Teaching Practice and the Gender Gap in Reading Achievement

Esteban Aucejo, Jane Cooley Fruehwirth, Sean Kelly and Robert Bringe

Boys persistently lag behind girls in reading achievement by as much as 0.29 of a standard deviation in National Assessment of Education Progress scores, which is associated with approximately a year of learning (Loveless, 2015). Reardon et al. (2016) finds that the gap in state standardized tests of reading achievement (or English language arts (ELA)) is on average 0.23 of a standard deviation or about two-thirds of a year of learning. Reading skills are essential building blocks for learning in all subjects. Recent evidence also suggests that verbal skills may be more important determinants of college attendance than math skills (Aucejo and James, 2016).

We study a key potential school-based explanation for the gender gap in reading that has nonetheless received little support in the literature—that teaching practices favor girls to the detriment of boys. Reardon et al. (2016) documents substantial heterogeneity in gender gaps across school districts, which could suggest a role of school policies in contributing to literacy gaps, beyond other developmental explanations for the gap.

There are two candidate explanations for how teaching practices could explain the ELA gap. First, it could be that recommended teaching practices produce higher value-added in ELA for girls than for boys. In this case, a given practice simply has higher returns for girls relative to boys. Second, it could be that within the classroom, teachers differentially engage, care for, or focus efforts on girls relative to boys. In this case, boys and girls who share the same teacher receive different benefits through how the teachers differentiate their instruction.

The Measures of Effective Teaching (MET) Longitudinal database provides a unique lens for studying this given rich measures of teacher effectiveness based on principal surveys, student surveys and ratings from observation of trained raters based on popular general protocols, CLASS (Hamre et al., 2013; La Paro, Pianta, & Stuhlman, 2004) and Danielson Framework for Teaching (The Danielson Group, 2013), along with the reading-specific protocol PLATO (Grossman et al., 2014). For the first hypothesis, we simply test whether the benefits of the measured practices or ratings for ELA differ based on the gender of the student. With this hypothesis, one key potential confound is matching of teachers with certain types of practices to more male or female students. We test whether there is evidence of this in the data using balancing tests based on observable characteristics and find very little evidence of matching. That said, we show our estimates are robust to controlling for class fixed effects which would deal with any selection based on teacher or classroom unobservables. A second potential confound is that teacher video-observation ratings may inadvertently reflect the gender composition of the classroom (Campbell and Ronfeldt, 2018). We find some mixed evidence of this in the data. To address this, we use correlations between contemporaneous and prior teaching practices to ensure that they are not reflecting current classroom composition. Finally, random measurement error in teaching practices could bias the interactions between the male dummy and teaching practice toward 0. We address this by instrumenting contemporaneous with prior practices.

For the second hypothesis, we consider whether student-level reports of teaching practice (based on Tripod) explain the ELA gap. Boys tend to consistently report lower values

relative to girls across most domains of the Tripod. We consider whether these gaps are productive, in the sense that they explain boy/girl ELA gaps. Tripod captures different dimensions of teaching instruction. For instance, boys may relate differently to the teachers than girls do, e.g., they may feel less cared for by the teacher or less challenged by the teacher to perform well (pertaining to domains *care* and *challenge*). Alternatively, boys may perceive different use of instructional practices, e.g., boys may not feel as included in student-centered approaches or not respond in the same way and/or they may not find the same activities to be clarifying as girls do if they learn differently (pertaining to domains *confer* and *clarify*). Finally, boys may find certain types of course work less interesting than girls pertaining to the dimension of *captivate*. The key confounds for this analysis are potential reverse causality, in that the student's report about the teacher is more a reflection of their innate engagement than the teacher's practice, and measurement error. We deal with this by instrumenting student reports with lags of the teaching average scores from the previous year when she/he taught a different set of students in each of these domains. By doing so, we are able to speak to the potential that the gap is explained through teaching practices and not through innate characteristics of the student they make him/her report lower values of captivate, challenge, etc.

Despite being able to exploit unprecedentedly rich measures of teaching practice, we find very little evidence that boys benefit differentially from teaching practices. Thus, we reject hypothesis 1, that differences in the benefits of a given teaching practice explain the achievement gap. For hypothesis 2, we find that the student reports of teaching practice tend to reflect meaningful gaps in the experience of boys and girls in the classroom. Most strikingly, we find that domains related to engagement tend to explain the ELA achievement gap. This has important implications for approaches to narrowing the ELA achievement gap between boys and girls and suggests that adapting teaching practices that better engage boys would help to keep boys from losing ground relative to girls as they progress in school. Optimistically, our empirical strategy suggests that this is feasible given that some teachers are significantly better-able to engage boys relative to others, though more work needs to be done to understand the best strategies for this.