DISCUSSION PAPERS IN ECONOMICS

Wo king Pa e No. 23-04

! "#$%&'()%#*(+,-."/#011&(1#0 2 3,4%#5(,4+(/#6(%()%.")7#
8-.'()4(#9/"2#:'2.).1%/,%.-(#',%,#")#$%&'()%#;99()1(1#

! "#$%&'#()*+)

Do Student Behavior Issues Impact Teacher Retention? Evidence from Administrative Data on Student Offenses*

Alexander H. Bentz

University of Colorado, Boulder
Economics Building
256 UCB
Boulder, Colorado 80309-0256
alexander.bentz@colorado.edu

October 30, 2023

Most recent version (link)

Abstract

This paper provides evidence that student behavior issues contribute to teacher turnover among U.S. middle school teachers. Using detailed administrative data on student behavior, discipline,

1 Introduction

Schools' failure to retain effective teachers is common and costly. Prior to the COVID-19 pandemic in 2020, about 16% of K–12 teachers in the U.S. left their school in a given year.[1] At high-poverty schools and among first-year (novice) teachers, more than 20% left each year, an issue that is more acute in some places since 2020 (Bruno, 2022). The cost of replacing teachers is high, both in terms of direct costs (e.g., recruitment, hiring, and training) (Barnes, Crowe, & Schaefer, 2007) and indirect costs (e.g., lower student achievement when losing high-quality teachers) (Chetty, Friedman, & Rockoff, 2014). Further, teacher attrition–leaving the teaching profession altogether–is particularly costly for school systems, as teaching effectiveness grows with experience (Wiswall, 2013).

Extensive survey evidence suggests that poor school climate and administrative support are major reasons why teachers leave a school (Carver-Thomas & Darling-Hammond, 2019; Ingersoll, 2001; Nguyen, Pham, Crouch, et al., 2020). However, the exact factors that contribute to school climate are less clear. One under-studied factor suggested by teacher surveys is student behavior: teachers who report behavioral problems at their school are more likely to leave teaching (Kukla-Acevedo, 2009; Nguyen, Pham, Crouch, et al., 2020). Behavior issues may be particularly problematic for novice teachers, who typically receive minimal classroom management training before becoming teachers. While teachers can improve their classroom management abilities, novice teachers who struggle with classroom management are more likely to leave teaching early in their career (Bartanen, Bell, James, et al., 2023).

However, existing evidence linking student behavior to teacher attrition or mobility (moving to another school) may be biased by using unreliable measures of student behavior. Teachers' perceptions, while important in individual decision-making, may not objectively reflect differences across classrooms or schools. Meanwhile, administrative data on reported offenses in schools, collected by many states, are an incomplete record of student behavior. These data include offenses reported to principals and added to the state database, but reporting policies vary by state and school. In North Carolina, the setting for this study, many offenses–such as disruptive behavior, dress code violations, cutting class, and insubordination–require reporting only

[1] The latest nationally-representative data are from the 2012–13 school year (U.S. Department of Education, 2019). Teachers who leave the profession (attrition) account for roughly half of overall turnover while teachers who move to another school (mobility) account for the other half.

if they result in an out-of-school suspension. If teachers who are more (less) likely to report these discretionary offenses leave at higher (lower) rates, descriptive evidence comparing reported offenses and teacher turnover will overstate (understate) the effect of student behavioral offenses (Feng, 2009).[2]

In this paper, I use administrative panel data from North Carolina on student behavior, discipline, and teacher turnover in public middle schools from 2009–2019 to show whether student behavior affects teacher turnover. I measure student behavior both at the school and in the grade that a teacher is assigned, allowing punitiveness by calculating a school's “propensity to remove” (PTR), as proposed by Sorensen, Bushway, and Gifford (2022), who show that a more punitive discipline policy leads to negative outcomes for students. By conditioning on a detailed description of the offense, PTR measures a school's response to offenses instead of the severity of the underlying behavior. To separate the effect of discipline from the effect of student behavior, I control for the number and severity of offenses in a teacher's school or grade.

This analysis builds on the mixed existing evidence on the effect of school discipline policy on teacher turnover, which does not control for student behavior. Penner, Liu, and Ainsworth (2023) find that when schools discipline a higher percentage of students, they have higher teacher turnover but do not control for student behavior. Pope and Zuo (2023), studying a district-wide policy to lower suspension rates, find that higher suspension rates lead to lower teacher turnover but do not investigate whether changes in suspension policy also lead to changes in student behavior that may affect teacher turnover. In this paper, I show whether schools' disciplinary choices affect teacher turnover while controlling directly for a rich set of student behavior measures.

I find that higher levels of student offenses in a given year lead to higher teacher turnover. In particular, a one standard deviation (SD) increase in offenses per student leads to a 0.5–0.8 percentage point (p.p.) increase in teacher turnover. These effects are driven by low- and middle-severity o

My results provide evidence that the relationship documented in previous literature that compares student

dent behavior across schools. However, I show that estimates from regressions that do not control for all school-level factors likely overstate the relationship between student behavior and teacher attrition. Third, I use school- and grade-level instead of classroom-level student behavior measures, reducing the potential for bias from unobserved confounding variables. As I show in section 4.7.3, student and teacher characteristics

2.1 Sample

I create a sample of 37; 781 teachers at 618 NC public middle schools in 112 school districts between the 2008–09 and 2018–19 school years.[7] My main specifications use 141; 435 teacher-by-year observations. I focus on middle schools because issues of behavior and discipline are most frequent at the middle and high school levels, and because students in middle schools are consistently given standardized tests that can be used to control for student academic performance.[8] I assign teachers to a school each year based on the school where they are employed the most hours per week in that school year, and; if hours are unavailable, I assign teachers to the school in which they earn the highest salary.

2.2 Student Behavior

I create variables measuring student behavior each year at both the school level and in the grade(s) each teacher is assigned to teach in that year. While a teacher might be most responsive to student behavior among the students they teach, they also have the most direct control over the behavior of these students and the reporting of offenses committed by these students. Therefore, I measure behavior at a higher level: among students in the grade(s) they are assigned or at their school. In Section 4.7.3, I assess the sensitivity of my results to this choice.

To measure school behavior, I use incident-level data on student infractions and the resulting discipline assigned. These data include records for all offenses requiring mandatory reporting by state or federal law, any offenses resulting in an out-of-school suspension or, in rare cases, an expulsion[9], and any additional incidents that each school chooses to record.

To create consistent measures across schools, I focus on mandatory offenses, which schools are required to record according to state or federal law. Appendix C lists the offenses recorded in the data and whether they are mandatory under state or federal law. Offenses with required reporting under state law consist of more serious offenses such as assault and firearm possession while offenses with required reporting only under federal law are somewhat less serious, such as possession of tobacco and property damage. As I show in

[7] Charter schools are counted as a separate “district.”

[8] Figure A1 shows the frequency of student offenses by school level. Middle and high schools have higher levels of student offenses than elementary schools.

[9] In my analysis, I group out-of-school suspensions and expulsions.

Section 2.2.1, there is significant variation in the severity of discipline applied to each mandatory offense. To include both more serious offenses and less serious offenses, I measure student behavior using the number of offenses per student requiring mandatory reporting under either state or federal law.

2.2.1 Measuring Offense Severity

There is significant variation in the severity of discipline applied to each offense. Table B1 shows the percentage of incidents in each offense category that result in an out-of-school suspension. The most common offense–fighting–results in an out-of-school suspension in 84% of incidents, while the third most common offense–bullying–results in an out-of-school suspension in only 43% of incidents. There is also variation in the average length of out-of-school suspensions: the average out-of-school suspension length for fighting is 4 days, while the average out-of-school suspension length for bullying is 1 day.

To categorize offenses by severity, I use statewide variation in the length of out-of-school suspensions ap-

I designate “low”, “medium”, and “high” severity offenses based on terciles of the estimated offense fixed effects. Figure 1 shows the percentage of offenses in each category that result in an out-of-school suspension. The lowest severity offenses result in an out-of-school suspension in 43% of incidents, medium severity offenses result in an out-of-school suspension in 70% of incidents, and high severity offenses result in an out-of-school suspension in 82% of incidents. The average out-of-school suspension length–conditional on receiving a suspension–for low-severity offenses is 3 days, for medium-severity offenses is 6 days, and for high-severity offenses is 5 days.

2.4 Disciplinary Policy

I measure student discipline policy based on the probability of a suspension being applied in each incident. The discipline assigned in a particular incident reflects both the severity of student behavior and a student's behavioral history as well as how punitive a school chooses to be for a given incident. Therefore, I measure school discipline policy conditional on both a student's prior offenses and the offense type.

In particular, I adapt Sorensen, Bushway, and Gifford (2022), who estimate principal-specific punitiveness, by estimating a school's "propensity to remove" (PTR) using incident-level regressions

$$r_{ijkt} = {}_{kst} + H_{it} + {}_{g} + {}_{t} + \varepsilon_{ijkt} \tag{2}$$

where r_{ijkt} is an indicator for whether a student received an out-of-school suspension for the incident j involving student i (each student may have more than one offense in a year) for offense-type k in year t. H_{it} is a vector of the student's prior disciplinary record (the number of mandatory offenses in the current year and the number of offenses in the prior year) and ${}_{g}$ and ${}_{t}$ are grade and year fixed effects respectively. The parameters of interest– ${}_{kst}$–are a vector of offense-school-year fixed effects, along with their standard errors. These fixed effects estimate the conditional probability that an out-of-school suspension is applied for a given offense type in a given year.[11]

To reduce the influence of offense fixed effects that are estimated with low precision, I adjust these estimates using empirical Bayes weights, which place more weight on offense fixed effects that are estimated with more precision. Specifically, I use the following empirical Bayes estimator:

$${}_{kst} = \left(1 - \frac{\widehat{\ }_{skt}^{2}}{\widehat{\ }_{skt}^{2} + V(\widehat{\ }_{kst})}\right)\widehat{\ }_{kst} + \frac{\widehat{\ }_{skt}^{2}}{\widehat{\ }_{skt}^{2} + V(\widehat{\ }_{kst})}\overline{\ }_{kt} \tag{3}$$

where $\widehat{\ }_{kst}$ is the estimated standard error of the school-year-offense fixed effects, $\overline{\ }_{kt}$ is the average school-year-offense effect, and $V(\widehat{\ }_{kst})$ is the variance of the set of estimated fixed effects.

[11] Sorensen, Bushway, and Gifford (2022) use a similar strategy to estimate static principal-specific punitiveness using principal-offense fixed effects instead of school-year-offense fixed effects.

Finally, to ensure that PTR estimates are not biased by the mix of offenses in each school, I create a single PTR measure for each school-by-year combination by weighting the individual offense-level estimates based on the sample proportion of offenses faced by all schools in a given year. Specifically, I calculate the following weighted average:

$$\widehat{P_{st}} = \sum_{k} {}_{kst}\frac{n_{kt}}{n_t} \tag{4}$$

where n_{kt} is the number of offenses of type k across all schools in year t and n_t is the total number of offenses in year t. As shown in Table 1, the standard deviation in PTR is 0:12.

2.5 Teacher Outcomes

I measure teacher turnover as whether a teacher is not employed at the same school in the next year. Movement can be the result of mobility–moving to a new school–or attrition–no longer teaching in NC public schools.[12] As shown in Table 1, 9% of teachers in my sample change schools and 10% leave teaching each year. Similar to national statistics, mobility and attrition account for a similar proportion of overall turnover in my sample; however, overall turnover in my sample is somewhat higher than national statistics.

2.6 Student Achievement

To control for the effect of student achievement on teacher turnover, I use student-level data on standardized Math and Reading assessments. I use each student's score on the end-of-grade (EOG) Math and Reading assessments for students in grades 6–8.[13] I standardize each student's score relative to other students in the same grade and year. Because teachers may have preferences for achievement levels at different points in the achievement distribution, I measure the 10th, 25th, 50th, 75th, and 90th percentiles of achievement separately in Math and Reading. I also measure achievement at the classroom-, grade-, and school-level.

[12] I cannot distinguish teaching outside of NC public schools from not teaching at all.

[13] To calculate teacher value added, I also use scores on end-of-course Math and Reading assessments for students in grade 5 to control for prior student performance. See Appendix Appendix D.

student behavior measures in equation 5 with grade-level behavior measures. Standard errors are clustered at the district level.

The main identification assumption is that changes in student behavior or discipline at a school are unrelated to other factors driving teacher retention. This is plausible since most schools have little control over the students they enroll. Indeed, while student behavior and discipline are strongly correlated with student body characteristics *across* schools, they are less strongly correlated with these characteristics *within* schools over time. As shown in Table 2, schools with more offenses have more male, economically disadvantaged, and non-white students; and lower median Math and Reading test scores. However, these correlations are much weaker when looking at within-school variation over time. Similar patterns hold for PTR. However, some correlation between behavior and student body characteristics persists within schools over time, potentially biasing my estimates. To limit the ability of this correlation to affect my results, I control for a rich set of school, grade, and classroom-level covariates and, in some specifications, a detailed set of controls for student achievement.

School fixed effects also rule out the potential for bias from teacher selection of schools based on their preferences for average behavior. As shown in Table 2, schools with more student offenses have teachers who are younger and more likely to be Black. However, teachers likely select schools based on average characteristics and not in anticipation of changes in behavior in the coming year. Indeed, after removing across-school variation using school fixed effects, teacher characteristics are not related to the quartile of student behavior or PTR.

3.2 School-by-Year Fixed E ects

Because unobserved time-varying school-level variables could still explain the relationship between student behavior and teacher attrition, I also use grade-level measures of student behavior to isolate within-school variation in student behavior using school-by-year fixed effects models

[14] Demographic covariates include percentage of female stduents, percentage of economically disadvantaged students, percentage of Black students, percentage of Hispanic students, and percentage of Asian students. Teacher characteristics include teacher experience (novice, 1 – –2 years, 3 – –4 years, or 5 – –10 years), age (five-year bins from age 20 – –65 and an indicator for 65 or older), gender, and race.

$$y_{igt} = {}_1S_{gt} + {}_3X_{gst} + {}_4X_{it} + {}_{st} + {}_{g} + "_{igt} \qquad (6)$$

where S_{gt} is the measure of student behavior for the grade that teacher i teaches in year t and ${}_{st}$ is a school-

PTR and school and district-year fixed effects (see equation 5). On average, I find that changes in the number of mandatory offenses are associated with higher overall turnover: as shown in column 1, an additional mandatory offense per student is associated with 8:6 p.p. higher turnover. This magnitude is significant: an additional standard deviation (SD) in the number of mandatory offenses per student is associated with 0:8 percentage points (p.p.) higher turnover, or 4% of annual turnover. Estimates of the effect of offenses on in middle-severity offenses per student at a school is associated with $0{:}5$ p.p. higher attrition–but estimates for mobility are similar in magnitude. However, in estimates that measure student behavior in the grade that a teacher is assigned and including school-by-year fixed effects–as shown in column 6–low-severity offenses–but not middle- or high-severity offenses–are associated with higher turnover. An additional SD in low-severity offenses in a teacher's grade is associated with $1{:}3$ p.p. higher turnover, accounting for 6% of overall turnover. For middle-severity offenses in a teacher's grade, I can rule out effects larger than $0{:}8$ p.p. This evidence suggests that both middle- and low-severity offenses are associated with higher teacher turnover but I cannot differentiate between the importance of measuring behavior at the school- or grade-level and the importance of school-level confounding variables.

4.2 Student Discipline

Despite the importance of student behavior for teacher turnover, I find limited evidence that a more punitive discipline policy affects teacher turnover independently of student behavior. As shown in Table 3, an additional SD in a school's PTR is associated with $0{:}1$ p.p. higher overall turnover. This estimate is statistically insignificant and precise enough to rule out effects larger than $0{:}7$ p.p. and smaller than $-0{:}5$ p.p. (across all specifications in Table 3). Estimates are particularly small for attrition: an additional SD in a school's PTR is associated with at most a $0{:}3$ p.p. increase or decrease in attrition.

4.3 Controls for Student Achievement

My main specifications control for school-level changes in student composition that may drive both changes in student behavior and discipline in addition to teacher turnover, but do not fully control for all classroom- or grade-level student characteristics. While I include controls for a rich set of student demographics at the school-, grade-, and classroom-level, unmeasured changes in student composition may still be correlated with changes in student behavior and discipline. For example, changes in academic achievement have an independent effect on teacher turnover (Karbownik, 2020) that may not be accounted for by changes in the school-level student body or classroom-level student demographics.

To investigate whether changes in student achievement are driving my results, I estimate my main specifications while adding a rich set of controls for student achievement. In particular, I add controls for school-, grade-, and classroom-level student achievement on end-of-grade Math and Reading assessments. Because teachers may react differently to the achievement levels of the best, worst, and average students in their class, grade, or school, I include controls for the test scores of students at the 10th, 25th, 50th, 75th, and 90th percentiles of a teacher's classroom, grade, and school.

Controls for student achievement have a minimal effect on my overall results. In columns 3 and 7 of Table 3, I present estimates from regressions including student achievement controls. Estimates are close to estimates from regressions that do not include controls for student achievement, suggesting that my results are not driven by correlated changes in student achievement.

4.4 Heterogeneity by Teacher Experience

First-year (novice) teachers react differently to student behavior than the average teacher. Table 4 shows estimates from my main school-level and grade-level specifications separately by teacher experience levels. In contrast to the average teacher, novice teachers are most responsive to high-severity offenses and not low- or middle-severity offenses. Among novice teachers, an additional SD in high-severity offenses in their school is associated with 1.8 p.p. higher turnover, in contrast to estimates for teachers with more experience. This effect accounts for 6% of the 30% overall turnover rate among novice teachers. Among novice teachers, this effect is driven by teacher mobility and not attrition: an additional SD in high-severity offenses in their school is associated with 1.2 p.p. higher mobility and only 0.7 p.p. higher attrition (an estimate that is statistically insignificant). This suggests that novice teachers are more likely to move schools in response to high-severity offenses but are not more likely to leave the profession. On average, novice teachers are not responsive to changes in school PTR.

While informative of teacher's decisions, comparisons of the effect of student behavior on teacher turnover by teacher experience level should be interpreted with caution due to the high level of overall turnover and attrition. Just 70% of teachers in my sample remain in their first school after one year and 14% leave the profession. Because novice teachers are more likely to move schools due to high-severity offenses, the remaining teachers may be less responsive to high-severity offenses not because of their teaching experience but because they were initially less responsive to high-severity offenses and thus did not previously change schools or leave teaching. Therefore, my results should not be interpreted as showing that teaching experience changes how teachers react to student behavior issues. In Section 4.6, I focus on novice teachers, who are least affected by potential selection out of teaching.

My overall results are driven by teachers with prior teaching experience. Among teachers with at least three years of experience, an additional SD in middle-severity offenses at their school is associated with 1:2–1:6 p.p. higher turnover. Among teachers with fewer than three years of experience, I find no statistically significant relationship between school-level middle-severity offenses and teacher turnover. Similarly, grade-level estimates show that the relationship between grade-level student behavior and teacher turnover is driven by teachers with 1–2 years of experience: among these teachers, an additional SD in low-severity offenses in their grade is associated with 3:8 p.p. higher turnover, accounting for 15% of turnover among that group. Among teachers with no experience or more than 2 years of experience, I find no statistically significant relationship between grade-level middle-severity offenses and teacher turnover.

4.5 Heterogeneity by Gender

Teachers differ in their reactions to student behavior based on gender. Despite having similar levels of overall turnover, estimates of the effect of student behavior on teacher turnover are generally larger for male teachers than for female teachers. Table 5 shows estimates from my main school-level and grade-level specifications separately by teacher gender. Among male teachers, an additional SD in middle-severity offenses per student at their school is associated with 1:7 p.p. higher turnover while among female teachers, an additional SD is associated with only 0:7 p.p. higher turnover. For male teachers, this effect is driven by mobility: an additional SD in middle-severity offenses at their school is associated with 1:2 p.p higher mobility and 0:6 p.p. higher attrition. Among female teachers, estimated effects are smaller but driven by attrition: an additional SD in middle-severity offenses is associated with 0:7 p.p. higher turnover, 0:4 p.p. higher attrition, and 0:3 p.p. higher mobility (not statistically significant).

4.6 Novice Teachers

As I show in Section 4.4, novice teachers are most responsive to high-severity student offenses; however, these results are driven by female teachers and by teachers with above-median effectiveness. Table 6 shows estimates of my main specifications for novice teachers overall, and separately by teacher gender and teacher effectiveness (value-added). While an additional SD in high-severity offenses is associated with .81 p.p. higher turnover among novice teachers, among novice female teachers an additional SD in high-severity offenses is associated with .62 p.p. higher turnover. In contrast, high-severity offenses are not associated with higher turnover among novice male teachers.

Among novice teachers with above-median value added (compared to other novice teachers), an additional SD in high-severity offenses is associated with 1.93 p.p. higher turnover. Among teachers with below-median value added, I find no statistically significant relationship between high-severity offenses and turnover. This effect is driven by mobility and not attrition: among novice teachers with above-median effectiveness, an additional SD in high-severity offenses is associated with .63 p.p. higher mobility and not higher attrition. This suggests that effective novice teachers are more likely to move schools in response to high-severity offenses but are not more likely to leave the profession.

While novice teachers are on average not responsive to school discipline, I find evidence that male teachers have higher mobility–but lower attrition–when schools are more punitive. Among novice male teachers, an additional SD in PTR is associated with 1.42 p.p. higher mobility. However, an additional SD in PTR is also associated with .42 p.p. lower attrition, offsetting the effect of mobility on overall turnover. This suggests that student discipline affects the composition of novice teachers who remain at their school or in the profession but does not affect the overall level of turnover.

4.7 Comparison to Previous Research

My analysis diverges from the existing literature on the relationship between student behavior and teacher turnover in three primary ways. First, I measure student behavior using only offenses requiring mandatory reporting, while Feng (2009) uses all offenses. Second, I include school or school-by-year fixed effects to control for school-level confounding variables while Feng (2009) includes only district and year fixed effects. Third, to limit the influence of teachers on behavior and offense reporting, I measure student behavior at the school and grade level. This contrasts with Feng (2009), who uses administrative data from Florida, finding that more student offenses at the classroom level are associated with higher attrition. In this section, I investigate the sensitivity of my results to these choices.

4.7.1 Mandatory vs. Discretionary Offenses

Using a specification similar to Feng (2009), I confirm that the number of mandatory and discretionary offenses in a teacher's classroom is associated with higher attrition. In column 1 of Table 7, I show these results, which are from regressions of teacher attrition and mobility on classroom-level measures of student behavior that include both mandatory and discretionary offenses, PTR, district and year fixed effects, as well as a rich set of controls included in my primary specifications.[15] Consistent with Feng (2009), I find that additional offenses are associated with higher attrition and, in contrast to Feng (2009), find that it is also associated with higher mobility.

This relationship persists when limiting behavior measures to offenses requiring mandatory reporting. In column 2 of Table 7, I show estimates from regressions that measure classroom-level behavior using only mandatory reporting offenses. As with estimates that include discretionary offenses, the estimates are positive and statistically significant. However, the magnitude of estimates from specifications measuring behavior using only offenses requiring mandatory reporting should not be compared directly to estimates from specifications measuring behavior using both mandatory and discretionary offenses. Excluding discretionary offenses primarily removes offenses–such as dress code violations, cutting class, and insubordination–that are less severe than the offenses with mandatory reporting requirements, meaning that an additional mandatory offense is likely to represent a more severe offense than an additional offense in specifications that include discretionary offenses.[16] In column 3, I include measures separately for both mandatory and discretionary o[illegible]-380(posubor79ally)-323(signi79alnt.)-530(Ho

4.7.2 School Fixed E ects

Next, I show that estimates are sensitive to the inclusion of school fixed effects. In columns 4 and 5 of Table 7, I add school and district-by-year fixed effects to the specifications in columns 2 and 3. Estimates for the effect of mandatory offenses are somewhat smaller than in specifications without school and district-by-year fixed effects and–for attrition–no longer statistically significant. Estimates for the effect of discretionary offenses are similar in magnitude to estimates in column 3, while estimates for the effect of mandatory offenses is smaller in specifications that include school and district-by-year fixed effects. In columns 6 and 7, I include school-by-year fixed effects instead of school and district-by-year fixed effects. For mandatory off

In Table 7, I show estimates from regressions of teacher attrition and mobility on classroom-, school-, and grade-level measures of student behavior separately for mandatory and discretionary offenses. In columns 4 and 5, I show estimates from regressions using classroom-level behavior without and without discretionary offenses and including school and district-by-year fixed effects. In columns 8 and 9, I show estimates from regressions using school-level behavior and the same fixed effects. Estimates using classroom-level behavior are somewhat smaller, particularly when not including measures for discretionary offenses. Similarly, in columns 6 and 7, I show estimates from regressions using classroom-level behavior with and without discretionary offenses and including school-by-year fixed effects. In columns 10 and 11, I show estimates from regressions using grade-level behavior and the same fixed effects. Estimates using classroom-level behavior are somewhat larger and estimates using grade-level measures of mandatory offenses are statistically insignificant. These results suggest that my estimates are sensitive to measuring behavior at the classroom level instead of at the school or grade level.

5 Conclusion

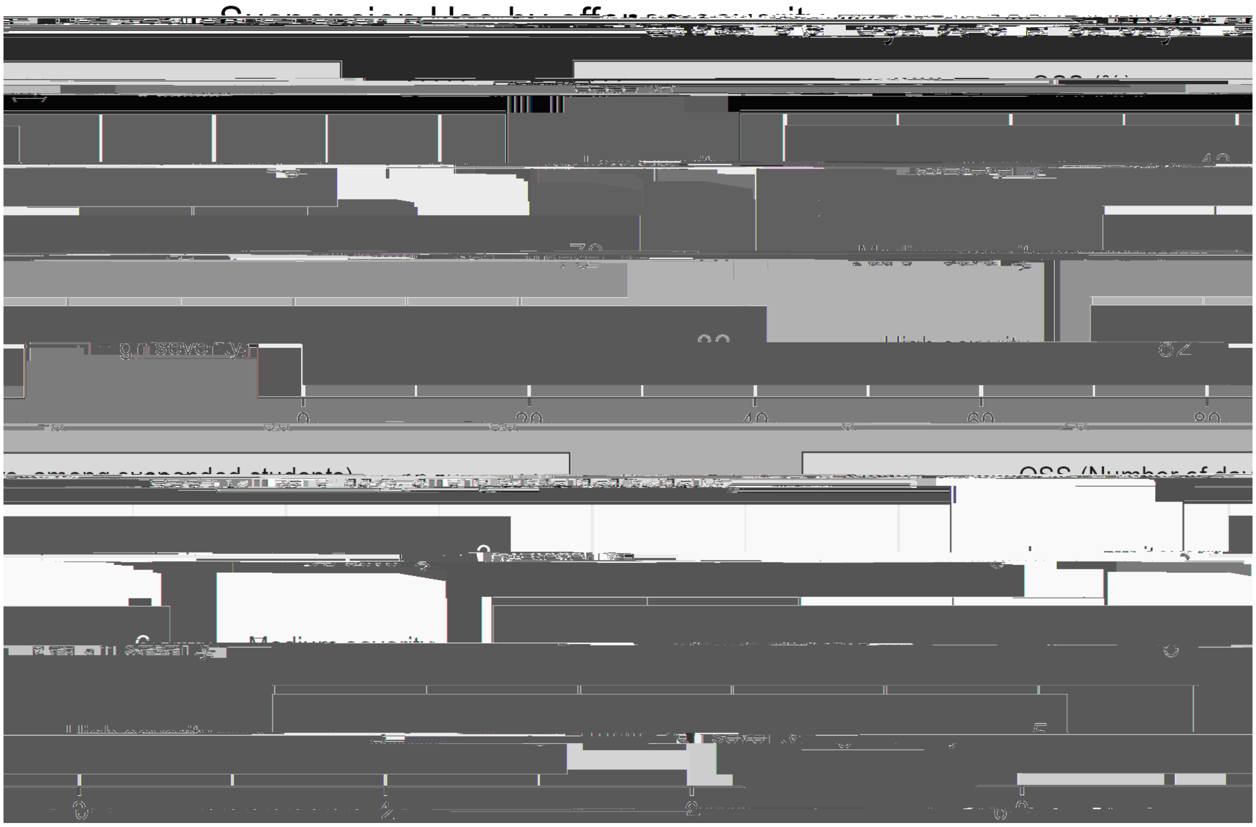
The retention of effective teachers is a critical issue for education policy, with significant equity considerations created by the movement of experienced teachers away from high-poverty schools (Boyd, Grossman, Lankford, et al., 2008). While poor school climate and administrative leadership are often cited as major reasons for teacher turnover, student characteristics also play a significant role (C. K. Jackson, 2009; Karbownik, 2020). In this paper, using detailed administrative data from North Carolina, I show that more student offenses lead to higher teacher turnover. Among novice teachers–who have the high turnover rates–this effect is driven by the high-severity offenses, while among more experienced teachers, the effect is driven by middle- and low-severity offenses. These results suggest that schools and teacher preparation programs need to focus on strategies to support teachers in responding to student behavior issues. Existing evidence suggests that training teachers on effective classroom management strategies can reduce attrition but more evidence is needed to show whether these programs can scale (Bartanen, Bell, James, et al., 2023).

When faced with student behavior issues, many schools have adopted more punitive discipline policies. While there is limited evidence that these policies reduce student offenses (Sorensen, Bushway, & Gifford, 2022), they may also have effects on teacher turnover. However, I find that a more punitive discipline policy does not lead to higher or lower teacher turnover among most teachers. In light of extensive evidence showing that more punitive discipline policies have adverse effects on student outcomes (Bacher-Hicks, Billings, & Deming, 2019; Sorensen, Bushway, & Gifford, 2022), this suggests that efforts aimed at reducing the use of out-of-school suspensions will have limited direct effect on teacher turnover. However, given that I find that more student off

References

Figures

Figure 1



Note: Offense variables include only offenses requiring mandatory reporting under state or federal law. Severity categories are based on the average days of out-of-school suspensions applied to each offense type. See Section 2.2.1 for details.
School-year PTR is the school-year propensity to remove (see Section 3 for details.).

Tables

Table 1: Descriptive statistics

Table 2: Average school characteristics by measures of student behavior and discipline

	Quartile				Wald statistics		
	1st (lowest)	2nd	3rd	4th (highest)	Across schools	Within schools	Within school by year
School: number of mandatory offenses per student							
School Characteristics							
Female students	0:49	0:49	0:48	0:47	12:59***	0:60	
Non-white students	0:38	0:43	0:50	0:65	60:03***	3:24**	
Economically disadvantaged students	0:36	0:43	0:49	0:58	48:87***	3:86***	
Median Math score (z-score)	0:16	0:15	0:13	0:07	28:23***	8:07***	
Median Reading score (z-score)	0:16	0:15	0:13	0:06	32:28***	9:00***	
Grade Characteristics							
Female students	0:49	0:49	0:48	0:47	17:39***	1:67	
Non-white students	0:38	0:43	0:50	0:65	59:19***	2:55*	
Economically disadvantaged students	0:36	0:43	0:49	0:58	50:07***	3:39**	
Median Math score (z-score)	0:16	0:15	0:13	0:07	25:74***	8:77***	
Median Reading score (z-score)	0:16	0:15	0:13	0:06	29:76***	7:53***	
Teacher Characteristics							
Age	42:00	41:59	41:32	40:92	4:81***	0:28	
Female	0:74	0:76	0:75	0:74	12:59***	0:60	
Black	0:09	0:13	0:17	0:28	24:54***	1:93	
White	0:87	0:83	0:80	0:67	24:89***	0:53	
Hispanic	0:02	0:02	0:01	0:02	4:94***	2:44*	

Continued on next page

Table 2, continued

	Quartile				Wald statistics		
	1st (lowest)	2nd	3rd	4th (highest)	Across schools	Within schools	Within school by year
Grade: number of mandatory offenses per student							
School Characteristics							
Female students	0:49	0:49	0:48	0:47	12:79***	0:92	
Non-white students	0:38	0:43	0:50	0:64	53:61***	5:17***	
Economically disadvantaged students	0:37	0:44	0:49	0:58	47:66***	3:97***	
Median Math score (z-score)	0:16	0:15	0:13	0:07	31:16***	2:03	
Median Reading score (z-score)	0:16	0:15	0:13	0:06	35:57***	1:30	
Grade Characteristics							
Female students	0:49	0:49	0:49	0:47	21:41***	7:61***	10:77***
Non-white students	0:39	0:43	0:50	0:64	54:95***	6:18***	5:46***
Economically disadvantaged students	0:36	0:43	0:48	0:58	52:46***	8:07***	3:51**
Median Math score (z-score)	0:16	0:15	0:13	0:07	32:94***	7:09***	11:11***
Median Reading score (z-score)	0:16	0:15	0:13	0:06	38:14***	2:79**	8:07***
Teacher Characteristics							
Age	41:91	41:59	41:44	40:90	3:90***	0:86	1:44
Female	0:75	0:75	0:75	0:75	12:79***	0:92	0:00
Black	0:10	0:13	0:17	0:27	15:21***	0:69	2:04
White	0:86	0:84	0:79	0:68	16:75***	1:07	2:21*
Hispanic	0:02	0:02	0:02	0:02	4:61***	1:11	0:82

Continued on next page

Table 2, continued

	Quartile				Wald statistics		
	1st (lowest)	2nd	3rd	4th (highest)	Across schools	Within schools	Within school by year
	School-year propensity to remove						
School Characteristics							
Female students	0:48	0:48	0:48	0:48	1:29	2:80**	
Non-white students	0:41	0:46	0:52	0:56	13:94***	1:13	
Economically disadvantaged students	0:44	0:46	0:48	0:49	2:07	0:33	
Median Math score (z-score)	0:13	0:13	0:13	0:12	0:33	0:96	
Median Reading score (z-score)	0:13	0:12	0:12	0:12	1:25	2:13*	
Grade Characteristics							
Female students	0:49	0:49	0:48	0:48	1:91	3:96***	
Non-white students	0:41	0:46	0:52	0:56	13:94***	1:11	
Economically disadvantaged students	0:44	0:46	0:48	0:49	2:12*	0:49	
Median Math score (z-score)	0:14	0:13	0:13	0:13	0:18	0:67	
Median Reading score (z-score)	0:13	0:12	0:12	0:12	0:76	1:11	
Teacher Characteristics							
Age	41:49	41:49	41:42	41:43	0:12	0:75	
Female	0:75	0:75	0:75	0:75	1:29	2:80**	
Black	0:11	0:15	0:19	0:22	10:78***	0:11	
White	0:86	0:82	0:77	0:73	10:46***	0:23	
Hispanic	0:01	0:02	0:02	0:02	1:10	0:28	

Note: Wald statistics test the null hypothesis that indicators for quartile of each measure of student behavior or discipline are equal to zero. The within-school estimates come from regressions with school and district-year fixed effects and indicators for the grade taught. The within-school by year estimates come from regressions with school-year fixed effects and indicators for the grade taught.
Offense variables include only offenses requiring mandatory reporting under state or federal law. Severity categories are 3 8.96:

Table 3: Main regression results: school- and grade-level student behavior

	School-level behavior			Grade-level behavior			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Mobility							
Mandatory offenses per student	0:048			0:017	0:006		
	(0:036)			(0:016)	(0:022)		
School-year PTR	0:014	0:011	0:011	0:012			
	(0:016)	(0:017)	(0:017)	(0:017)			
Offenses per student: low severity		−0:041	−0:047			0:087**	0:087*
		(0:047)	(0:051)			(0:043)	(0:044)
Offenses per student: middle severity		0:090	0:094*			−0:029	−0:029
		(0:055)	(0:055)			(0:029)	(0:029)
Offenses per student: high severity		0:227	0:226			0:014	0:001
		(0:150)	(0:154)			(0:118)	(0:116)
Attrition							
Mandatory offenses per student	0:038**			0:014*	0:038**		
	(0:018)			(0:008)	(0:019)		
School-year PTR	−0:002	−0:004	−0:004	−0:003			
	(0:013)	(0:013)	(0:013)	(0:012)			
Offenses per student: low severity		−0:029	−0:023			0:078	0:072
		(0:043)	(0:043)			(0:053)	(0:052)
Offenses per student: middle severity		0:084***	0:078***			0:020	0:023
		(0:026)	(0:025)			(0:036)	(0:034)
Offenses per student: high severity		0:028	0:046			0:040	0:028
		(0:128)	(0:134)			(0:114)	(0:118)
Any turnover							
Mandatory offenses per student	0:086***			0:031*	0:044		
	(0:030)			(0:016)	(0:031)		
School-year PTR	0:013	0:007	0:007	0:009			
	(0:023)	(0:024)	(0:023)	(0:023)			
Offenses per student: low severity		−0:070	−0:070			0:165**	0:159**
		(0:055)	(0:057)			(0:063)	(0:063)
Offenses per student: middle severity		0:174***	0:172***			−0:009	−0:006
		(0:055)	(0:056)			(0:050)	(0:048)
Offenses per student: high severity		0:255	0:272			0:054	0:029
		(0:183)	(0:191)			(0:165)	(0:166)
Obs.	141; 435	141; 435	141; 435	141; 435	141; 435	141; 435	141; 435
School FE	X	X	X	X			
District-year FE	X	X	X	X			
School-year FE					X	X	X
School-level controls	X	X	X	X			
Grade-level controls	X	X	X	X	X	X	X
Classroom-level controls	X	X	X	X	X	X	X
Teacher demographic controls	X	X	X	X	X	X	X
Grade taught indicators	X	X	X	X	X	X	X
School-level achievement controls			X				
Grade-level achievement controls			X				X
Classroom-level achievement controls			X				X

Note: Offense variables include only offenses rew2J/F103 10.9091 Tf 2.(dB3:1(indicator1 Tf 5.454 0 Td [(:)]T

Table 4: Heterogeneity by experience level

	School-level	Grade-level

Table 5: Heterogeneity by gender

	School-level		Grade-level	
	Male	Female	Male	Female
Mobility				
Offenses per student: low severity	0:033	−0:063	0:223**	0:023
	(0:079)	(0:050)	(0:106)	(0:056)
Offenses per student: middle severity	0:184**	0:048	−0:123**	−0:011
	(0:091)	(0:054)	(0:060)	(0:035)
Offenses per student: high severity	0:017	0:294	0:284	−0:041
	(0:173)	(0:219)	(0:184)	(0:176)
School-year PTR	−0:004	0:014	−0	

Table 6: Novice teachers

	School-level	Grade-level

Table 7: Alternative specifications

	Classroom-level behavior							School-level behavior		Grade-level behavior	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Mobility											
Mandatory and discretionary offenses per student	0:004***										
	(0:001)										
School-year PTR	−0:009	−0:010	−0:010	0:015	0:014			0:016	0:014		
	(0:011)	(0:011)	(0:011)	(0:017)	(0:017)			(0:016)	(0:016)		
Mandatory offenses per student		0:042***	0:041***	0:032***	0:037***	0:020**	0:036***	0:036	0:048	−0:006	0:006
		(0:010)	(0:011)	(0:010)	(0:009)	(0:009)	(0:009)	(0:037)	(0:036)	(0:023)	(0:022)
Discretionary offenses per student			0:001	0:002		0:007***		0:004***		0:009**	
			(0:001)	(0:002)		(0:002)		(0:001)		(0:005)	
Attrition											
Mandatory and discretionary offenses per student	0:006***										
	(0:001)										
School-year PTR	0:002	−0:002	0:002	0:000	−0:002			0:001	−0:002		
	(0:011)	(0:012)	(0:011)	(0:012)	(0:012)			(0:013)	(0:013)		
Mandatory offenses per student		0:030***	0:019***	0:008	0:021***	0:003	0:022***	0:022	0:038**	0:016	0:038**
		(0:006)	(0:006)	(0:006)	(0:006)	(0:008)	(0:007)	(0:019)	(0:018)	(0:020)	(0:019)
Discretionary offenses per student			0:005***	0:006***		0:008***		0:006***		0:017***	
			(0:001)	(0:001)		(0:002)		(0:001)		(0:005)	
Any turnover											
Mandatory and discretionary offenses per student	0:010***										
	(0:002)										
School-year PTR	−0:007	−0:012	−0:008	0:016	0:012			0:018	0:013		
	(0:014)	(0:015)	(0:015)	(0:024)	(0:024)			(0:023)	(0:023)		
Mandatory offenses per student		0:072***	0:059***	0:040***	0:058***	0:023*	0:057***	0:058*	0:086***	0:010	0:044
		(0:014)	(0:014)	(0:013)	(0:012)	(0:012)	(0:013)	(0:032)	(0:030)	(0:032)	(0:031)
Discretionary offenses per student			0:005**	0:008***		0:015***		0:011***		0:026***	
			(0:002)	(0:002)		(0:003)		(0:002)		(0:007)	
Obs.	141; 435	141; 435	141; 435	141; 435	141; 435	141; 435	141; 435	141; 435	141; 435	141; 435	141; 435
District FE	X	X	X								
Year FE	X	X	X								
School FE				X	X			X	X		
District-year FE				X	X			X	X		
School-year FE						X	X			X	X
School-level controls	X	X	X	X	X	X	X	X	X	X	X
Classroom-level controls	X	X	X	X	X	X	X	X	X	X	X
Grade-level controls	X	X	X	X	X	X	X	X	X	X	X
Teacher demographic controls	X	X	X	X	X	X	X	X	X	X	X
Grade taught indicators	X	X	X	X	X	X	X	X	X	X	X

Note: Offense variables include only 106.569 Td [(Note:)-369 X X X X X X X

Appendix A Appendix Figures

Figure A1: Offenses and Suspensions by School Level

Note: Offense variables include only offenses requiring mandatory reporting under state or federal law. Severity categories are based on the average days of out-of-school suspensions applied to each offense type. See Section 2.2.1 for details.
School-year PTR is the school-year propensity to remove (see Section 3 for details.).

Appendix B Appendix Tables

Table B1: O ense severity by œnse category

O ense	% Suspended	Suspension Days	Suspension Days (Avg. > 0)	N
Highest severity				
Assault involving a weapon	89	25	31	812
Assault on school personnel not resulting in an injury	83	9	11	; 08153
Assault resulting in an injury	84	20	26	; 0125
Bomb threat	77	23	32	; 1106
Controlled substance use or possession	89	11	13	; 36514
Distribution of a controlled substance	83	15	20	; 3833
Possession of a rearm	90	9	10	; 1433
Possession of a weapon (non- rearm)	83	10	13	; 62895
Robbery without a dangerous weapon	91	17	19	; 3716
Medium severity				
Alcohol use or possession	87	7	9	; 10369
Assault not resulting in an injury	72	4	6	1; 27045
Communicating threats of attack with a rearm	72	5	7	866
Communicating threats of attack with a weapon (non- rearm)	69	4	6	; 1171
Extortion	76	4	6	663
Fighting	84	4	4	606; 904
Gang activity	78	7	10	1; 3737
Posession of drug paraphernalia	87	7	8	; 9686
Possession of another's prescription drug	85	9	12	; 4148
Sexual Assault	81	7	10	; 2403
Lowest severity				
Bullying	43	1	3	93; 513
Communicating threats	71	4	6	; 8401
Communicating threats of attack without a weapon	55	2	4	; 925
Discrimination	36	1	3	; 1205
Harrassment - other	46	2	3	; 9293
Possession of tobacco	51	1	3	; 5926
Property damage	44	2	4	; 50320
Sexual Harassment	71	3	4	; 3958
Tobacco use	44	1	3	5; 9864
Verbal harassment	43	1	3	; 48338

Includes o enses with at least 100 observations.

Note: O ense variables include only œnses requiring mandatory reporting under state or federal law. Severity categories are based on the average days of out-of-school suspensions applied to each œnse type. See Section 2.2.1 for details.

School-year PTR is the school-year propensity to remove (see Section 3 for details.).

Table B2: Average student and teacher characteristics by classroom-level measures of student behavior and discipline

	Quartile				Wald statistics		
	1st (lowest)	2nd	3rd	4th (highest)	Across schools	Within schools	Within school by year
Classroom: number of mandatory offenses per student							
School Characteristics							
Female students	0:49	0:49	0:48	0:47	14:43***	1:25	
Non-white students	0:41	0:44	0:50	0:60	35:92***	6:76***	
Economically disadvantaged students	0:39	0:44	0:49	0:55	42:80***	6:03***	
Median Math score (z-score)	0:15	0:15	0:13	0:09	20:53***	1:55	
Median Reading score (z-score)	0:14	0:15	0:12	0:08	23:68***	0:63	
Grade Characteristics							
Female students	0:49	0:49	0:49	0:47	20:48***	7:39***	9:12***
Non-white students	0:42	0:44	0:50	0:60	36:65***	8:28***	9:39***
Economically disadvantaged students	0:39	0:44	0:49	0:55	45:52***	7:99***	6:37***
Median Math score (z-score)	0:15	0:16	0:13	0:09	22:54***	4:07***	7:22***
Median Reading score (z-score)	0:14	0:15	0:12	0:08	25:22***	2:40*	5:98***
Classroom Characteristics							
Female students	0:49	0:49	0:48	0:43	176:41***	238:71***	257:80***
Non-white students	0:40	0:44	0:51	0:63	58:67***	27:86***	34:17***
Economically disadvantaged students	0:38	0:43	0:50	0:60	126:88***	89:02***	131:18***
Median Reading score (z-score)	−0:57	0:09	0:01	−0:24	114:39***	94:13***	91:42***
Median Math score (z-score)	−0:51	0:10	0:01	−0:28	116:23***	103:28***	98:50***
Teacher Characteristics							
Age	42:12	41:39	41:11	41:20	22:65***	13:49***	12:09***
Female	0:78	0:74	0:74	0:75	14:43***	1:25	0:00
Black	0:10	0:13	0:17	0:27	16:99***	15:43***	14:44***
White	0:85	0:83	0:79	0:69	18:76***	16:56***	16:88***
Hispanic	0:02	0:02	0:01	0:02	3:92***	9:57***	8:66***

Note: Offense variables include only offenses requiring mandatory reporting under state or federal law. Severity categories are based on the average days of out-of-school suspensions applied to each offense type. See Section 2.2.1 for details.

School-year PTR is the school-year propensity to remove (see Section 3 for details.).

Note: Wald statistics test the null hypothesis that indicators for quartile of each measure of student behavior or discipline are equal to zero. The within-school estimates come from regressions with school and district-year fixed effects and indicators for the grade taught. The within-school by year estimates come from regressions with school-year fixed effects and indicators for the grade taught. “N” presents the number of teacher-year observations used in each calculation.

*p<0.10; **p<0.05; ***p<0.01

Table B3: Average classroom characteristics by measures of student behavior and discipline

	Quartile				Wald statistics		
	1st (lowest)	2nd	3rd	4th (highest)	Across schools	Within schools	Within school by year
School: number of mandatory offenses per student							
Classroom characteristics 45*** 0							
Female students	0:48	0:48	0:47	0:46	17:65***	1:96	
Non-white students	0:39	0:44	0:50	0:65	54:39***	1:81	
Economically disadvantaged students	0:38	0:44	0:50	0:59	49:27***	3:41**	
Median Reading score (z-score)	–0:12	–0:15	–0:18	–0:24	4:41***	0:09	
Median Math score (z-score)	–0:12	–0:15	–0:18	–0:24	4:45***	0:24	

Grade.137 0 Td [(:)]TJ/F103 10.9091 Tf 2.727 0 Td [(09)]TJ -411.09 -13.8a573 4 [(45***)-1596(0)]TJ/F145 10.9091 Tf xs83 10

Appendix C List of O enses

O enses Requiring Reporting Under State Law

Alcohol use or possession
Assault involving a weapon
Assault on school personnel not resulting in an injury
Assault resulting in an injury
Bomb threat
Burning of a school building
Controlled substance use or possession
Death by other than natural causes
Distribution of a controlled substance
Distribution of a prescription drug
Homicide
Kidnapping
Posession of a controlled substance
Posession of a firearm
Posession of a weapon (non-firearm)
Possession of a firearm
Possession of a weapon (non-firearm)
Possession of another’s prescription drug
Rape
Robbery with a dangerous weapon
Robbery without a dangerous weapon
Sexual Assault

O enses Requiring Reporting Under Federal Law (and not State Law)

Assault not resulting in an injury
Bullying
Communicating threats
Communicating threats of attack with a firearm
Communicating threats of attack with a weapon (non-firearm)
Communicating threats of attack without a weapon
Discrimination
Extortion
Fighting
Gang activity
Harrassment - other
Posession of drug paraphernalia
Possession of tobacco
Property damage
Sexual Harassment
Tobacco use
Verbal harassment

O enses Not Requiring Reporting

Aggressive behavior
Alcohol intoxication
Being in an unauthorized area
Bus misbehavior
Cell phone use
Controlled substance intoxication
Cutting class
Dangerous acts
Discipline action violation
Disorderly conduct
Disrespect of faculty/staff
Disruptive behavior
Dress code violation
Excessive display of affection
Excessive tardiness
False fire alarm
Falsification of information
Gambling
General rule violation
Hazing
Honor code violation
Inappropriate behavior
Inappropriate items on school property
Indecent exposure
Insubordination
Intimidation
Misuse of technology
Mututal sexual contact between students
No immunization
Other
Physical Exam
Posession of own prescription drug
Possession of counterfeit items
Possession of drug paraphernalia
Profanity
Staff Offense
Theft
Threats
Truancy
Unlawfully setting a fire
Use of counterfeit items

Appendix D Teacher Value Added Estimation

To assess heterogeneity by teacher quality, I estimate teacher value added (VA) on standardized Math and ELA exams following Chetty, Friedman, and Rockoff (2014). These estimates use student scores from end-of-course Math and Reading assessments for the students each teacher is assigned.

I estimate teacher VA using the following steps. First, I capture residuals from the following teacher fixed effects regression:

$$a_{ijt} = {}_i + {}_t + {}_g + {}_1 X_{jt} + {}_2 X_{it} + \text{"}_{ijt} \tag{7}$$

where a_{ijt} is the Math or Reading test score for student j assigned to teacher i (students are assigned to more than one teacher) in year t, standardized relative to other students in that subject, grade, and year. ${}_i$, ${}_t$, and ${}_g$ are teacher, year, and grade fixed effects respectively. X_{jt} is a vector of one-year-lagged student test score controls in both Math and Reading. X_{it} is a vector of classroom-, grade-, and teacher-level controls for student demographics.

Second, I create “modified residuals” for each observation, ignoring the teacher fixed effects in equation 7:

$$\widetilde{\text{"}}_{ijt} = a_{ijt} - (\widehat{\ }_1 X_{jt} + \widehat{\ }_2 X_{it} + \widehat{\ }_i + \widehat{\ }_t + \widehat{\ }_g) \tag{8}$$

Third, I average these residuals by teacher and year, weighted by the number of students assigned to each teacher in each year.

Finally, I estimate estimate estimate4250(year)55(.)]TJ 0 (controls)-229(f -21.923 Td [(ELA)-373g4stimates)-37

Because much of my analysis focuses on the attrition of novice teachers, I estimate VA for novice teachers using a modified version of equation 9 that includes only the constant term. Additionally, I create an "overall" teacher VA variable that averages the Math and Reading VA estimates for each teacher in each year (when available).