Abstract: Classroom observation ratings make up the largest component of summary ratings given to teachers in the multiple-measure teacher evaluation systems many states have implemented in the last decade, but little research has examined observation ratings in a high-stakes setting. Using data from the first six years of implementation of teacher evaluation in Tennessee, we investigate whether nonrandom sorting of students and teachers and other potential sources of bias systematically lower the observation scores of some groups of teachers. In particular, we test the potential for biases that affect teachers differentially by race and gender. We find that White and female teachers outscore their Black and male colleagues, even when comparing teachers with otherwise similar characteristics in the same school with similar value-added scores. These gaps appear across rubric domains. A small portion of the Black-White gap is attributable to sorting of students within schools. Also, we find evidence that teachers receive somewhat higher ratings from raters of the same race. In contrast, we find no same-gender rater effects, and are in fact able to explain only a small fraction of the gender gap in ratings with contextual factors.

As for people to meet with during my visit: it would be great to meet with Steve, if he’s available, and with Elaine Allensworth, whom I know. But otherwise you all have a great group and I’d be happy to meet with any of your faculty or fellows. Also, I have a project just getting off the ground with Lauren Sartain (CCSR) and Molly Gordon (NORC). I will let them know I am coming to Chicago, and perhaps some time could be scheduled with them (maybe when I arrive on Monday since it sounds like Tuesday will be pretty full).