Carrell and West, “Does Professor Quality Matter?”

How would you measure “professor quality”?

Student evaluation scores?
* may be affected by (expected) grades, prior interest in course, and/or other irrelevant factors
* may be affected by “endogenous selection” – different students self-select into courses of different professors (so evaluations won’t tell us how good a given professor is at teaching a given student)

Student “value-added” (gain in student test scores between two dates)?
* may be affected by endogenous selection (as above)
* may not be long-lasting
* may be affected by “teaching to the test”
* hard to control for all relevant background factors (prior preparation)
Air Force Academy (AFA) offers unique situation/data

- students randomly assigned to intro professors (no self-selection)
- syllabus is identical for all faculty in the same course
- standardized exams given to all students in the same course
- students randomly assigned to numerous follow-on courses
  (again, no self-selection)
  (allows research to distinguish between initial and later impact)
- large enrollments, lots of different variables are recorded, over a period of many years
- small class size (average $\approx 20!$), prof can potentially have big impact
Basic regression setup:

\[ Y^1_{i,t,j1,j2,s1,s2} = X_{i,t,s1} \beta^1 + \gamma^1_i + \lambda^1_{j1} + \lambda^1_{j2} + \xi^1_{t,s1} + \xi^1_{t,s2} + \epsilon^1_{i,t,j1,j2,s1,s2} \]  
(intro courses)

\[ Y^2_{i,t,j1,j2,s1,s2} = X_{i,t,s2} \beta^2 + \gamma^2_i + \lambda^2_{j1} + \lambda^2_{j2} + \xi^2_{t,s1} + \xi^2_{t,s2} + \epsilon^2_{i,t,j1,j2,s1,s2} \]  
(follow-on courses)

where \( Y^1_{i,t,j1,j2,s1,s2} \) = points earned by student \( i \) in semester \( t \)

with intro prof \( j1 \) in section \( s1 \) and follow-on prof \( j2 \) in section \( s2 \)

for an intro course (superscript = 1);

similarly for follow-on (superscript = 2)

\( X_{i,t,s1} \) = student-specific and classroom-mean characteristics (SAT scores, etc.)

(note: allows for "peer effects")

\( \gamma^1_i \) = dummy variables for course and semester

(controls for grading standards/achievement over time)

\( \lambda^1_{j1} \) = dummy variables for intro prof \( j1 \) effect in intro course

\( \lambda^1_{j2} \) = dummy variables for follow-on prof \( j2 \) in intro course

\( \xi^1_{t,s1}, \xi^1_{t,s2} \) = dummy variables for section of intro/follow-on course

\( \epsilon^1_{i,t,j1,j2,s1,s2} \) = conventional "error term"

...and similarly for follow-on courses.
Basic results:
Carrell and West report the *variance components* for the regressions. For example, the variance in intro professor quality in the intro course is

\[ \text{Var}(\lambda_{j1}^1) = 0.0028, \text{ so } \text{SD}(\lambda_{j1}^1) = 0.052 \]

"This result indicates that a one-standard deviation change in professor quality results in a 0.05-standard-deviation-change in student achievement. In terms of scores, this effect translates into about 0.6 percent of the final percentage of points earned in the course." (p. 421)

But is this effect long-lasting? Or simply due to teaching to the test? (NB: AFA instructors know what the exam will look like (everyone takes a common exam)!!)

Now consider the variance in intro professor quality in the follow-on course, which is

\[ \text{Var}(\lambda_{j1}^2) = 0.0025, \text{ so } \text{SD}(\lambda_{j1}^2) = 0.050 – \text{ just about the same as for } \text{SD}(\lambda_{j1}^1). \]

But what is the relationship between \( \lambda_{j1}^2 \) and \( \lambda_{j1}^1 \)? For this we need their *covariance*.

This covariance is not significantly different from zero! Thus, "...our estimates... indicate that one set of calculus professors produce students who perform relatively better in calculus and another set of calculus professors produce students who perform well in follow-on related courses, and these sets of professors are not the same." (p. 424)
Useful (and reassuring) check: effect of follow-on course professors... on intro course scores is zero (!), but... on follow-on course scores is significant. (pp. 425-426)

Likewise, section-specific shocks in follow-on course don’t affect intro scores, but section-specific shocks in intro course do affect follow-on scores (p. 426).

**Relation between prof characteristics, student evaluations, and student achievement**

Regress intro professor’s effect on student achievement \((\lambda_{j_1}^1, \lambda_{j_1}^2)\) on...

(a) professor-characteristics variables (Table 5, upper panel)
(b) student-evaluation ratings (Table 5, lower panel)

Basic results:

highly-rated intro profs have positive, significant effect on intro achievement but have negative, significant effect on follow-on achievement “senior” profs have negative (not usually signif) effect on intro achievement but have positive (sometimes signif) effect on follow-on achievement
Selected results from Table 5 (p. 427):

<table>
<thead>
<tr>
<th>independent variable:</th>
<th>dependent variable:</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Associate/full professor</td>
<td>( \lambda^1_{j_1} )</td>
<td>-0.69</td>
<td>0.70</td>
</tr>
<tr>
<td></td>
<td>( \lambda^2_{j_1} )</td>
<td>(0.41)</td>
<td>(0.40)</td>
</tr>
<tr>
<td>Over three years’ teaching experience</td>
<td>( \lambda^1_{j_1} )</td>
<td>-0.79</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td>( \lambda^2_{j_1} )</td>
<td>(0.29)</td>
<td>(0.29)</td>
</tr>
<tr>
<td>Instructor provided clear well-organized instruction</td>
<td>( \lambda^1_{j_1} )</td>
<td>0.50</td>
<td>-0.46</td>
</tr>
<tr>
<td></td>
<td>( \lambda^2_{j_1} )</td>
<td>(0.19)</td>
<td>(0.20)</td>
</tr>
</tbody>
</table>

Rank-order of intro profs’ value-added on the intro course, \( \lambda^1_{j_1} \),
- is negatively-correlated with rank-order of intro profs’ value-added on the followup course, \( \lambda^2_{j_1} \)

Rank-order of student evaluation of instructor’s “effectiveness” is...
- positively-correlated with rank order of intro profs’ value-added for intro course, \( \lambda^1_{j_1} \),
- but negatively-correlated with rank-order of intro profs’ value-added for the followup course, \( \lambda^2_{j_1} \)

Rank-order of intro profs’ value-added on the intro course, \( \lambda^1_{j_1} \),
- is negatively-correlated with rank-order of intro profs’ value-added on the follow-up course, \( \lambda^2_{j_1} \)
How to account for these results?
all instructors get an advance copy of exams before they are given
junior profs may stick to the “regimented curriculum being tested,”
whereas senior profs may “broaden the curriculum, produce students with
deeper understanding of the material” (pp. 429-430)
students who are “taught to the test” may not develop good study habits
students of low value-added profs may have to work harder in follow-ons
(note that this means that $Y^1$ may affect $Y^2$ directly!)

So are student evaluations a good guide to professor quality?