Evaluating Teacher Quality

Jan Bietenbeck

Lund University

Madrid, May 4, 2017
What is teacher quality and why is it important?

- **Teacher quality** = the ability to increase students’ knowledge and skills (economists: “students’ human capital”), including:
  - math and reading skills,
  - critical thinking and reasoning skills,
  - personality traits (ability to work in a team, grit, ...), etc.

- **General agreement among researchers**: teachers are the most important school-based factor affecting learning.
  - Much more important than physical resources (books, computers, ...).
  - Different teacher quality ⇒ cross-country differences in PISA scores?
To measure teacher quality need to measure students’ skills

- Evaluating teacher quality = evaluating how good teachers are at raising students’ human capital ⇒ need to measure this!

- Problem: no all-encompassing measure of human capital exists.

- Most research in economics: approximate human capital by students’ performance on standardized tests which assess cognitive skills (in math, reading, ...).
Test scores as a measure of human capital

- **Key advantage of standardized test scores**: objective, comparable across students in different classes and schools (unlike grades).

- **Standardized tests already used in many school systems** (in the United States, England, Madrid; as part of PISA, PIRLS, TIMSS).

- **An informative measure**: test scores predict economic growth at the country level and individual earnings later in life.
Standardized test scores and later-life earnings

Source: adapted from Chetty et al. (2011)
Teacher quality \(\approx\) ability to raise test scores

- For much of the remaining talk: teacher quality = ability to raise students' performance on standardized tests.

- But keep in mind that while very informative, test scores are no perfect measure of human capital:
  - Do not measure personality traits (ability to work in teams etc.).
  - Will discuss the importance of this later on.
Outline for the remainder of the talk

- **Remainder of this talk:**
  - What makes a good (=high-quality) teacher?
  - How can we identify high-quality teachers in data?
  - Which policies can be used to raise teacher quality?

- **Summarize research from economics of education.**
  - Distinguishing feature: focus on causal relationships.
  - Evidence mostly from outside of Spain because of data availability.
Talk outline

1. What is teacher quality and why is it important?

2. “CV measures” of teacher quality

3. A direct measure of quality: teacher value-added

4. Teaching practices and classroom observations

5. Conclusion
Talk outline

1. What is teacher quality and why is it important?

2. “CV measures” of teacher quality

3. A direct measure of quality: teacher value-added

4. Teaching practices and classroom observations

5. Conclusion
“CV measures” of teacher quality: what are they?

“CV measures” = teacher characteristics which are easily observed by employers when hiring a teacher – those usually stated on an applicant’s curriculum vitae (CV).

Three measures frequently used for hiring and salary decisions:

1. Teacher experience: years of tenure on the job.
2. Educational credentials: master’s degree, doctorate (Ph.D.).
3. Test scores from a teacher entrance exam (e.g. oposición in Spain).
Teacher experience as a measure of teacher quality

- In many school systems around the world, teacher salaries rise in lockstep with experience.

- But research has generally found that experience matters for student learning only during the first few (3-5) years.

- Thus, experience is of limited use as a measure of teacher quality and paying teachers by experience ≠ paying for quality.
Teacher experience and student math scores

Source: own figure based on data from Wiswall (2013)
Educational credentials as a measure of teacher quality

- Many school systems partly base hiring and salary decisions on whether a teacher has a master’s / Ph.D. degree.

- Research again shows that such educational credentials are not consistently related to student performance ⇒ they are not good measures of teacher quality.
Scores from teacher qualification exams (such as the oposición) are often used for hiring decisions. While not usually reported on CVs, they are still observable to employers.

At best mixed track record of scores as a measure of quality:
- Performance on tests which measure mostly general cognitive skills, IQ, and general personality traits usually not predictive of student scores.
- Even if related to student performance, teacher test scores explain very little of its variation ⇒ not a good measure of teacher quality.
CV measures” of teacher quality: summing up

“CV measures” such as experience and educational credentials are often used to inform hiring decisions and linked to pay.

But they are actually not measures of teacher quality:
- They don’t/hardly predict students’ performance on standardized tests.
- Recent research: also not predictive of students’ personality traits.
Talk outline

1. What is teacher quality and why is it important?
2. “CV measures” of teacher quality
3. A direct measure of quality: teacher value-added
4. Teaching practices and classroom observations
5. Conclusion
Around 15 years ago, economists started to develop methods to directly quantify individual teachers’ effectiveness (rather than searching for specific characteristics which predict student scores).

They came up with teacher value-added ("TVA"): very roughly, this is a measure of how much a teacher improves her students’ test scores from one year to the next, on average.
Imagine a primary school with two fourth-grade classes. All students in this school are tested in math at the end of third and fourth grade, with the following results:

<table>
<thead>
<tr>
<th></th>
<th>Class 1</th>
<th>Class 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg. score in 3rd grade</td>
<td>100</td>
<td>150</td>
</tr>
<tr>
<td>Avg. score in 4th grade</td>
<td>150</td>
<td>175</td>
</tr>
<tr>
<td>4th-grade teacher value-added</td>
<td>50</td>
<td>25</td>
</tr>
</tbody>
</table>

Teacher value-added is then computed as the difference between the average fourth-grade and third-grade scores in each class. Here, the teacher in Class 1 has higher value-added = “has higher quality.”
Teacher value-added: a fair measure

- Key idea of TVA: a fair comparison of teachers’ effectiveness.

- This is why it measures test score growth, not levels.
  - In terms of test score levels, students in Class 2 have higher scores at the end of fourth grade (175 points vs. 150 points in Class 1).
  - But they were also smarter to begin with, so their teacher didn’t really have to do much to bring them up to that level.
  - By looking at test score growth, we are leveling the playing field. We now see that the teacher of Class 1 is actually more effective.
Note that in reality, the statistical calculations behind TVA are far more complicated and take into account a lot of other concerns.

Recent efforts to validate the usefulness of TVA as a measure of teacher quality show that:

- Comparisons based on teacher value-added are indeed fair.
- TVA is highly predictive of a teacher’s future ability to raise scores.
- TVA is highly predictive of students’ earnings later in life.
Teacher value-added and students' earnings

- Having a teacher that is at 84th vs. 50th percentile of TVA for a single grade raises undiscounted lifetime earnings by ca. $25,000.

Source: adapted from Chetty et al. (2014)
As of today, most teachers’ contracts are completely unrelated to their TVA ⇒ effective teachers don’t get rewarded for their good work, and ineffective teachers don’t face any consequences.

How can TVA be used for policy?
- Cannot use TVA in hiring decisions for first-time teachers, as one needs to observe their previous students to calculate TVA.
- But can pay teachers or even fire them based on their calculated TVA...
Value-added policies in practice

- **Washington D.C.**: teachers get fired and offered salary bonuses based on a metric that puts 50% weight on TVA. Similar policies in some other school districts in the United States.

- **These policies are highly controversial**, not least because very recent evidence shows that some low-TVA teachers are actually good at raising students’ behavioral skills (but not their test scores).
  - Means that teachers who actually add value are underpaid or fired.
Teacher value-added: some final thoughts

- To implement TVA policies, need to set up an extensive student testing system. While costly, the benefits likely outweigh the costs.

- TVA helps us distinguish effective from ineffective teachers. But it doesn’t tell us what exactly makes a good teacher!
  - Cannot learn anything about how to better train teachers.
  - Policies that pay/fire teachers based on TVA implicitly assume that teachers themselves know how to improve - but this might not be true.
Talk outline

1. What is teacher quality and why is it important?

2. “CV measures” of teacher quality

3. A direct measure of quality: teacher value-added

4. Teaching practices and classroom observations

5. Conclusion
Some recent studies try to link teacher quality to teaching practices – what teachers actually do in the classroom.

Example: lecturing versus student group work – which is better?

Studies show that teaching practices seem to matter. But they suffer from important limitations:

- Difficult to measure teaching practices well in existing data.
- Causality is very difficult to establish.
Classroom observation measures of teacher quality

- Related research tries to identify effective teachers using classroom observations by teacher peers or experts:
  - Teacher peers sit in the back of the classroom and evaluate.
  - Videotaped lessons get sent to independent experts for quality review.

- Peer evaluation already takes place in many schools, but
  - it is usually very infrequent, and
  - hardly any teacher gets bad grades, which means that grades are not closely related to actual teacher quality (as measured by TVA, for ex.).
How classroom observations can work

- **Classroom observations *can* help identify effective teachers if**
  - teachers are evaluated based on more than one lesson,
  - there are clear-cut criteria on which teachers are evaluated,
  - and evaluators are well-trained and willing to also give bad grades.

- **Teacher quality measures from such well-designed systems are**
  - highly correlated with teacher value-added on test scores,
  - but also predictive of some student behaviors ⇒ are thus a potentially more complete measure of teacher quality.
Classroom observations: some thoughts

- **In practice, evaluation systems based on classroom observations are likely difficult and costly to implement.**
  - Disadvantage compared to teacher value-added systems: learn only about teacher quality, not about student achievement.

- **An advantage of classroom observations:** evaluation criteria may give us insights into what makes a good teacher!
  - Can use these insights in teacher training and peer mentoring programs.
1. What is teacher quality and why is it important?

2. “CV measures” of teacher quality

3. A direct measure of quality: teacher value-added

4. Teaching practices and classroom observations

5. Conclusion
Currently, teacher salaries and hiring/firing decisions are often based on “CV measures.” But these are actually not (or only very weakly) related to teacher quality.

Can distinguish good from bad teachers using value-added.

- Can improve overall teacher quality by making pay and/or tenure decisions conditional on value-added.
- But value-added should not be the only quality measure, as it misses impacts on human capital not captured by test scores ⇒ potentially combine with classroom observations.
