Teacher Value-Added and Credentials as Tools for School Improvement

Douglas N. Harris
University of Wisconsin at Madison
Summary

Discuss the logic of the teacher credential and value-added strategies

Discuss how well teacher credentials and value-added measure actual teacher effectiveness

Interpret this evidence for the purpose of educational policy (e.g., accountability)

Conclusion: A mixture of credentials and value-added measures is probably the best way to improve schools.
The Traditional “Credentials Strategy” to Teacher Quality

- Until 1990s, the education system had a “resource focus”—finance, class size, and others . . .
- Teacher credentials also fall within the input approach
  - Undergraduate Education and Test Scores: Used as a frequent basis for initial certification
  - Professional Development: Used as a basis for maintaining certification and improving teacher effectiveness
  - Graduate Education and Experience: Used as a basis for compensation
Arguments For Change

- Teachers with more education, higher test scores, and more advanced levels of certification don’t seem much better than others.
- Schools don’t focus on student outcomes and have insufficient incentives for improvement.
- Larger perceived failures of school system—need to “do something”
Possible Directions for Change

- Change the specific teacher credentials but maintain the traditional strategy.
- Alternatively, switch to an “accountability strategy,” holding teachers accountable for:
  (a) “subjective” performance measures
  (b) “objective” performance measures
- The choice of strategies and policies depends partly on how well the measures predict actual teacher performance.
Analogy: The Performance of Politicians

What is a “high-performing” politician?
- Definition for illustrative purposes: One who makes the average constituent “better off” than they would have been otherwise—better than expected (Alternatively, better off than she would have been if the politician’s opponent had won the last election?!

What if we relied on credentials?
- University degrees? Number of bills passed? Number of campaign endorsements?

What about measuring performance directly?
- Changes in social and economic conditions?
Analogy Continued

- Hard to know what would have happened if someone else had been elected

- Changes in social and economic conditions are not entirely within the control of politicians; likewise, student achievement scores are not entirely within the control of teachers

- Even when we know what is expected, there is considerable “random error”—1992 and 2008 elections as examples

- Measuring performance is hard in any job
How Well Can Value-Added Modeling (VAM) Measure the Performance of Teachers?
Basics of VAM

- Annual student testing allows researchers to track the “growth trajectory” of students.
- Teachers whose students make greater than expected growth have “high value-added”.
- “Expected growth” is based on each student’s average growth trajectory, adjusted for school resources students receive along the way.
- Value-added and federal “growth-to-proficiency” models are not the same.
Why Value-Added is Important

- Researchers generally worried about the problem of “non-random assignment”

- Specifically, in this case: certain types of students assigned to certain types of teachers
  - some teachers assigned to students with lower achievement levels

- Point-in-time snapshots (levels) of student achievement are therefore highly inaccurate measures of teacher performance
  - Problem in NCLB-AYP and school report cards
In Terms of Value-Added, Teachers A and B are Performing Equally.

Teacher A: “Success” on Ach. Levels
Teacher B: “Failure” on Ach. Levels
A Teacher with Low-Proficiency Students Can Still be High Value-Added (and vice versa)
Two Uses of VAM

(1) VAM for Program Evaluation (VAM-P)
   - Evaluate educational programs, such as professional development
   - Identify characteristics of effective teachers

(2) VAM for Accountability (VAM-A)
   - Evaluate individual educational personnel, especially teachers
   - e.g., Bill Sanders, Tennessee model; Dallas
What VAM-P Research Says about Teacher Credentials
## Summary of Evidence on Credentials

<table>
<thead>
<tr>
<th>Teacher Credential</th>
<th>Gain Score Studies</th>
<th>VAM-P/Experiment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Undergraduate</td>
<td>5  4</td>
<td>0  3</td>
</tr>
<tr>
<td>Graduate</td>
<td>3  10</td>
<td>3  6</td>
</tr>
<tr>
<td>Prof. Develop.</td>
<td>0  1</td>
<td>2  1</td>
</tr>
<tr>
<td>Experience</td>
<td>7  8</td>
<td>8  1</td>
</tr>
<tr>
<td>Teacher test sc.</td>
<td>5  2</td>
<td>1  1</td>
</tr>
</tbody>
</table>
A few good signs, e.g., teacher experience, but:

- Little evidence that graduate degrees are associated with teacher value-added

- Evidence is mixed on validity of certification tests

- National Board (NBPTS) certification seems to be a modest predictor of effectiveness

- Overall, credentials are not very good predictors of teacher effectiveness
What About Evidence on Value-Added for Accountability (VAM-A)?

- Research on VAM-A is in its infancy, but . . .
- Differences between the lowest and highest value-added teachers seem large
- VAM-A measures have been partly validated by a random assignment experiment
- VAM-A measures of teacher effectiveness are positively correlated with principals’ subjective assessments of teachers
BUT Teacher VAM-A measures also:

- Are imprecise
  - hard to say that one teacher is clearly better than another based on VAM-A
  - this problem also arises with other measures as well

- Vary across tests (same subject)

- May not totally address the “assignment problem”
Also, Problematic Assumptions

- VAM based on assumptions that may not be true:
  - student tests are interval scaled?
  - assignment based on fixed student characteristics?
  - teachers equally effective with all students?
  - data missing at random?

- The first two assumptions seem especially problematic
What Does It All Mean for Education Policy?
Factors to Consider

- All measures have their advantages and disadvantages
- Question is how they compare, on three criteria:
  1. how well they predict performance
  2. purpose
  3. cost
Purposes of Quality Measures

- **Signaling** is about determining who starts off in the profession
  - examples: certification; entry requirements to teacher education programs

- **Improvement** is about raising effectiveness for those who have started in the profession
  - type #1: summative, to provide incentives for improvement (accountability)
  - type #2: formative, to identify a path to improvement
Different Teacher Quality Measures Serve Different Purposes

- Consider some policies using teacher quality measures:
  - recommended for entry into profession (signal)
  - required for entry into profession (signal)
  - basis for compensation (improve)
  - dismissal from a job (improve)

- Recommendation for entry involves lower stakes than dismissal from a job
  - appropriate level of stakes should be proportional to the measures usefulness as a signal or improvement measure (and costs)
Costs of Quality Measures

- Often neglected topic in all areas of education
- Some measures much more costly than others:
  - a master’s degree for a teacher costs at least $80,000 in budgetary and non-budgetary costs
  - paper and pencil tests cost little
- VAM-A calculations are cheap but providing incentives based on them can be expensive
- An especially important topic in the current economic climate
Some Probably Inappropriate Uses in Policy

- Based on this logic, compensation based on graduate degrees is the least valid common use of a teacher quality measure
  - given moderate-high weight (thru compensation) despite high cost and low statistical validity

- Attaching very high stakes to teacher value-added scores would be equally invalid
  - value-added is inexpensive and more directly focused on effectiveness, but . . .
  - untested assumptions and disconcerting findings
Some Potentially Valid Uses

- Give preference to teachers with a master’s degree when selecting mentor or master teachers
- Teacher value-added as one of several indicators of teacher effectiveness for purposes of hiring, compensation, and dismissal
- Include pedagogical content knowledge and classroom experience as requirements for state-approved teacher education programs
- In some sense, this is a call for “multiple measures,” but not just any measures or uses
Teacher value-added is not the only way to use student achievement scores. Two examples...

School VAM-A doesn’t provide information about how each teacher is doing, but:

(a) school VAM-A creates pressures; and
(b) perhaps “everybody knows” who the good teachers are, creating peer/administrator

Provide raw student sub-scores to teachers (along with school/district averages)
- provides summative measures and path to improvement
- but non-random assignment; and no incentives
Conclusions: Part I

- In deciding how to use VAM-A, we must:
  1. compare it to the alternatives; and
  2. consider predictive validity in relation to specific purposes and costs

- Despite the problems, teacher VAM-A appears more cost-effective than credentials as signals and summative assessments

- However, we need a path to improvement, and existing credentials may serve that function
  - we know little about this, except experience
Key Conclusion: We Should Move Forward with Careful Experimentation

- Value-added shows some real potential, but we’ll only know whether it is useful if some school districts or states try some different approaches AND rigorously evaluate them.

- For this reason, the approach taken by Congressman Miller (D-CA) is better than that of the New York State Legislature.
Papers and References

- My web site:
  http://www.education.wisc.edu/eps/faculty/harris.asp

- Web site focused on teacher quality research:
  http://www.teacherqualityresearch.org

- Ed Week Commentary (June, 2008)

- National Conference on Value-Added