Disruptive Peers and the Estimation of Teacher Value Added

Irina Horoi
Department of Economics
University of Illinois at Chicago

Ben Ost
Department of Economics
University of Illinois at Chicago

Abstract

Classroom disruption is often cited as an obstacle to effective teaching, yet little is known regarding how disruptive students influence classroom learning and teacher evaluation. In this study, we show that students with serious behavioral difficulties substantially reduce the academic performance of their peers. Since standard value-added models fail to account for these peer effects, we find that some teachers' value-added is penalized because of the students she is assigned. Importantly, we show that the assignment of disruptive students to teachers is non-random, so these peer effects do not impact the evaluation of all teachers equally.

Keywords: Teacher Quality, Value-Added, Disruptive Peers

Disruptive Peers and the Estimation of Teacher Value Added

I. Introduction

Understanding classroom peer effects is important both for determining optimal student grouping patterns and for generally understanding the educational production function. While classroom peer effects have been studied extensively, most research has focused on how the existence or absence of peer effects influences whether students should be tracked or placed in heterogeneous classrooms. While these considerations are first order, the existence of peer effects also implies that the educational production functions typically estimated in the literature omit an important input. To the extent that these unmeasured peer inputs are correlated with other school and classroom inputs, estimates of non-peer inputs will be biased. This point is illustrated theoretically by Lazear (2001) in the context of estimating the returns to class size, but little research has examined the issue empirically.

In this study, we consider the extent to which peer effects bias the estimated impact of other inputs by showing how disruptive peers influence the estimation of teacher value added. While teachers are just one input whose estimated impact could be biased by peer effects, the use of value-added estimates in high stakes personnel decisions makes it particularly important to correctly estimate teachers' impact.[1]

Many different forms of peer interactions have the potential to bias value-added estimation; we illustrate the issue in the context of disruptive students for several reasons. First, surveys of teachers and administrators frequently mention disruption

[1] As of 2013, 40 states require that a teacher's annual evaluation is based in part on her value added (Doherty and Jacobs 2013).

as a major obstacle to learning (Figlio 2007). Second, while researchers have controlled for average peer demographic and peer academic performance when estimating teacher value-added, we are aware of no study that controls for measures of disruption. Similarly, to the best of our knowledge, none of the value-added models currently in use to make high-stakes personnel decisions control for measures of classroom disruption. Third, while there is a large literature on classroom peer effects, most of this research has focused on how peer academic performance impacts one's academic performance, and few studies explore how the non-cognitive attributes of one's peers impact one's academic performance.

Though disruption is frequently reported as an issue by teachers and administrators, datasets typically do not include direct measures of disruption and so researchers necessarily use student characteristics that proxy for disruption (Carrell and Hoekstra 2010; Fletcher 2009[a], 2009[b]; Figlio 2007). We follow this approach by using the diagnosis of an emotional disability to proxy for disruption. In the institutional context that we study, emotional disabilities are diagnosed primarily because students exhibit disruptive behaviors in school, and we show that emotional disability correlates strongly with disciplinary action such as suspension. To the extent that some emotionally disabled students are not disruptive, our estimates will tend towards zero.[2]

[2] ED transfer students represent a relatively small fraction of all students. As such, if ED students were the only students with disruptive behavior, then the overall importance of disruption for teacher evaluation would likely be small. However, research suggests that serious class disruption is a common occurrence, particularly in urban schools (Johnston, 2013; OECD, 2013). While many students who are not diagnosed as emotionally disabled may be disruptive, the extent of the disruption might differ between ED students and other disruptive students. As such, we view our study as providing evidence that

This article expands the literature on classroom peer effects in several ways. First, we provide carefully identified evidence that peer non-cognitive attributes can influence academic achievement. Second, we use matched longitudinal data on students and teachers over a six-year period to show that the existence of these non-cognitive peer effects systematically influences the estimation of teacher value added. We show that for a variety of value-added models currently being used in policy, teaching emotionally disabled (ED) students reduces a teacher's estimated value added.

Identifying the impact of disruptive students on their peers is difficult because of the well-known issues of homophily, reflection and common shocks. Our study addresses these concerns in several ways. First, we are able to address the possibility that students are non-randomly placed into classrooms by aggregating peer groups to the school-grade-year level and including a school-by-year fixed effect. Second, we focus on transfer students who were previously diagnosed as emotionally disabled to address concerns regarding reflection and common shocks (correlated unobservables). Finally, we test for non-random sorting into grades and find that the arrival of an emotionally disabled transfer student is uncorrelated with all observable pre-determined characteristics, suggesting that homophily is unlikely to bias estimates of the peer effects we document.

Educational production functions invariably omit important inputs and we do not argue that this incompleteness necessarily leads to biased estimates of teacher quality.

---

classroom disruption has the potential to meaningfully impact teacher value added, but we cannot provide empirical evidence as to the total impact of all forms of disruption on teacher evaluation.

For example, parental and neighborhood inputs are rarely controlled for in value-added models, but since these inputs are likely to be highly correlated over time, controlling for a lagged test score or student fixed effect plausibly addresses many concerns regarding these omitted inputs.

Compared to omitting family or neighborhood characteristics, failing to control for peer effects presents a potentially more serious issue for value-added modeling for several reasons. First, since classmates change each year, peer effects will be time varying, and thus lagged test score will not control for current peer effects. Second, the majority of value-added models emphasize individual rather than peer controls, and these individual controls are unlikely to be good proxies for peer characteristics. While some researchers have controlled for average peer achievement and demographics when estimating teacher value added, few school districts collect or use data on peer quality in measuring teacher quality (Kane 2014).

If disruptive students were randomly assigned to teachers, then the peer effects we document would make the estimation of yearly teacher value added more noisy, but these estimates would remain unbiased. Conversations with principals suggest, however, that the classroom placement of disruptive students is a non-random decision, and our data bear this out. We find that within a school-grade-year, emotionally disabled students are more likely to be placed with male teachers, black teachers and more experienced teachers. While the assignment of difficult students to certain teachers may be optimal for student learning, our study suggests that the practice imposes a cost on these teachers, particularly if value added is being used for high-stakes personnel decisions.

While our study is focused on teachers, the tension we highlight between worker evaluation and task assignment is applicable to a variety of occupations. For example, financial analysts are often times rewarded for accurate forecasts, but some analysts are assigned more difficult markets than others. Similarly, universities evaluate professors based on teaching evaluations, but the material in certain courses may be more easily accessible and appealing to students. Though pay-for-performance compensation schemes are theoretically effective at eliciting optimal effort, a critical difficulty in implementation is adjusting for task assignment difficulty. In contexts where identifying task difficulty is imperfect, randomly assigning tasks to workers ensures a more fair assessment of worker productivity, but may reduce total productivity by failing to capitalize on the comparative advantage of workers when assigning tasks. Pay-for-performance schemes that fail to adjust for task difficulty create perverse incentives in which workers with a comparative advantage in difficult tasks aim to hide this information from employers.

Relative to the evaluation systems in many other occupations, value-added models include substantial adjustment for task difficulty. Teachers who are assigned low-achieving students are not penalized simply because their students perform below average at the end of the year. That said, our study demonstrates that even value-added models are unable to fully adjust for task difficulty and as such, certain teachers are systematically misevaluated. In contexts where value-added models are used for high stakes teacher evaluation, there is an important balance to strike between fair assessment and optimal task assignment.

II. Related Literature

A growing body of literature studies the impact of non-cognitive peer characteristics on cognitive outcomes (Neidell and Waldfogel 2010; Hoxby 2000; Lavy and Schlosser 2011;Fletcher 2009[a], 2009[b]; Friesen, Hickey, and Krauth 2010; Imberman, Kugler, and Sacerdote 2012; Carrell and Hoekstra 2010; Figlio 2007). Of this literature, the majority of studies have focused on disruptive students at the elementary level (Fletcher[a,b] 2009; Carrell and Hoekstra 2010) with a few studies examining disruption at the middle school level (Friesen , Hickey, and Krauth 2010; Figlio 2007).

There are several obstacles to studying the impact of disruption on peer outcomes. First, disruption is likely endogenous to teacher and peer quality and thus it is necessary to use pre-determined or exogenous determinants of disruption to instrument or proxy for disruptive students. Second, as in all peer effects studies, it is necessary to address the possibility of common shocks (correlated unobservables) and homophily (sorting).

In a series of related papers, Fletcher proxies for disruption using the diagnosis of an emotional disability and uses school fixed effects and student fixed effects to address concerns of non-random student placement. While these controls likely address part of the sorting of students to classrooms, because he has access to only a single cohort of data, he is unable to aggregate the analysis to the grade level, and therefore the time-varying systematic placement of ED students has the potential to bias estimates. Figlio (2007) likewise uses student fixed effects to address sorting into classrooms. He addresses the reflection problem by using data on males with female sounding names as an instrument for disruptive behavior, since these boys have higher propensities to act out.

As in our paper, Carrel and Hoekstra (2010) address the possibility of sorting of students to classrooms by aggregating peer groups to the school-grade-year level and use variation across grade cohorts within a school over a time. Their paper solves the reflection problem using data on parental domestic disputes as an instrument for disruptive behavior. Using cohort variation similar to that of Carrel and Hoekstra (2010), Friesen, Hickey, and Krauth (2010) explores spillovers of various disabled peers on a given student's academic achievement. Unlike the Carrel and Hoekstra article, Friesen, Hickey and Krauth are not able to address the possibility of reflection since they only observe students after they have had many years of interaction.

While there are several studies investigating the impact of student disruption, none of these studies consider how teacher value-added is impacted by peer effects and only Carrel and Hoekstra (2010) and Figlio (2007) are able to address the reflection problem.

III. Institutional Background

Serious emotional disturbance, or emotional disability, is one of the disabilities covered by the Individuals with Disabilities Education Act (IDEA), which governs how states provide interventions and services to disabled students. While students diagnosed with this disorder are a heterogeneous group, many of the behaviors typically used to diagnose an emotional disability can be directly linked to classroom disruption. For example, the screening and evaluation for emotional disability guidelines provided by the North Carolina Department of Public Instruction lists the following behaviors: "aggressive and authority challenging behaviors, overreaction to environmental stimuli, markedly diminished interest in activities, agitated, and physical manifestation of fear that have psychosomatic origin."[3]

Our data confirm the link between emotional disability and disruptive school behavior: students diagnosed with an emotional disability are 333% more likely to be suspended during 6th grade compared to other students. Relative to other proxies for disruption, emotional disability is much more strongly related to school suspension. For example, males with female sounding names are 25% more likely to be suspended (Figlio 2007) and exposure to domestic violence increases the number of disciplinary incidents by approximately 110% (Carrel and Hoekstra 2010).

IV. Data

This study uses restricted access student-teacher-matched data provided to us by the North Carolina Education Research Data Center (NCERDC). North Carolina's public school data contains rich information on students, classrooms, teachers, schools, and districts. It includes this information for all public school students in the state of North Carolina from 1995-2012. However, course membership information necessary for matching students accurately to classrooms is unavailable before 2006 and incomplete for that year. Therefore, we use data from 2007-2012 for the present analysis.

[3] This is not an exhaustive list of the behaviors that fit into one or more of the federally defined characteristics. For more examples and information see: http://ec.ncpublicschools.gov/instructional-resources/behavior-support/resources/screening-and-evaluation-for-serious-emotional-disability.

To create our estimation sample we start with student test score data from 2006-2012. We restrict it to fourth and fifth graders in the years 2007-2012, using 2006 to obtain baseline test scores. Additionally, we include only students who have taken mainstream standardized tests in math and reading and who have a baseline test score. Finally, to determine math and reading classrooms and their associated teachers, we use the 2007-2012 course membership data. These data allow us to match students to their official subject-specific classrooms and teachers. While our administrative data minimizes the extent of measurement error compared to a survey, there is likely some degree of measurement error in classroom assignments because students occasionally move between classrooms mid-year and some elementary schools use informal ability grouping that may not be reported in our data.

Table 1 shows the descriptive characteristics for the entire student sample, the sample of emotionally disabled (ED) students, and the ED transfer students that we use to identify disruption. Emotionally disabled transfer students have very different characteristics than the average student in our sample, and are somewhat lower performing than the average emotionally disabled student. For example, compared to the average student, ED transfer students are thirty-four percentage points more likely to be male, twenty-eight percentage points more likely to be African American, and thirty percentage points more likely to come from an economically disadvantaged home. ED students also perform worse academically, scoring 0.98 standard deviations below average in math and 0.87 standard deviations below average in reading.[4]

While different from the average student, ED transfer students share more similarities with the average ED student, as can be seen by comparing columns 3 and 4 to 5 and 6. For example, ED transfer students are only two percentage points more likely to be male, six percentage points more likely to be African American, perform only 0.1 standard deviations lower on their baseline math and reading tests, and tend to be placed in slightly smaller classrooms. The descriptive statistics in Table 1 suggest that ED transfer students are quite different from the average student and consequently might not be assigned to the same types of classrooms as the average student.

In order to directly examine the degree to which ED transfer students are non-randomly placed into classrooms, we examine the characteristics of classrooms with and without an ED transfer student. Comparing columns 1 and 3 in Table 2 shows that classrooms with an ED transfer student are considerably different than classrooms without an ED transfer student: about six percentage points more male, ten percentage points more African American, and eleven percentage point more economically disadvantaged.

While students in classrooms with an ED transfer student score 0.35 standard deviations lower on their math achievement test, it would be wrong to interpret this difference as the causal effect of the ED transfer student. This difference must be partly driven by student sorting since students who are in classrooms with an ED

[4] Test scores have been normalized such that grade-by-year test score have mean zero with a standard deviation of one.

transfer student also scored 0.30 standard deviations worse on their math test in the previous year.

Simple empirical models using classroom variation will fail to control for the non-random selection of students in classrooms. Furthermore, even models that include school fixed effects may be biased since students can be sorted to classrooms within schools. For this reason we use grade variation in exposure to an ED transfer student combined with school-by-year fixed effects to identify estimates. In Section VI. we test for and fail to find evidence that student sorting drives grade variation in ED transfers.

V. Empirical Approach

A. Peer Effect on Student Achievement

Identifying and estimating the impact of high-needs children on their peers is complex due to issues of correlated unobservables, reflection, and homophily (Manski 1993; Moffitt 2001). To illustrate these issues in our context, consider estimating the naïve peer effects model shown in (1).

$$(1)\ Y_{icgst} = \alpha ClassEDTS_{cst} + \omega Y_{icgs(t-1)} + X_{it}\beta + C_{cst}\delta + \varepsilon_{icgst}$$

Y_{icgst} is a subject-specific test score of individual *i* in classroom *c* in grade *g* in school *s* at time *t*, $ClassEDTS_{cst}$ is an indicator set to one if the subject specific class in school *s* at time *t* has an emotionally disabled transfer student, $Y_{icgs(t-1)}$ is the lagged subject specific test score , X_{it} is a vector of student demographic information at time

t, $\mathrm{C_{cst}}$ is a vector of classroom level characteristics of classroom *c* in school *s* at time t, and $\varepsilon_{\mathrm{icgst}}$ is an error term. The parameter of interest in equation (1) is α, which reflects the mean difference in achievement between classrooms with and without an ED transfer student, conditional on observable student and classroom characteristics. The implicit assumptions required for α to be an unbiased estimate of exposure to an ED student are:

1) Conditional on student and classroom controls, students who are exposed to an ED transfer student are comparable to those who are not in classrooms with an ED transfer student (no homophily).
2) Emotional disability status, i.e. behavior associated with this classification, and peers' academic achievement is not simultaneously determined (no reflection).
3) Conditional on student and classroom controls, the error term is orthogonal to $\mathrm{ClassEDTS_{cst}}$, i.e. there exist no correlated unoberservables with exposure to an ED transfer student and student achievement.

In our context, the three assumptions required for α to be unbiased are unlikely to hold with the specification shown in equation (1). Assumption 1 requires that students and teachers are randomly assigned to classrooms. As highlighted in Section IV, classrooms with an ED transfer student look quite different on observables both across and within schools. This evidence suggests that unless we have an instrument that creates random variation in classroom formation, classroom variation should not be used to identify the effect of exposure to an ED transfer student on student achievement. We address this endogenous peer formation by aggregating exposure to an ED transfer student to the grade rather than the classroom level. Past studies have used this same aggregation strategy to address homophily (Friesen, Hickey, and Krauth 2010; Carrel and Hoekstra 2010). In Section VI we test directly for endogeneity of peers at the grade level, and find no evidence of it.

Assumption 2 requires that a transfer student's emotional disability is not simultaneously determined with his peers' achievement. Specification (1) uses the contemporaneous measure of an ED transfer student to generate $\mathrm{ClassEDTS}_{cst}$, which will lead to spurious peer effects if low-achieving students in a class cause a transfer student to be diagnosed as emotionally disabled. To limit the possibility of reflection, we follow past work (Hoxby and Weingarth 2005; Lavy and Schlosser 2011; Neidell and Waldfogel 2010), and we identify exposure to an ED transfer student by classifying an ED transfer student as a student that was diagnosed as emotionally disabled in the previous year, before being transferred. In our context, reflection is particularly unlikely since new transfer students are unlikely to have ever had exposure to their current peers.[5]

Lastly, the simple specification in equation (1) is unlikely to satisfy assumption 3. For instance, if parents transfer their ED students to schools because of a particularly effective new principal, estimated peer effects will be biased since this new principal will also directly impact the performance of all students in the school. We address this concern by controlling for school-by-year fixed effects. Since we also control for grade-by-year fixed effects, common shocks will only bias our estimates if parents

[5] Student relocations have been previously used in the peer effect literature to resolve reflection (Imberman, Kugler, and Sacerdote, 2012).

transfer their ED students to particular schools due to school-by-grade-by-year specific factors.

In light of the above issues, our preferred model is:

$$(2)\ Y_{igst} = \alpha GradeEDTS_{gst} + \omega Y_{igs(t-1)} + X_{it}\beta + G_{gst}\delta + \theta_{st} + \lambda_{gt} + \varepsilon_{igst}$$

where $\mathrm{Y_{igst}}$ is a subject-specific test score of individual *i* in grade *g* in school *s* at time *t*, $\mathrm{GradeEDTS_{gst}}$ is an indicator set to 1 if the grade in school *s* at time *t* has an emotionally disabled transfer student, $\mathrm{Y_{igs(t\text{-}1)}}$ is the lagged subject specific test score, $\mathrm{X_{it}}$ is a vector of student demographic information at time *t*, $\mathrm{G_{gst}}$ is a vector of grade-level peer characteristics of grade *g* in school *s* at time *t*, $\mathrm{\theta_{st}}$ is a school-by-year fixed effect, $\mathrm{\lambda_{gt}}$ is a grade-by-year fixed effect, and $\mathrm{\varepsilon_{igst}}$ is an error term. Empirical specifications of education production functions with student lagged test scores on the right hand side are widely used because they are more flexible than gains models and can partly control for dynamic achievement-based sorting (Kane and Staiger 2008).

While equation (2) is our preferred specification, in the results section, we also show estimates that exclude the school fixed effects and the peer controls to help describe the robustness of the result and establish the basic patterns in the data. In all of our models we cluster the standard errors at the school-by-year-by-grade level, as this is the level of identifying variation.

B. Peer Effect on Teacher Value-Added

To evaluate how classroom disruption impacts estimated value added, we implement a two-step procedure. First, we estimate value added for every teacher in each year that they teach.[6] We allow value added to be time varying to mimic the value-added models typically used in policy. We then use these teacher-by-year value-added estimates as the dependent variable to assess whether teacher value added differs in years when a teacher teaches in a grade with an ED transfer student. We use grade, rather than classroom variation to address concerns that teachers are being sorted to particular classrooms within a grade (Rivkin et al. 2005).

Since different value-added models may yield different estimates, we consider three policy-relevant value-added models. The first model, which we refer to as the gains model, is very similar to the value-added models used in the Dallas DVASS model. In this model, we predict test score gains, adjusted for student-level covariates, to generate estimates of teacher-by-year value added. The second model is the value-added model currently in use by New York City and is very similar to the model used by the Washington, DC public schools as well. This model controls for lagged test scores, student-level covariates and basic mean peer characteristics. The third model is based on one of the models used by the *Los Angeles Times* in their release of individual-level value-added estimates for teachers in the LAUSD. This model includes peer characteristics and student fixed effects in estimating teacher-by-year value added.

The three value-added models that we estimate are shown explicitly in equations (3a)-(3c):

$$(3a)\ \Delta Y_{ijcgst} = X_{it}\beta + \gamma_t + \mu_{jt} + \varepsilon_{icjgst}$$

[6] Since many school districts only evaluate regular classroom teachers based on value-added, we drop special education teachers from all analyses of teacher-value added. Analyses that include special education teachers yield slightly larger point estimates and similar standard errors.

$$(3b)\ Y_{ijcgst} = \omega Y_{ijcgs(t-1)} + X_{it}\beta + C_{jcst}\delta + \gamma_t + \theta_g + \mu_{jt} + \varepsilon_{icjgst}$$

$$(3c)\ \Delta Y_{ijcgst} = X_{it}\beta + C_{jcst}\delta + \mu_{jt} + \pi_i + \varepsilon_{icjgst}$$

Y_{ijcgst} is subject-specific test score of individual *i* matched to subject-specific teacher *j* in classroom *c* in grade *g* in school *s* at time *t,* $Y_{ijcgs(t-1)}$ is lagged subject-specific test score of individual *i* matched to subject specific teacher *j* in classroom *c* in grade *g* in school *s,* X_{it} is a vector of student demographic information at time *t,* C_{jcst} is a vector of classroom-level characteristics of classroom *c* with teacher *j* in school *s* at time *t*, γ_t are year dummies, θ_g are grade dummies, μ_{jt} are teacher-by-year fixed effects, π_i is a student fixed effect and ε_{icjgst} is an error term. [7]

Using the teacher-by-year value-added estimates as the dependent variable, we examine the impact of exposure to an ED transfer student. Analogous to equation (2), we estimate (4):

$$(4)\ \mu_{jst} = \alpha TeachGradeEDTS_{jst} + \theta_{js} + \varepsilon_{jst}$$

where μ_{jst} is the estimated subject-specific teacher effectiveness for teacher *j* in school *s* at time *t,* $TeachGradeEDTS_{jst}$ is an indicator equal to 1 if the teacher *j* in school *s* at time *t* teaches in a grade with an ED transfer student. Depending on the specification, we also include a school fixed effect, a school-by-year fixed effect or a

[7] Because classroom composition is perfectly collinear within a teacher-year, we estimate the classroom composition effect, , using a three-step procedure as described by Isenberg and Walsh (2013) and used by NYC and DC. First we run a specification similar to (3), except that the teacher-by-year fixed effects are replaced with teacher-by-school fixed effects. This allows us to compare multiple classrooms for a teacher over time, which breaks the perfect colinearity between teacher and classroom composition. In step two, we use the estimated impact of classroom composition to calculate an adjusted subject-specific test score that nets out the contribution of the classroom characteristics. In the final step, we use the adjusted subject-specific scores in place of the actual test scores, and estimate (3), omitting classroom variables from the specification.

teacher-by-school fixed effect. All standard errors are clustered at the school-by-year-by-grade level.

VI. Results

A. Peer Effect on Student Achievement

Table 3 reports the effects of grade exposure to an ED transfer student on the math and reading achievement of other students. The first four columns show the effect for math achievement, and the last four columns report the estimates for reading achievement. Column 1 reports OLS estimates controlling for student-level demographics, lagged student achievement in math, and grade-by-year dummies.[8] In columns 2 and 6, school fixed effects are added, in columns 3 and 7, school-by-year fixed effects are included in place of school fixed effects, and in columns 4 and 8, grade-level peer characteristics are added.[9]

The results shown in Table 3 provide evidence that emotionally disabled students impact the math scores of their peers. Estimates for math drop by 45 percent when school fixed effects are added (column 2) but remain fairly stable (and statistically indistinguishable) when subsequently adding school-by-year and peer characteristics. The relative stability of the estimates in columns 2 through 4 is reassuring since it suggests that ED students do not systematically transfer to grades with different student characteristics. The fact that the estimated peer effect falls considerably when

[8] The student-level demographics include a male gender dummy, race dummies (African American, Hispanic, and White), a dummy for limited English proficiency, and a dummy for being economically disadvantaged.

[9] The grade-level peer controls include proportion male, percent African American, percent Hispanic, percent White, percent limited English proficiency, percent economically disadvantaged, and average pretest.

adding school fixed effects suggests that ED students tend to transfer to schools with lower-achieving students.

For reading, the estimates drop by more than half when school fixed effects are included (column 6), and stay statistically the same in magnitude, but become insignificant across more controlled specifications. In our preferred specification, shown in columns 4 and 8, a single emotionally disabled student causes the average performance of other students in the grade to be reduced by 0.017 standard deviations in math and 0.006 standard deviations in reading.[10]

To put our estimates in perspective, it is worth converting our estimates to approximate days of learning, as in Reardon (2011). Based on his estimates, average learning is approximately 0.3 standard deviations per year, and thus an effect of 0.016 corresponds to approximately a 5 percent reduction in school-year equivalents, or about two weeks less learning.

These findings are consistent with the literature on academic externalities associated with disruptive peers (Carrel and Hoekstra 2010; Figlio 2007; Friesen, Hickey, and Krauth 2010; Fletcher[a,b] 2009). That said, in comparing our estimates to some papers in the literature (e.g. Fletcher[b] 2009), it is important to keep in mind that we estimate our effects at the grade, rather than classroom level. [11]

[10] Finding a smaller impact on reading scores is a consistent finding in the literature. The most likely explanation is that most math learning occurs at school, whereas reading is more likely to be learned both at school and at home.

[11] Estimating peer effects at the grade level is similar to using grade-level variation to instrument for having an ED transfer student in one's classroom. In fact, our preferred specification is simply the reduced form from that IV specification. To estimate the IV specification, one simply scales up our estimate of 0.016 based on the inverse probability that a student is in classroom with an ED transfer student, given that they are in a grade with an ED transfer student. We opt not to estimate the peer effects using the IV

We view disruption as the most likely channel through which emotionally disabled students impact their peers, but we cannot rule out the possibility that some characteristic that is correlated with emotional disability actually drives our estimates. In results not shown, we find that very low-achieving transfer students (at least 1 standard deviation below average) who are not emotionally disabled have a negative, but much less severe, impact on the academic performance of their peers. Since low-achieving transfer students have lower test scores than the ED transfer students we consider, this provides suggestive evidence that it is not just the poor academic performance of ED students that hurts their peers' performance. In any case, for the analysis of teacher value added, it is unimportant whether classroom disruption or some other channel drives the peer effects, so long as a peer effect exists.

B. Specification Tests

To test for whether emotionally disabled transfer students endogenously enter particular grades in a school, we examine whether predetermined student characteristics predict whether or not one's school-grade-year has an ED transfer student. Specifically, we regress an indicator for whether or not the student is in a grade with an ED transfer student on student characteristics, school-by-year fixed effects and grade-by-year fixed effects. While our preferred estimates will only be biased if ED students sort to particular grades, we also examine whether ED students

specification because this specification assumes that students have no interaction with students in their grade outside of their classroom, an assumption we find implausible. This is particularly implausible at the elementary level where students often mix across self-contained classrooms for special math or reading classes. In our data, we have no information regarding this sort of informal classroom assignment. That said, when the IV specification is used, the estimates are on par with the literature and are quite close to Fletcher[b] (2009).

sort into particular classes within a school using the same approach. The idea behind these tests is to see whether ED transfer students enter grades or classes that were likely to perform poorly in any case.

These results are presented in Table 4. Column 1 shows results for the classroom-level regression, and column 2 shows the results for the grade-level specification. We find that a number of predetermined student characteristics such as gender, Hispanic ethnicity, previous emotional disability, and pretest scores are predictive of the probability of being assigned a classroom with an ED transfer student within a school and year. For example, a student classified with an emotional disability in a previous year is 7.5 percentage points more likely to get assigned to a class with an ED transfer student. These results suggest that ED transfer students are systematically placed into certain classrooms and thus, across-class variation cannot be used to identify the impact of peer effects.

Column 2 of Table 4 shows that when we use grade variation within a school and year, we find that none of the student characteristics predict assignment to grades with an ED transfer student. This suggests that conditional on school-by-year fixed effects, ED transfer students are not systematically placed into grades with certain types of students.[12]

C. Teacher Value-Added

[12] Since not all students take both math and reading classes, there are slightly different samples for the math and the reading analyses. The specification tests shown in Table 4 use the sample of students in math classes but the results are qualitatively and quantitatively the same for the sample of students in reading classes. These results are available upon request from the authors.

Table 5 presents the results from estimating equation (4): the effect of having an ED student in the grade on teacher value added. We measure ED exposure at the grade, rather than classroom level, since Table 4 shows that ED students are systematically sorted towards certain classes. In these specifications, we restrict the analysis to regular education teachers to ensure that special education teachers do not drive the results. This focus makes sense from a policy perspective since most special education teachers are not evaluated based on value added.[13]

Columns 1 through 3 show that regardless of the choice of value-added model, teachers are evaluated as less effective when they are in a grade with an ED student. Based on these results alone, however, it would be wrong to conclude that the ED students are causing the reduction in value added, because it is possible that ineffective teachers are more likely to work in schools that have more ED transfer students. To address this possibility, in columns 4 and 5 of Table 5, we use the same dependent variable as in column 2, but we add school or school-by-year fixed effects to the model relating ED students to teacher value-added. These models use within-school variation in exposure to ED students to test whether teachers in grades with an ED student are evaluated as worse than other teachers in the same school.

Though student sorting to teachers within a grade cannot bias our estimates (since we aggregate to the school-grade-year level) it remains possible that lower-quality teachers are placed into grades that will have ED transfer students. Though we view this scenario as unlikely, we investigate the possibility by adding teacher-by-school fixed effects to the model that predicts teacher value-added. Essentially, this

[13] All estimates are nearly identical when we include special education teachers in the analysis.

specification uses across-time variation in exposure to ED transfer students to compare teachers to themselves. Column 6 of Table 5 shows that teachers are evaluated as having lower value added (relative to themselves) in years that they are in a grade with an ED student. Past research has found that there exists substantial within-teacher year-to-year variation in value-added estimates (McCaffrey et al. 2009). The result shown in column 6 suggests that a portion of this variation in teacher value-added is attributable to peer effects.

In interpreting the magnitudes of the estimates shown in Table 5, it is important to keep in mind that the dependent variable is teacher value added – not student test scores. Our estimates imply that teacher value added for math is approximately 0.02 *student* standard deviations lower because of the ED student. Since the standard deviation of teacher value added is approximately one-fifth of the standard deviation of student test scores, our effect size corresponds to approximately one-tenth of a standard deviation decrease in teacher value added. The magnitude of our estimate suggests that few teachers will be grossly misevaluated as a result of the peer effects that we study, but the estimate is large enough to be of substantive significance. Also, we are identifying the impact of just one type of peer effect, so it remains possible that the overall importance of peers in the estimation of value added is substantial.

D. Systematic placement of ED students

If ED students were randomly assigned to teachers, peer effects might cause biased assessments of teacher quality in particular years, but over the long run no teacher would by systematically penalized. For a principal interested in maximizing the learning of her students, however, it makes little sense to randomly assign ED students to teachers since certain teachers may be better equipped to handle these students. If some teachers are repeatedly assigned more difficult students and value-added estimates fail to account for disruption, certain teachers will have lower estimated value-added, even when measured over many years. Importantly, the value-added measure for these teachers would accurately reflect the amount of learning that occurs in their classrooms, but it would not accurately reflect the teachers' skill or effort.

To investigate whether certain teachers are systematically assigned ED students, we use two distinct approaches. First, we investigate whether the observed distribution of assignments could be random based on the mean frequency with which each teacher is assigned an ED student. Second, we investigate whether measurable teacher and school characteristics are predictive of being assigned to an ED student. As in the previous section, we restrict the analysis to regular education teachers to ensure that results are not driven by special education teachers.[14] [15]

To test whether the distribution of ED student assignments are plausibly random, we regress an indicator for ED student assignment on a set of individual teacher fixed effects. If the assignment of ED students were random, the teacher fixed effects would be jointly indistinguishable from zero, but an F-test strongly rejects this possibility (p-value<0.01). One potential concern with this test is that it is so powerful that differences in assignment caused by chance will lead the F-test to reject

[14] We find nearly identical estimates if we include special education teachers in these analyses.
[15] Since our analysis of sorting is not subject to concerns regarding reflection or common shocks, we include all ED students in these analyses.

the null. While technically, the F-test should account for this possibility, we explore this issue by simulating random assignment of ED students to teachers.[16]

Though the F-test rules out the possibility that ED students are randomly assigned to teachers, it does not provide information regarding which types of teachers are more or less likely to be assigned ED students. To explore this question, we investigate two conceptually different forms of sorting. First, ED students might sort towards certain types of schools, making teachers at these schools more likely to teach ED students than teachers at other schools. Second, within a school-year-grade that has an ED student, certain teachers may be more likely to be assigned that ED student. While across-school sorting cannot bias value-added models that control for school fixed effects, in practice, most school districts aim to compare teachers across schools and thus both within and across sorting is relevant for understanding whether the evaluation of certain teachers will be systematically impacted by ED students.

To investigate sorting across schools, we regress an indicator for whether an ED student is present in a particular school on a vector of school-level characteristics. To investigate within-school sorting, we restrict the analysis to school-year-grades with an ED student and regress an indicator for whether a teacher was assigned to teach that ED student on characteristics of that teacher and school-by-year-by-grade fixed effects. In interpreting the results of these regressions, we are focused on correlations

[16] We implement this simulation by starting with our analysis dataset but ignoring true ED assignment. For each ED student observed in the data, we randomly assign that student to a teacher in their school and grade. We then simulate random assignment 1,000 times and for each simulation, we calculate the F-statistic testing random assignment. The F-statistic calculated from the actual data is an extreme outlier in the distribution of F-statistics from the simulation: of the thousand simulated F-statistics, 100% are smaller than the F-statistic found in the actual data.

rather than a causal interpretation since we are simply describing how a variety of factors correlate with ED assignment.

Column 1 of Table 6 shows the results from a regression of whether a school-year has an ED student on various school characteristics. Schools with smaller class sizes and schools with a higher percentage of black students are slightly more likely to have ED students, but there is little evidence that schools serving poorer populations are differentially likely to have ED students. ED students are substantially more likely to be found at charter schools compared to traditional public schools and are less likely to be at small schools compared to large schools. An important caveat in interpreting the results shown in column 1 is the possibility that ED diagnosis, but not necessarily ED behaviors, differ across schools.

Column 2 of Table 6 adds school fixed effects to examine how across-time variation in school characteristics related to ED student enrollment. Some of the descriptive patterns shown in column 1 do not appear to hold across time within a school, suggesting that class size and student demographics are likely not causally related to the placement of ED students into schools. On the whole, we find little evidence that time-varying school characteristics meaningfully impact the probability of having an ED student enroll.

Conditional on enrolling in a particular school, ED students may be systematically assigned to certain teachers within that school. We investigate this possibility by regressing whether a teacher is assigned an ED student on observable characteristics and school-by-year-by-grade fixed effects. Unlike the analysis of school sorting, the results for our study of within-school sorting cannot be driven by differences in diagnosis across schools.

We find that teachers in their first year of teaching are much less likely to be assigned ED students, but there is little relationship between experience and being assigned an ED student beyond the first year of teaching.[17] Male teachers are nearly 6 percentage points more likely to be assigned ED students and black teachers are approximately 3 percentage points more likely to be assigned ED students. We find no evidence that ED student assignment within a school-grade-year is systematically related to whether a teacher has an advanced degree or a teacher's estimated value-added.[18]

Columns 4-6 of Table 6 demonstrate the basic robustness of these results. First, we replace the school-by-grade-by-year fixed effects with school-by-year fixed effects. Though we prefer the specification with school-by-grade-by-year fixed effects, the fact that the results are very similar regardless of which fixed effect is included supports the general robustness of the result, and the school-by-year fixed effects specification provides slightly more precise estimates. In columns 5 and 6 we investigate whether there is a relationship between experience and ED assignment for teachers with more than 5 years of experience, and find little evidence that there is.

[17] Feng (2010) finds that more experienced teachers are less likely to be assigned special education students in general, but does not specifically consider ED students. Feng's findings imply that early career teachers may be penalized if special education students impose negative peer effects on their classmates.

[18] We calculate value added for each teacher using only years in which they are not teaching an ED student to avoid building a mechanical relationship between ED assignment and the portion of value added attributable to the ED student.

VII. Conclusion

The landscape for high-stakes teacher evaluation policies has changed dramatically over the last five years. Since 2009, 25 states and the District of Columbia have adopted policies that require teacher evaluation to include objective measures of student achievement. More strikingly, the number of states that require student growth to be the major criterion in teacher evaluation increased by 500%, going from 4 to 20 states including D.C. in 2013 (Doherty and Jacobs 2013). As evaluations of teachers continue to rely more heavily on teacher value-added estimates, it is important that policy makers are aware of the limitations and strengths of these estimates.

Given the difficulty of credibly identifying the impact of peer effects, we do not attempt to give a full characterization of how teacher value added is impacted by all types of peer effects. Instead, we show that a particular type of peer effect – namely the impact of emotionally disabled students – moderately biases the evaluation of teacher value added. While we only provide empirical evidence for this one peer effect, it is likely that other forms of peer effects also influence the estimation of teacher value added, such that the total bias caused by peer interactions could be quite large.

In a recent influential paper, Chetty, Friedman and Rockoff (2014) provide evidence that value-added models yield approximately unbiased estimates of teacher quality and that these value-added estimates correlate with long-run student outcomes. Though these results have recently been challenged (Rothstein 2014), a correlation between long-run outcomes and value added is consistent with our results for several reasons. First, our results imply that value-added estimates will be only modestly biased – well within the standard error of the Chetty et al. estimates. Second, Chetty et al. (2014) find relatively weak correlation across years within a teacher, allowing for the possibility that year-to-year variation is partly driven by factors such as changes in peer composition not captured by their controls. Finally, the correlation between long-run outcomes and teacher value-added is completely consistent with the notion that students learn less when their peers are disruptive. The smaller learning gains made by these students could plausibly impact long-run outcomes, and as we show, teacher value added is reduced as well. In the presence of important peer effects, basic value-added models may still correctly identify student learning, but they do not necessarily identify teacher quality.

As discussed in Lavy, Paserman, and Schlosser (2011), students may impact their peers through a variety of channels. While beyond the scope of the current paper, it would be interesting for future work to disentangle whether the peer effects we estimate are attributable to direct classroom disruption or to teachers altering their instruction (or time allocation) as a result of the ED student.

While our study demonstrates one limitation of value-added estimates, it is important to note that we provide little evidence on the question of whether school districts should use value added for high-stakes teacher evaluation. First, it is very possible that observation-based evaluations are also subject to bias from peer effects. Though observers aim to evaluate teacher quality, observer perception of quality may be influenced by classroom composition (Whitehurst, Chingos and Lindquist 2014). Second, the magnitude of the bias we document is sufficiently modest so that the cost of unfairly evaluating some teachers may be outweighed by other benefits of value-added evaluation. Finally, regardless of any limitations in the estimation of teacher value added, policies that evaluate teachers based on value-added may induce effort that improves student achievement. (Dee and Wyckoff 2013)

As school districts increasingly rely on value-added models for high stakes personnel decisions, principals should be aware that these models do not fully adjust for classroom composition. Teachers that are consistently given difficult classrooms may be evaluated to be less effective than teachers given less difficult students, even if their true quality is equivalent. Random assignment of teachers to difficult students would avoid penalizing any particular teacher, but may lower student achievement by reducing the match quality between teachers and students.

References

Carrell, Scott and Mark L. Hoekstra, 2010. “Externalities in the Classroom: How Children Exposed to Domestic Violence Affect Everyone’s Kids,” *American Economic Journal: Applied Economics*, 2(1): 211-228.

Raj Chetty & John N. Friedman & Jonah E. Rockoff, 2014. "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates," *American Economic Review*, 104(9): 2593-2632.

Dee, Thomas, and James Wyckoff. 2013. "Incentives, Selection, and Teacher Performance: Evidence from IMPACT." NBER Working Paper WP19529. Cambridge, MA: National Bureau of Economic Research (October).

Doherty, Kathryn M. and Sandi Jacobs, The National Council on Teacher Quality. 2013. State of States 2013 Connect the Dots: Using Evaluations of Teacher Effectiveness to Inform Policy and Practice (Research Report). Retrieved from http://www.nctq.org/dmsView/State_of_the_States_2013_Using_Teacher_Evaluations_NCTQ_Report

Feng, L. (2010). Hire today, gone tomorrow: New teacher classroom assignments and teacher mobility. *Education Finance and Policy*, 5(3), 278-316.

Figlio, David N. 2007. “Boys Named Sue: Disruptive Children and their Peers.” *Education Finance and Policy,* 2(4): 376-394.

Fletcher, Jason M. 2009[a]. “The Effects of Inclusion on Classmates of Students with Special Needs: The Case of Serious Emotional Problems.” *Education Finance and Policy,* 4(3): 278-299.

Fletcher, Jason M. 2009[b]. “Spillover Effects of Inclusion of Classmates with Emotional Problems on Test Scores in Early Elementary School.” *Journal of Policy Analysis and Management,* 29(1): 69-83.

Friesen, Jane, Ross Hickey, and Brian Krauth. 2010. “Disabled Peers and Academic Achievement.” *Education Finance and Policy,* 5(3): 317-348.

Hoxby, Caroline M. 2000. “Peer Effects in the Classroom: Learning from Gender and Race Variation.” NBER Working Paper No. 7867.

Hoxby, Caroline M. and Gretchen Weingarth. 2005. “Taking Race Out of the Equation: School Reassignment and the Structure of Peer Effects.” Mimeo, Harvard University.

Imberman, Scott A., Adriana D. Kugler, and Bruce I. Sacerdote. 2012. "Katrina's Children: Evidence on the Structure of Peer Effects from Hurricane Evacuees." *American Economic Review*, 102(5): 2048-2082.

Isenberg, Eric and Elias Walsh. 2013. Measuring Teacher Value Added in DC, 2012-2013 School Year (Report No. 06038.502).Washington, DC: Mathematica Policy Research.

Johnston, Howard (2013). *Disruptive Behavior: School Based Interventions*. Retrieved from Education Partnerships Research into Practice: http://www.gearup.oous.edu/sites/default/files/research-briefs/researchbriefdisruptivebehavior.pdf

Thomas J. Kane. "Do Value-Added Estimates Identify Causal Effects of Teachers and Schools?" The Brown Center Chalkboard, Brookings Institution, October 30, 2014.

Kane, Thomas J., and Douglas O. Staiger. 2008. "Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation." National Bureau of Economic Research." NBER Working Papers 14607.

Lavy, Victor, and Analia Schlosser. 2011. "Mechanisms and Impacts of Gender Peer Effects at School." *American Economic Journal: Applied Economics*, 3(2): 1-33.

Lavy, Victor, Daniele M. Paserman, and Analia Schlosser. 2011. "Inside the Black Box of Ability Peer Effects: Evidence from Variation in the Proportion of Low Achievers in the Classroom." *The Economic Journal,* 122(559): 208-237.

Lazear, Edward. 2001. "Education Production," *Quarterly Journal of Economics*, 116(3): 777-803.

Manski, Charles. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *Review of Economic Studies*, 60(3): 531-542.

McCaffrey, D. F., Sass, T. R., Lockwood, J. R., & Mihaly, K. (2009). The intertemporal variability of teacher effect estimates. *Education Finance and Policy*, 4(4), 572-606.

Moffitt, Robert. "Policy Interventions, Low-Level Equilibrium, and Social Interactions," in S. Durlauf and P. Young (Eds.), *Social Dynamics*. Cambridge, MA: MIT Press, 2001.

Neidell Matthew, and Jane Waldfogel. 2010. "Cognitive and Noncognitive Peer Effects in Early Education." *The Review of Economics and Statistics,* 92(3): 562-576.

OECD, 2013 *How the quality of the learning environment is shaped.* Retrieved from http://www.oecd.org/pisa/keyfindings/Vol4Ch5.pdf

Rivkin, Steven, Eric Hanushek and John Kain (2005) "Teachers, Schools and Academic Achievement" Econometrica Vol. 73, No. 2, pp. 417-458.

Rothstein, Jesse (2014) “Revisiting the Impacts of Teachers.” *Working paper*. Retrieved from: http://eml.berkeley.edu/~jrothst/workingpapers/rothstein_cfr_oct2014.pdf

Whitehurst, Grover J., Matthew M. Chingos and Katharine M. Lindquist (2014) “Evaluating Teachers with Classroom Observations: Lessons Learned in Four Districts” Brown Center on Education Policy at Brookings.

Table 1: Summary Statistics

	All Students		*Emotionally Disabled Students*		*Emotionally Disabled Transfer Students*	
	mean	sd	mean	sd	mean	sd
Male	0.50	0.50	0.82	0.39	0.84	0.37
African American	0.26	0.44	0.48	0.50	0.54	0.5
Hispanic	0.12	0.32	0.02	0.15	0.03	0.16
White	0.54	0.50	0.43	0.49	0.35	0.48
Limited english proficiency	0.07	0.25	0.01	0.11	0.02	0.13
Economically disadvantaged	0.51	0.50	0.76	0.43	0.81	0.39
Suspended in 6th grade	0.12	0.32	0.52	0.5	0.54	0.5
Reading score	0.00	1.00	-0.76	1.02	-0.87	0.97
Math score	0.00	1.00	-0.83	0.99	-0.98	0.93
Reading pretest	0.04	0.97	-0.70	1.00	-0.8	0.96
Math pretest	0.04	0.97	-0.75	0.95	-0.85	0.89
Teacher experience	11.34	9.16	11.34	9.00	10.86	8.83
Observations	1311480		3902		1128	

Table 2: Summary Statistics of Classrooms With and Without an Emotionally Disabled Transfer Student

	Classroom without EDTS		*Classroom with EDTS*	
	mean	sd	mean	sd
Proportion male	0.51	0.14	0.57	0.18
Proportion African American	0.28	0.27	0.40	0.31
Proportion Hispanic	0.12	0.15	0.11	0.13
Proportion White	0.53	0.31	0.42	0.32
Proportion limited englich proficiency	0.07	0.12	0.07	0.10
Proportion economically disadvantaged	0.54	0.28	0.66	0.27
Teacher experience	11.59	9.20	10.92	8.69
Class size	22.65	11.48	20.39	12.77
Math pretest score	-0.02	0.61	-0.32	0.63
Reading pretest score	-0.02	0.59	-0.29	0.62
Math score	-0.07	0.64	-0.42	0.68
Reading score	-0.07	0.61	-0.39	0.66
Observations	59359		677	

Table 3: Estimates of Exposure to an Emotionally Disabled Transfer Student on Math and Reading Test Scores

	(1) Math	(2) Math	(3) Math	(4) Math	(5) Reading	(6) Reading	(7) Reading	(8) Reading
Grade has an EDTS	-0.0344***	-0.0191***	-0.0156**	-0.0168***	-0.0241***	-0.0097**	-0.0061	-0.0063
	(0.0064)	(0.0056)	(0.0062)	(0.0061)	(0.0044)	(0.0039)	(0.0039)	(0.0039)
Student controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grade controls	No	No	No	Yes	No	No	No	Yes
School FE	No	Yes	No	No	No	Yes	No	No
School-by-year FE	No	No	Yes	Yes	No	No	Yes	Yes
Observations	1164356	1164356	1164356	1164356	1156750	1156750	1156750	1156750

Notes: All models include student controls (male, African American, Hispanic, White, limited English proficiency, economic disadvantage, and the subject specific pretest score), and grade-by-year dummies. Grade controls refers to percent male, African American, Hispanic, White, limited English proficiency, economically disadvantaged, and the average subject specific pretest score. Standard errors clustered at the school-by-year-by-grade level are shown in parentheses. (* $p<0.10$ ** $p<0.05$ *** $p<0.01$)

Table 4: Predicting Probablility of Exposure to an Emotionally Disabled Tranfer Student

	(1) *Estimated at Class*	(2) *Estimated at Grade*
Male	0.0006***	-0.0001
	(0.0001)	(0.0002)
African American	-0.0002	-0.0006
	(0.0004)	(0.0005)
Hispanic	-0.0010**	0.0004
	(0.0005)	(0.0006)
White	-0.0002	-0.0003
	(0.0004)	(0.0004)
Previouse emotional disability	0.0753***	-0.0035
	(0.0073)	(0.0030)
Math pretest	-0.0009***	0.0000
	(0.0002)	(0.0003)
Reading pretest	-0.0004**	-0.0001
	(0.0002)	(0.0002)
Grade has an emotionally disabled transfer student	0.1677***	-
	(0.0044)	
School-by-year FE	Yes	Yes
Observations	1154589	1154589

Notes: All models include grade-by-year dummies. The samples refer to students in math classes. Standard errors clustered at the school-by-year-by-grade level are shown in parentheses. (* p<0.10 ** p<0.05 *** p<0.01)

Table 5: Relation Between Various Teacher Quality Estimates and Whether or Not the Teacher Teaches in a Grade with an Emotionally Disabled Student

	(1)	(2)	(3)	(4)	(5)	(6)
Model used to caluclate value added:	Gains	Student fixed-effect	Lagged achievement			
Panel A. Math Teachers						
Teaches math in grade with an emotionally disabled student	-0.0241***	-0.0256**	-0.0282***	-0.0161***	-0.0192***	-0.0189***
	(0.0072)	(0.0253)	(0.0069)	(0.0062)	(0.0066)	(0.0056)
School FE	No	No	No	No	Yes	No
School-by-year FE	No	No	No	Yes	No	No
Teacher-by-school FE	No	No	No	No	No	Yes
Observations	45867	45363	45363	45363	45363	45363
Panel B. Reading Teachers						
Teaches reading in grade with an emotionally disabled student	-0.0137***	-0.0357***	-0.0198***	-0.0089**	-0.0128***	-0.0153***
	(0.0050)	(0.0185)	(0.0047)	(0.0044)	(0.0044)	(0.0043)
School FE	No	No	No	No	Yes	No
School-by-year FE	No	No	No	Yes	No	No
Teacher-by-school FE	No	No	No	No	No	Yes
Observations	49909	49345	49345	49345	49345	49345

Notes: Panel A. shows results from regressions of math teacher value-added on a dummy for whether a teacher teaches math in a grade with an emotionally disbaled student. Panel B. estimates the same specifications but for reading teachers and their value-added. In column 1, the dependent variable is value-added calculated using the gains model. In column 2, the dependent variable is value-added calculated using the student fixed effects model and in columns 3-6, the dependent variable is value-added calculated using the lagged model. (see text for details on the calculation of value added in each of these models) Standard errors clustered at the school-by-year-by-grade level are shown in parentheses (* p<0.10 ** p<0.05 *** p<0.01) .

Table 6: Sorting of Emotionally Disabled Students Across and Within Schools.

	(1)	(2)		(3)	(4)	(5)	(6)
	School has ED Student			Teacher has Emotionally Disabled Student			
Student teacher ratio	-0.006	0.0036	Teacher has 0 years	-0.0963***	-0.0992***	-0.1154***	-0.1215***
	(0.0037)	(0.0032)	of teaching exper.	(0.0200)	(0.0191)	(0.0235)	(0.0226)
% on free or reduced lunch	-0.0001	-0.0007	Teacher has 1 years	-0.0088	-0.0125	-0.0267	-0.0333
	(0.0004)	(0.0008)	of teaching exper.	(0.0217)	(0.0207)	(0.0243)	(0.0233)
Charter	0.2200***	-	Teacher has 2 years	-0.021	-0.0286	-0.0369	-0.0472**
	(0.0299)		of teaching exper.	(0.0219)	(0.0209)	(0.0242)	(0.0232)
Magnet program	0.0444	-	Teacher has 3 years	-0.0159	-0.0177	-0.031	-0.0352
	(0.0328)		of teaching exper.	(0.0243)	(0.0230)	(0.0265)	(0.0252)
% Black	0.0019***	0.0016	Teacher has 4 years	-0.0045	-0.0048	-0.0183	-0.0209
	(0.0003)	(0.0028)	of teaching exper.	(0.0236)	(0.0226)	(0.0251)	(0.0241)
% Hispanic	0.0005	-0.0047	Teacher has 5 years	-0.0133	-0.0158	-0.0261	-0.0307
	(0.0007)	(0.0037)	of teaching exper.	(0.0235)	(0.0224)	(0.0250)	(0.0238)
Small school	-0.1544***	-	Male teacher	0.0595***	0.0596***	0.0574***	0.0572***
	(0.0203)			(0.0192)	(0.0184)	(0.0192)	(0.0184)
Medium school	-0.0455**	-	Asian teacher	0.0026	0.0178	-0.0005	0.0152
	(0.0206)			(0.0915)	(0.0833)	(0.0916)	(0.0836)
			Black teacher	0.0308*	0.0329*	0.0321*	0.0345**
				(0.0179)	(0.0169)	(0.0180)	(0.0170)
			Hispanic teacher	0.0303	0.0333	0.0283	0.0314
				(0.0887)	(0.0852)	(0.0882)	(0.0847)
			Masters or higher	-0.0147	-0.0169	-0.0125	-0.0143
				(0.0124)	(0.0120)	(0.0126)	(0.0121)
			Teacher value-added	0.0173	0.0205	0.0171	0.0201
				(0.0306)	(0.0279)	(0.0306)	(0.0278)
			Experience (linear)	-	-	-0.0013	-0.0016*
						(0.0009)	(0.0008)
School fe	No	Yes		No	No	No	No
School-by-year fe	No	No		No	Yes	No	Yes
School-by-grade-by-year fe	No	No		Yes	No	Yes	No
Year fe	Yes	Yes		No	No	No	No
Observations	4855	4855		8238	8238	8238	8238

Notes: In columns 1-2 the dependent variable is whether a school in a given year has enrolled an emotionally disabled student. In columns 3-6 the dependent variable is whether a teacher in a given year teaches an emotionally disabled student. Standard errors clustered at the school-by-year-by-grade level are shown in parentheses. (* p<0.10 ** p<0.05 *** p<0.001)