Decades worth of research on education production have narrowed in on the importance of teachers to student outcomes (Hanushek and Rivkin 2010; Murnane and Phillips 1981; Todd and Wolpin 2003). Over the last several years, these studies have coalesced around two key findings. First, teachers vary considerably in their ability to improve students’ academic performance (Hanushek and Rivkin 2010; Nye, Konstantopoulos, and Hedges 2004), which in turn influences a variety of long-term outcomes including teenage pregnancy rates, college attendance, and earnings in adulthood (Chetty, Friedman, and Rockoff 2014b). Second, experimental and quasi-experimental studies indicate that “value-added” approaches to estimating teachers’ contribution to student test scores are valid ways to identify effective teachers (Bacher-Hicks et al. 2015; Chetty, Friedman, and Rockoff 2014a; Glazerman and Pratik 2015; Kane et al. 2013; Kane and Staiger 2008). In other words, these teacher effect estimates are not confounded with the non-random sorting of teachers to students, the specific set of students in the classroom, or factors beyond teachers’ control. Policymakers have taken notice of these important findings, leading to widespread changes in teacher evaluation (Dee and Wyckoff 2015), compensation (Podgursky and Springer 2011), and promotion (Loeb, Miller, and Wyckoff 2015).

While the studies described above have focused predominantly on teachers’ contribution to students’ academic performance, the research community is starting to have evidence that teachers also impact a variety of other student outcomes in ways that are only weakly related to their effects on test scores (Blazar and Kraft forthcoming; Gershenson 2016; Jackson 2012; Jennings and DiPrete 2010; Kraft and Grace 2016; Ruzek et al. 2015). For example, in earlier work drawing on the same dataset used in this paper, Blazar and Kraft (forthcoming) found that teacher effects on students’ self-reported behavior in class, self-efficacy in math, and happiness in class were similar in magnitude to effects on math test scores. However, teachers who were effective at improving math test scores often were not equally effective at improving students’ attitudes and behaviors, with correlations between measures no higher than 0.19. Similarly, Jackson (2012) found that teacher effects on a composite measure of observed school behaviors, including suspensions, absences, grade point average, and on-time grade progression, explained five percent or less of the variation in teacher effects on students’ academic performance. Together, these findings lend empirical evidence to the multidimensional nature of teaching and, thus, the need for policymakers and researchers to account for this sort of complexity.

Given that the research base examining teachers’ contributions to student outcomes beyond test scores is relatively new, important questions remain about the validity of these measures. In particular, it is not clear whether the key identifying assumption underlying the estimation of teacher effects – that estimates are not biased by non-random sorting of students to teachers – holds when test scores are replaced with other student outcomes. Researchers who estimate value-added to students’ test scores typically control for prior achievement because it captures many of the pre-determined factors that also affect current achievement, including the schools they attend, the neighborhoods they live in, and the family members with whom they interact (Chetty, Friedman, and Rockoff 2014a; Kane et al. 2013). In a few instances (Blazar and
Kraft forthcoming; Backes and Hansen 2015; Gershenson 2016), studies also controlled for prior measures of the outcome variable when these measures were available at more than one point in time. However, it is quite possible that there are additional factors not captured either by prior test scores or by prior measures of students’ attitudes or behaviors that, in turn, could bias teacher effects on these outcomes.

In this paper, I examine this concern by drawing on a unique dataset from the National Center for Teacher Effectiveness in which participating students completed a survey that asked about a range of attitudes and behaviors in class. In the third year of the study, a subset of participating teachers were randomly assigned to class rosters within schools. Together, these data allow me to examine the extent to which teachers vary in their contribution to students’ attitudes and behaviors, even after random assignment; the sensitivity of teacher effects on students’ attitudes and behaviors to different model specifications, including those that control for students’ prior academic performance versus prior attitudes and behaviors; and, ultimately, whether non-experimental estimates of teacher effects on these attitudes and behaviors predict these same outcomes following random assignment, which produces a measure of forecast bias.

Findings indicate that teachers have causal effects on their students’ self-reported behavior in class, self-efficacy in math, and happiness in class. The magnitude of the teacher-level variation on these outcomes is similar to that on test scores. However, value-added approaches to estimating these teacher effects are insufficient to account for bias in many cases. One exception is teacher effects on students’ behavior in class, where predicted differences come close to actual differences following random assignment. Interestingly, teacher effects are not particularly sensitive to models that control from students’ prior achievement, student demographic characteristics, or prior survey responses. Given that these are the tools and data typically available to the econometrician, it likely will be necessary to continue to rely on random assignment in order to identify teachers who are effective at improving students’ attitudes and behaviors, as well as to find ways to help teachers improve in these areas.