflect on findings
   • Use findings from data analysis for annual program improvement efforts
   • Document use of data for continuous improvement
8. Train faculty and supervising teachers in the use of well-validated observations systems
   - Develop a system for regular reliability checks so observations are conducted with a high degree of fidelity
   - Implement observation systems at appropriate points in the preparation pathway and use data for feedback to candidates and program

9. Identify and develop student surveys that predict preK-12 student achievement
   - Develop baseline data with large enough samples to conduct psychometric analysis that lead to benchmark performance levels
   - Use data for continuous feedback to candidates and programs
10. Develop and validate developmental benchmarks and multiple metrics for graduation decisions to ensure graduates are proficient teachers who can influence student learning.

11. Develop curricula that prepare teacher candidates in the use of data so that candidates can continue to self-assess and faculty can assess their students’ progress.

12. Report to the public regularly on any adverse impact of implementation of assessments on the teaching force or preK-12 learning.
Questions