EdWorkingPaper No. 25-1146

# The Peer Effects of Grade Retention

NaYoung Hwang
University of New Hampshire

Cory Koedel
University of Missouri

We study the peer effects of grade retention in the context of Indiana's statewide third-grade retention policy. When a retention occurs, it changes the peer group for two cohorts: rising fourth graders who lose a peer and rising third graders who gain a peer. We identify peer effects in both cohorts by leveraging plausibly exogenous variation in cohort-level retention rates caused by a discontinuity in the rule that determines which students are retained. We find that when a student is retained, rising fourth graders who lose the student as a peer are less likely to have a disciplinary incident. However, the effect fades out after one year, and there is not consistent evidence of peer effects on other outcomes for promoted students. For rising third graders who gain the retained student as a peer, we find no evidence of peer effects. We conclude that the total peer effects of grade retention are limited.

VERSION: March 2025

Suggested citation: Hwang, NaYoung, and Cory Koedel. (2025). The Peer Effects of Grade Retention. (EdWorkingPaper: 25 -1146). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/be3d-6t30

# The Peer Effects of Grade Retention

NaYoung Hwang
Cory Koedel

We study the peer effects of grade retention in the context of Indiana's statewide third-grade retention policy. When a retention occurs, it changes the peer group for two cohorts: rising fourth graders who lose a peer and rising third graders who gain a peer. We identify peer effects in both cohorts by leveraging plausibly exogenous variation in cohort-level retention rates caused by a discontinuity in the rule that determines which students are retained. We find that when a student is retained, rising fourth graders who lose the student as a peer are less likely to have a disciplinary incident. However, the effect fades out after one year, and there is not consistent evidence of peer effects on other outcomes for promoted students. For rising third graders who gain the retained student as a peer, we find no evidence of peer effects. We conclude that the total peer effects of grade retention are limited.

Affiliations: Hwang is an assistant professor in the Department of Education at the University of New Hampshire. Koedel is a professor in the Department of Economics and Truman School of Government and Public Affairs at the University of Missouri.

Acknowledgements: We thank the Indiana Department of Education for providing access to administrative records and supporting our independent analysis. This paper was supported by Notre Dame's Center for Research on Educational Opportunity (CREO), the Institute for Educational Initiatives, and their partnership with the Indiana Department of Education. We also appreciate the support of Mark Berends and Roberto Peñaloza. The opinions expressed are those of the authors and do not necessarily reflect the views of their institutions. Any errors are solely the responsibility of the authors.

## 1. Introduction

Forty-one states and the District of Columbia have early literacy policies, with 22 either recommending or mandating third-grade retention for struggling readers (Westall and Cummings, 2023). These policies reflect a growing emphasis on early literacy as a gateway to more advanced academic skills. Many studies have examined the effects of grade retention on the outcomes of retained students themselves, but little research explores its impacts on their peers. We contribute to the thin literature in this area by studying the peer effects of grade retention in the context of Indiana's statewide third-grade retention policy.

When a retention occurs, it changes the peer group for two cohorts. Students in the retained student's original cohort lose a peer, and students in the cohort below gain one. While there is no direct evidence on this, previous research suggests students in the original cohort may benefit from the retention because retained students typically struggle academically and exhibit behavioral problems, which can negatively impact their peers (e.g., Burke and Sass, 2013; Carrell and Hoekstra, 2010; Hanushek et al., 2003; Imberman et al., 2012; Kristoffersen et al., 2015).

For students in the cohort below who gain the retained student as a peer, the expected effect is less clear. While it is reasonable to expect a negative peer effect for the same reasons to expect a positive peer effect in the original cohort, the retention directly affects the retained student, which could change the peer effect. Most notably, research shows retained students typically improve tremendously in the distribution of academic performance when placed in a new cohort, especially when the retention occurs in an early grade (e.g., Figlio, Karbownik, and Özek, 2023; Hwang and Koedel, 2025; Schwerdt, West, and Winters, 2017). This improvement likely limits the "low achievement" channel through which retained students negatively affect their peers.[1] Moreover, in

[1] The retention could also alter the retained student's behavior, though the literature is mixed on the behavioral impacts of early-grade retention. Hwang and Koedel (2025) and Martorell and Mariano (2018) find grade retention

our context of a statewide retention policy focused on the third grade, the presence of a retained student in a cohort of rising third graders could increase the salience of the retention threat. Students and parents could put forth more effort in response, leading to a positive peer effect. Therefore, the expected impact on students who gain the retained student as a peer is uncertain.

We estimate the peer effects of grade retention on four student outcomes: math achievement, English language arts (ELA) achievement, disciplinary incidents (suspensions and expulsions), and unexcused absences from school. We track these outcomes for four years after the retention event for both rising fourth graders, who lose retained students as peers, and rising third graders, who gain them as peers. To identify peer effects, we leverage a discontinuity created by Indiana's test-based retention rule. Our research design can be described as an aggregated version of a regression discontinuity (RD) model, which we implement using an instrumental variables (IV) framework.

We find little evidence of retained-student peer effects. Our most pronounced finding is that rising fourth graders who lose a peer due to retention are less likely to be suspended or expelled during the fourth grade. However, while our estimate of this effect is large relative to the baseline mean, it is small in an absolute sense and fades out after one year. Along other dimensions, we do not find evidence of peer effects on rising fourth graders, nor do we find any evidence of peer effects on rising third graders.

A limitation of our research design is that it only permits the identification of peer effects at the school-by-grade level. This raises the concern that our estimates may understate more localized effects within classrooms (Burke and Sass, 2013). We cannot directly isolate classroom-level effects, especially for rising 4$^{th}$ graders who lose a peer (because the potential 4$^{th}$ grade

in the early grades has no impact on behavior in Indiana and New York City, respectively; Özek (2015) finds negative impacts on behavior in Florida.

classroom of the retained student is unobserved). However, if we assume the peer effects are entirely concentrated within classrooms—i.e., cross-classroom spillovers are zero—we can convert our cohort-level estimates to classroom-level treatment effects. This exercise gives an upper bound of sorts on how large classroom-level peer effects can be. The implied classroom-level effect on rising fourth graders' disciplinary outcomes is even larger relative to the baseline mean; however, it remains substantively small. For the other outcomes, even if we take statistically insignificant point estimates at face value, our estimates imply small classroom-level peer effects. Moreover, the achievement estimates are offsetting across grades, implying small gains for rising fourth graders and small losses for rising third graders.

We interpret our findings holistically as showing that the total peer effects of grade retention are limited. If anything, they suggest slight overall improvements in peer outcomes, driven by immediate—albeit short-lived—reductions in disciplinary incidents in retained students' original cohorts.

# 2 Background

## 2.1 The Effects of Grade Retention on Retained Students

There is a large, long-standing literature on the effects of grade retention on retained students themselves. Relatively recently, a small but high-quality group of studies provides credible causal evidence by leveraging test-based discontinuities in the rules that determine which students are retained. The findings from the regression discontinuity (RD) literature are mixed—some studies find positive impacts of retention on retained students (Diaz et al., 2021; Greene and Winters, 2007; Figlio, Karbownik, and Özek, 2023; Hwang and Koedel, 2025; Schwerdt, West, and Winters, 2017), while others find mixed (Eren, Depew, and Barnes, 2017; Jacob and Lefgren, 2004, 2009; Roderick and Nagaoka, 2005), null (Martorell and Mariano, 2018), or negative impacts (Eren, Lovenheim, and Mocan, 2022; Manacorda, 2012; Mariano, Martorell, and Berglund, forthcoming; Özek, 2015). At first glance, it seems difficult to draw conclusions from the literature. However, a key pattern emerges upon closer review: the impacts of grade retention are significantly more positive—although not exclusively positive—when retention occurs in an earlier grade. Most of the negative impacts are found in instances where retention occurs in the sixth grade or later. A potential explanation is that the stigma and weakened sense of belonging associated with retention have larger effects on older students (Anderson et al., 2005; Ou and Reynolds, 2010).[2]

Among existing studies, Hwang and Koedel (2025) is most relevant to our setting, as these authors estimate the effects of third-grade retention on retained students in Indiana during the same period we study. They show that prior to the retention event, retained students are significantly behind their peers academically and more likely to be absent and suspended from school. When a retention occurs, rising fourth graders lose a peer with these characteristics. Rising third graders gain the retained student as a peer, inclusive of the effect of the retention itself. Like most RD studies of early-grade retention, Hwang and Koedel (2025) find large and positive effects of retention on the same-grade academic achievement of retained students. Retained students' academic gains diminish over time but remain large and statistically detectable through the sixth grade. Hwang and Koedel (2025) also test for impacts on behavioral and attendance outcomes, finding that while retained students are more likely to be absent and subject to disciplinary action than other students, retention itself does not cause changes to these behaviors.

[2] For more detailed discussions of the literature see Hwang and Koedel (2025) and Özek and Mariano (2023). For a review of older studies that use research designs with weaker causal designs, see Allen et al. (2009).

Hwang and Koedel's (2025) findings help to characterize the peer treatments we study. When a retention occurs, rising fourth graders lose a peer with very poor on-grade academic performance, frequent disciplinary incidents, and attendance problems. While the absenteeism and disciplinary issues do not change on average after the retention, retained students are no longer in the bottom tail of the distribution of academic performance in their new cohorts.

### 2.2 The Peer Effects of Retained Students

A rich body of research examines how school peers influence students' academic and behavioral outcomes. Most relevant for understanding the peer effects of retained students are studies that examine the impacts of low-achieving and disruptive peers.[3] Starting with the former, research generally finds increased exposure to low-achieving peers in the elementary grades leads to lower academic achievement, though the magnitude of peer-effect estimates varies considerably (Burke and Sass, 2013; Hanushek et al., 2003; Imberman et al., 2012; Lefgren, 2004; Sojourner, 2012). Findings also vary in terms of how the peer effects are distributed. Burke and Sass (2013) find that the negative peer effects of low achievers are concentrated among low- and middle-performing students. They also find high-performing students *benefit* from exposure to more low-achieving peers, which they suggest may be due to the self-segregation of high and low achievers within classrooms. In contrast, most other research finds negative impacts of low-achieving peers throughout the distribution of academic performance (e.g., Hanushek et al., 2003; Imberman et al., 2012; Sojourner, 2012).

Research also shows disruptive peers negatively affect student outcomes. A compelling example is students who are exposed to domestic violence. Carrell and Hoekstra (2010) show these

[3] We provide a brief review covering only a small portion of the peer effects literature, focusing primarily on studies of peer effects in early grades. For more comprehensive reviews of the peer effects literature, see Epple and Romano (2011) and Sacerdote (2011).

students have lower academic achievement and more behavioral incidents than their non-abused peers, and large negative peer effects. Similarly, Kristoffersen et al. (2015) and Fletcher (2019) find exposure to highly disruptive peers significantly reduces academic achievement. These studies focus on students who are likely to be more disruptive than retained students, on average, so it is reasonable to expect any peer effects of retained students operating through this channel to be more moderate.

In terms of direct evidence on the peer effects of retained students, there are just a handful of studies, all of which focus on the effect of gaining a retained peer (akin to our analysis of peer effects on rising third graders). Gottfried (2013a) and Gottfried (2013b) estimate the effects of exposure to retained peers on attendance and achievement, respectively, in Philadelphia elementary schools. Both studies use student fixed effects models that leverage variation within students over time in classroom exposure to retained students for identification. The results suggest exposure to an additional retained peer decreases student achievement and increases unexcused absences. However, a concern is that the identifying conditions of the models may not be met; for instance, if dynamic factors not captured by the fixed effects simultaneously increase retention and negatively impact other students, it could cause bias in the direction of overstating the negative peer effects of retained students.

Bietenbeck (2020) conducts a conceptually similar analysis but leverages random variation in exposure to retained students in kindergarten via Project STAR. He finds a large and immediate negative peer effect on math test scores for students with more kindergarten repeaters in their class, but the effect fades out after one year. He finds no evidence of a peer effect on reading test scores in the short or long run. He also finds *positive* long-run effects of exposure to retained peers in kindergarten on students' grade point averages and educational persistence in high school. He shows the positive effects cannot be explained by selection effects or differential access to resources. While he does not directly identify a mechanism, Bietenbeck (2020) hypothesizes that compensatory behavioral adjustments by teachers, parents, and students themselves may drive the long-term benefits.

In a related paper, Lavy et al. (2012) estimate the peer effects of students who are old for their grade in high school. They refer to these students as "repeaters" (or "low ability students") but acknowledge many did not actually repeat a grade and are old for their grade because of delayed school entry. Lavy et al. (2012) use cross-cohort, within-school variation in exposure to these students in high school to estimate their peer effects. In contrast to Bietenbeck (2020), they find more cohort-level exposure to "repeaters" has negative effects on students' academic outcomes in high school. An exploratory analysis of mechanisms suggests having more repeaters in a high-school cohort worsens teachers' pedagogical practices, lowers the quality of student-to-student and student-to-teacher relationships, and increases classroom disruptions.[4]

The literature on the peer effects of retained students is small and, from a policy perspective, incomplete, because no study has examined how promoted students are impacted by the *absence* of a retained peer. Moreover, we are not aware of any prior research in the context of a statewide, grade-gated retention policy. These policies are now common in the U.S. and unlike informal retention policies, they create a plausible new channel for peer effects: students rising into the gated grade—typically the third grade—may experience a more salient retention threat when a student is retained from the cohort above. If so, all else equal, they may respond positively to the presence of a newly retained student in an effort to avoid their own retention.

[4] Relatedly, Hill (2014) studies the peer effects of students who repeat particular courses in mathematics in high school. Noting that course repetition and grade repetition are different interventions, he finds exposure to more course repeaters increases the probability of course failure among students enrolled in the course for the first time.

## 3 Data and Context

Indiana's third-grade retention policy was first implemented during the 2011-12 school year. It requires students to attain what the Indiana Department of Education (IDOE) deems foundational reading skills by the end of the third grade, prior to promotion to the fourth grade. The state measures these skills using the Indiana Reading Evaluation and Determination (IREAD-3) test. Students with scores below the threshold for promotion are required to repeat the third grade unless they qualify for an exemption (the threshold is set at 446 points; the full range of scores is 200 to 650). Exempted students include English language learners, students with disabilities, and students who had previously been retained twice, all of whom can be promoted even if their IREAD-3 test scores do not reach the threshold. All third graders are given two chances to pass the test. The first chance is in March and the second is over the summer, in either June or July. Hwang and Koedel (2025) show that over 96 percent of students who fail the spring test take the summer test. Given this, they build their RD analysis around the summer test, and we do the same.[5]

We use an administrative data panel from the IDOE for our analysis. The data panel starts with the first year of the policy, in 2011-12, and continues through the 2016-17 school year. The data include basic demographic and socioeconomic information, IREAD-3 test scores, Indiana Statewide Testing for Educational Progress-Plus (ISTEP+) test scores, and attendance and disciplinary outcomes for students statewide. The IREAD-3 test is used exclusively to measure reading proficiency at the end of the third grade and is central to the retention policy. We use

[5] The 96 percent number understates the coverage of the summer test because the dominator includes students who exit Indiana Public Schools after the spring test. That said, in principle, the RD could also be built around the spring test. A limitation of using the spring test is that because retention decisions are based on the summer test, using the spring test makes for a (much) fuzzier design. Given the high rate of summer testing, the loss of statistical power that comes with using the spring test is difficult to justify.

IREAD-3 test scores in our models that leverage variation from the regression discontinuity around the test threshold for identification. The ISTEP+ is the state standardized exam administered to all students in ELA and math in grades 3-8. We use ISTEP+ test scores as outcomes when we estimate the achievement effects of peer retention.

We track outcomes for cohorts of rising third- and fourth-grade students from 2011-12 through 2015-16. Given that our data panel runs through 2016-17, the 2015-16 cohort is the last cohort for which we can observe post-retention outcomes. During our study period, third grade was the only formal grade gate in Indiana, and the retention policy was entirely test-based.[6] Beginning with the 2017-18 school year the retention policy changed, allowing schools to consider overall academic performance and the needs of individual students, in addition to IREAD-3 test scores for retention decisions (Indiana Department of Education, 2017).

The Indiana policy is designed to retain only very low-performing students. The effective retention rate during the period we study is just 1.8 percent. This is similar to the 3-percent rate in the New York City policy studied by Mariano and Martorell (2018) and Mariano, Martorell, and Berglund (forthcoming), but lower than in Florida, where about 10 percent of third graders are retained (Winters and Greene, 2012), and in Chicago Public Schools, where 21 percent of third graders are retained (Jacob and Lefgren, 2004). Some of these differences across locales reflect differences in student circumstances, such as the high concentration of disadvantage in Chicago Public Schools. But even accounting for this, the Indiana policy is more narrowly targeted than similar policies that have been studied previously. An implication is that the retained-peer treatments we study are likely drawn from lower in the distribution of academic performance, on average, than in many other contexts. We expect this feature of our evaluation to lead to stronger

[6] Although the third grade was the only policy-targeted grade gate, retentions did occur in other grades, albeit rarely. Over grades 4-7, no more than 0.2 percent of students were retained in any grade during the sample period.

peer effects on a per-treatment basis. However, the low retention rate also means that variation in student exposure to retained peers is more limited in our setting than in settings where the retention threshold is higher. Although this raises concerns about statistical power, below we show that our models are sufficiently powered to detect educationally meaningful peer effects.

Table 1 provides descriptive statistics for our sample. The top horizontal panel documents student characteristics by retention status. Students enrolled for free or reduced-price lunch (FRL), individualized education program (IEP) students, and Black and Hispanic students are overrepresented among retained students. Retained students also have much lower (first-time) third-grade test scores (by about 1.1 and 1.2 standard deviations in math and English language arts, respectively), about twice as many unexcused absences (3.7 versus 1.7 days), and are over three times as likely to have a disciplinary incident (11 percent versus 3 percent) compared to their promoted counterparts. Not shown in the table is the effect of retention on retained students' on-grade academic achievement. Hwang and Koedel (2025) find that in the fourth grade, the earliest grade they test for on-grade achievement effects, marginally retained students score roughly 0.50 standard deviations higher than their marginally promoted counterparts on the ISTEP+ in both math and English language arts.

The bottom panel of Table 1 documents features of third-grade cohorts in Indiana. The average cohort of rising fourth graders has 88 students and experiences 1.36 retained peers. Just over four students in the average cohort score within 25 points of the retention threshold on the IREAD-3 test; we leverage variation in retentions among these low-scoring students for identification in our preferred models.

## 4 Methods

We begin with a basic model linking the outcomes of promoted students—i.e., rising fourth graders—to their exposure to retained peers within schools. For these promoted students, exposure to a retained peer indicates the absence of this peer moving forward. Our initial approach follows Lavy, Paserman, and Schlosser (2012):

$$Y_{ist} = \gamma_0 + T_{st}\gamma_1 + \boldsymbol{X}_{it}\gamma_2 + Z_{st}\gamma_3 + \psi_s + \xi_t + \nu_{ist} \tag{1}$$

In Equation (1), $Y_{ist}$ is an outcome for promoted student $i$ in school $s$ who belongs to third-grade cohort $t$. We estimate the effects of retained peers on promoted students' standardized achievement in math and ELA, number of unexcused absences, and disciplinary incidents, which we measure with an indicator for whether student $i$ was suspended or expelled at least once. We estimate retained peer effects on promoted students in grades 4 through 7.

The treatment variable, $T_{st}$, captures the number of promoted-student $i$'s third-grade peers who are retained. In our preferred models we specify $T_{st}$ as a count variable (i.e., the number of retained peers); we also estimate models that specify $T_{st}$ as the fraction of peers who are retained. The vector $\boldsymbol{X}_{it}$ includes controls for student $i$'s racial/ethnic designation, sex, enrollment in the free and reduced-price lunch (FRL) program, English language learner (ELL) status, individualized education program (IEP) status, and third-grade (i.e., lagged) test scores, attendance, and disciplinary outcomes. $Z_{st}$ is a scalar variable that captures total enrollment in student $i$'s third grade cohort. We need this control because we specify $T_{st}$ as a count variable; it accounts for variation in cohort size correlated with $T_{st}$ (Fitzpatrick and Lovenheim, 2014). $\psi_s$ and $\xi_t$ are school and year fixed effects, respectively, and $\nu_{ist}$ is an idiosyncratic error, which we cluster at the school level.

Equation (1) leverages variation across cohorts within schools for identification. The identifying assumption is that cohort-to-cohort variation in the number of retained peers is exogenous within a school. Lavy, Paserman, and Schlosser (2012) argue this is a reasonable assumption, but identification could be compromised if there are cohort-level shocks that affect all students and the number of retained students. We alleviate this concern by expanding the model to isolate variation from the test-based discontinuity in Indiana. To do this, we add a control for the number of students in cohort *st* who score within a bandwidth around the retention threshold, then instrument for the number of retained peers with the number of students within the bandwidth who score below the cutoff. The instrumental variables (IV) model can be written as follows:

$$T_{ist} = \alpha_0 + BW_{st}\alpha_1 + LS_{st}\alpha_2 + X_{\mathrm{it}}\alpha_3 + Z_{st}\alpha_4 + \pi_s + \phi_t + \varepsilon_{ist} \tag{2}$$

$$Y_{ist} = \beta_0 + BW_{st}\beta_1 + \widehat{T}_{ist}\beta_2 + X_{\mathrm{it}}\beta_3 + Z_{st}\beta_4 + \delta_s + \tau_t + \eta_{ist} \tag{3}$$

Equation (2) is the first-stage model and equation (3) is the second-stage model. Repeated variables and analogous coefficients in equations (2) and (3) are as in equation (1).

The substantive difference between the model in equation (1), and the IV model described by equations (2) and (3), is the inclusion of two new variables in equation (2)—$BW_{st}$ and $LS_{st}$.[7] $BW_{st}$ measures the number of students within a bandwidth around the retention threshold, and $LS_{st}$ is the number of (low scoring) students below the threshold within the bandwidth. $LS_{st}$ is the excluded instrument. $BW_{st}$ controls for cohort-level shocks that push more students into the range of scores near the retention threshold, which would be missed by equation (1). The model identifies the effect of retained peers from variation in $LS_{st}$ conditional on $BW_{st}$.

[7] There is also a non-substantive difference: we add an *i* subscript to the retained-peers variable, $T_{ist}$, in equations (2) and (3). We make this adjustment because equation (2) technically allows $T_{ist}$ to be a function of the student-level control variables in $\boldsymbol{X_{it}}$. This adjustment is non-substantive because the IV model isolates variation in $T_{ist}$ from just the excluded instrument, $LS_{st}$, which only varies by cohort. Thus, like in equation (1), our estimates of retained peer effects in equation (3) are identified from within-school, cross-cohort variation.

The new variables in equation (2) essentially embed a standard, fuzzy regression discontinuity into the model, aggregated to the cohort level and without the use of a running variable. Cattaneo, Idrobo, and Titiunik (2019) describe RD estimation without a running variable as following the "local-randomization approach." Conceptually, this approach makes sense in the limit as the bandwidth around the discontinuity goes to zero, in which case there is no need to account for the running variable. The alternative to the local-randomization approach, and by far the more common approach in the literature, is what Cattaneo, Idrobo, and Titiunik (2019) describe as the "continuity-based approach." This approach relies on smooth functions of the running variable, approaching the discontinuity from both sides, to control for the relationship between the running variable and outcome leading into the discontinuity. The continuity-based approach is generally preferred because it is identified under weaker assumptions. Namely, unlike the local-randomization approach, it does not require the running variable functions to be flat on both sides of the discontinuity (i.e., that the coefficients on the running variable terms are zero).

Unfortunately, in our aggregated application, there is no clear way to construct a continuity-based RD analog. Thus, we use the local-randomization framework. This is a limitation conceptually. Additionally, because the test-score cutoff for retention in Indiana is well outside the thick part of the test distribution, it is infeasible to mitigate concerns about this approach by using a very tight bandwidth. However, while we acknowledge we cannot fully address concerns about the local-randomization RD approach, we provide two pieces of evidence suggesting it likely produces unbiased estimates in our setting. First, we conduct balancing tests analogous to those used in standard RD applications and show that our RD-based instrument balances observable covariates, including covariates that are highly predictive of student outcomes. Second, we show our findings are not sensitive to adding an additional control for the average test score of students who score within the bandwidth. When we add this control, it pushes our aggregated RD closer in concept to the standard, continuity-based RD approach.

In our preferred models we use a bandwidth of 25 test points to construct $BW_{st}$ and $LS_{st}$. The 25-point bandwidth is the same bandwidth used in Hwang and Koedel's (2025) standard RD analysis of the same policy. It is an average over outcomes and grades of the optimal bandwidth calculated following Calonico et al. (2020) and corresponds to just over half a standard deviation of the IREAD-3 test. That said, it is not obvious how to choose the "right" bandwidth in our aggregated RD application, and we confirm our findings are robust to a broad range of bandwidths from 15-45 points.

Finally, we estimate similar models for rising third graders who gain the retained students as peers. The treatment variable, $T_{st}$, and aggregated RD variables, $BW_{st}$ and $LS_{st}$, are identical to the variables we use for promoted students. The variable vector $\boldsymbol{X_{it}}$, and $Z_{st}$, are conceptually similar, but constructed based on the attributes of rising third graders and their cohorts. One small difference in our models for rising third graders is that we do not control for lagged achievement in $\boldsymbol{X_{it}}$ because like most states, standardized testing in Indiana does not begin until the third grade. But we do control for lagged disciplinary and attendance outcomes.

# 5 Results

## 5.1 Manipulation, Covariate Balance, and the First Stage

We begin by looking for evidence of manipulation in the underlying IREAD-3 test scores. Figure 1 shows a histogram of student scores on the summer test. Visually, the distribution is smooth through the retention cutoff at a score of 446. A formal density test following Cattaneo, Jansson, and Ma (2018) is consistent with the visual evidence, revealing no indication of running-variable manipulation.

Next, we conduct balancing tests like those used in the standard RD literature. For these tests, we pull the vector $\boldsymbol{X_{it}}$ from the right-hand side of equations (2) and (3), then use each element of $\boldsymbol{X_{it}}$ as a dependent variable in IV regressions that otherwise match our primary specification. If the aggregated RD is operating as intended, the number of retained peers, instrumented by the number of students who score below the cutoff, should not systematically predict the exogenous variables in $\boldsymbol{X_{it}}$.

Table 2 shows results from our balancing tests using the samples of rising fourth graders and rising third graders in the year immediately after the retention event. Each estimate in the table is from a separate regression. Only 3 of the 24 coefficients in the table are statistically different from zero. Moreover, it is notable that none of the coefficients from the lagged-outcome models in columns (10)-(13) suggest imbalance. In particular, the lagged math and ELA test scores for rising 4$^{th}$ graders are (a) highly differentiated and (b) highly predictive of outcomes for promoted fourth graders. Overall, the tests in Table 2 give no indication that the instrument is meaningfully imbalanced.

Table 3 shows results from first-stage regressions that predict the number of retained peers using the number of students who score below the retention threshold.[8] The first-stage coefficients are all well below 1.0. This is as expected based on Hwang and Koedel's (2025) earlier evaluation, which shows substantial policy non-compliance. In fact, our first stage coefficients around 0.30 are a close match to analogous first stage coefficients from their analysis. Also like in Hwang and Koedel (2025), our first stage is consistently strong across samples, with F statistics on the instrument between 78 to 86. These F statistics are within the range identified by Lee et al. (2022)

[8] We show first-stage results using the samples of rising fourth graders in the fourth grade, and rising third graders in the third grade. In the analysis below we also test for peer effects in later grades; the first stage coefficients are similar or larger in the later-grade samples and suppressed for brevity.

as requiring very little adjustment for significance testing—if anything, they imply our findings will very marginally overstate statistical significance.[9]

### 5.2 Results for Rising Fourth Graders

The first five columns of Table 4 show results from different versions of our model for rising fourth graders in the fourth grade. Column (1) documents the simple unconditional relationship between each outcome and the number of retained peers. Column (2) adds the student-level control variables and cohort enrollment ($\boldsymbol{X_{it}}$ and $Z_{st}$). With the addition of school and cohort fixed effects, column (3) shows results from equation (1) based on Lavy, Paserman, and Schlosser (2012), and column (4) shows output from our preferred IV model described by equations (2) and (3). Finally, column (5) extends our preferred model by adding the additional control for the average achievement of students within the bandwidth.[10] The results in columns (4) and (5) use a 25-point bandwidth to construct the variables $BW_{st}$ and $LS_{st}$; results using alternative bandwidths are substantively similar and shown in the appendix (Appendix Table A1).

Beginning in column (1), we show that unconditionally, there is a strong negative relationship between exposure to retained peers and the outcomes of rising fourth graders. Rising fourth graders with more retained peers have lower test scores in math and ELA, a higher likelihood of disciplinary incidents, and more unexcused absences, compared to their counterparts with fewer retained peers. This is unsurprising because exposure to more retained peers is likely indicative of a student's own disadvantage and the disadvantage of his or her school. The negative relationships in column (1) attenuate substantially in column (2) when we control for the rich set

[9] For instance, for a five percent test, Lee at al. (2022) show that with a first-stage F-statistic around 80, the typical second-stage standard error (which we report below) is about 98 percent (≈1/1.024) as large as the correct value.
[10] For cohorts with no students who score within the bandwidth, we impute the average bandwidth score to zero and include an indicator variable equal to one to indicate no students are within the bandwidth. Approximately 20 percent of students are in a school-by-year cohort where no students score within the bandwidth.

of observables in our data. In column (3), when we add school and cohort fixed effects, the peer effects turn positive for both achievement outcomes, become null for disciplinary outcomes, and attenuate but remain of the same sign for student attendance. That is, though small, the effects of peer retention on rising fourth graders begin to look more favorable.

Column (4) shows results from our preferred IV model. Relative to the model in column (3), the estimates for student achievement change little but become insignificant because we lose statistical power. In the model of student discipline, our estimate becomes more negative and is statistically significant, suggesting when rising fourth graders lose a peer due to retention, it leads to improved disciplinary outcomes. The disciplinary effect is about 25 percent of the baseline mean for promoted students (per Table 1), though in practical terms it is small. To give a sense of its magnitude, it implies that for the average cohort of 88 students in our sample, there is a 0.62 reduction in the number of students with a disciplinary incident in the fourth grade.

Column (5) shows that controlling for the average achievement level of students within the bandwidth has no substantive impact on our estimates. This reinforces the causal claim of our IV approach by showing that an additional control that provides more information about the attributes of students scoring within the bandwidth—i.e., the students who provide the identifying variation—does not affect our findings.

Next, columns (6) to (8) show estimates from our preferred model for three additional years, following the initial cohorts of rising $4^{th}$ graders into the $5^{th}$, $6^{th}$, and $7^{th}$ grades. This allows us to explore the possibility of peer effects that may emerge late, or in the case of our finding for student discipline, the persistence of the effect. Expectations about long-run peer effects are unclear. On the one hand, the limited evidence of short-run peer effects suggest long-run effects are less likely. Moreover, as students move into middle schools, typically in the $6^{th}$ grade in our sample, their elementary peer groups likely disperse, reducing the intensity of elementary-peer interactions. However, on the other hand, social dynamics change as children age, potentially altering how peer effects operate. Behavioral peer effects may be especially sensitive to the age at which they are assessed (Figlio, 2007).

All of that said, our findings provide no indication of persistent effects of losing a retained peer through the 7$^{th}$ grade. The immediate effect on student discipline fades out by the 5$^{th}$ grade and does not re-emerge. There are no peer effects on unexcused absences in any grade. While we find a negative and significant coefficient in the model of math achievement in the 6$^{th}$ grade, no other math or ELA coefficients are significant, and the ELA coefficients are inconsistent in sign across grades. We can offer no theoretical explanation for why there would be a peer effect on math achievement only, and in the 6$^{th}$ grade only, and thus we believe the most likely explanation for this coefficient is that it is spurious.

### 5.3 Results for Rising Third Graders

In Table 5 we follow the same approach to examine rising third graders. Like in Table 4, the unconditional relationships between gaining a retained peer and student outcomes are consistently negative, and they attenuate sharply in column (2) when we add control variables. When we condition on school and year fixed effects in column (3) the relationships largely disappear. In the IV models in column (4), we estimate null effects of gaining a retained peer on all four student outcomes. When we add the additional control for average achievement within the bandwidth in column (5), our estimates again change very little, though the small change in the ELA coefficient pushes it from marginally insignificant to marginally significant.

Like with rising fourth graders, we do not find evidence of persistent peer effects for rising third graders in columns (6) to (8). Interestingly, the coefficient in the math model in the third year after the retention event (the 5th grade for these cohorts) is significant and similar in magnitude to what we find in the third year after the retention event among rising fourth graders (the 6th grade). However, again, we believe the most likely explanation for this estimate is that it reflects a spurious correlation.[11]

### 5.4 Robustness and Effect Heterogeneity

#### 5.4.1 Robustness to How We Measure Retained-Peer Exposure

In Appendix Table A2 we replicate the results from our main models but define $T_{st}$ as a proportion of peers rather than using a count variable—i.e., $T_{st} \in [0,1]$. This allows the effect of the number of retained peers to vary depending on the size of the cohort (note we also test for effect heterogeneity by cohort size within our main framework below). The coefficients in the models that define $T_{st}$ as a proportion are much larger than in our count-based models because they capture the effect of a hypothetical shift from 0 to 100 percent retained peers. However, when rescaled to be comparable to our count-based estimates, they are substantively very similar to what we show above.

#### 5.4.2 Exploratory Evidence on Effect Heterogeneity

In this section we show results from tests for effect heterogeneity along several dimensions, with two caveats. First, our heterogeneity tests are generally not well-powered, which means we can only detect substantial heterogeneity. Second, for the most part, we do not find meaningful first-order peer effects in the analysis up to this point and true heterogeneity is usually smaller

[11] We do not have a definitive explanation, but this estimate—and its counterpart among rising 4th graders—is a clear outlier among our test-effect estimates in other grades and both subjects, and there is no *a priori* theoretical explanation for why there would be a peer effect in the third year after the retention and no other year. Also, the peer effects are in the same direction for rising 4th graders and rising 3rd graders, which is hard to explain. That is, taken at face value, the estimates imply very similar negative effects of retained peers, in the third year after the event, for cohorts that both lose and gain these students as peers. One possibility is a spurious correlation between retention events and something related to the math testing instrument, as these estimates align in calendar timing across cohorts. However, this is conjectural.

when there is little or no main effect (von Hippel and Schuetze, 2025). This suggests greater potential for false discovery in these tests. For these reasons, we interpret our heterogeneity findings cautiously and present them as exploratory only.

We begin by following on many prior papers that test for peer effect heterogeneity operating through achievement-level interactions. Evidence from Burke and Sass (2013) suggests the potential for substantial effect heterogeneity over the achievement distribution, whereas other research suggests effect heterogeneity is likely limited (e.g., Hanushek et al., 2003; Imberman et al., 2012; Sojourner, 2012). We test for achievement-level heterogeneity by dividing students into terciles based on the average of their third-grade math and ELA test scores. Then, we estimate equation (3) separately for rising fourth graders and rising third graders in each tercile.[12] Table 6 shows our results. The cross-tercile estimates are generally not statistically distinguishable from each other and the point estimates are inconsistent directionally—i.e., some estimates nominally imply top-tercile students are more affected than bottom-tercile students, and others imply the opposite. We conclude there is no evidence that retained students have heterogeneous peer effects across the achievement distribution.

Next, in Table 7 we test for effect heterogeneity across schools that differ by size. For these tests we divide schools into terciles based on the average enrollment in their third-grade cohorts. The average cohort sizes in the three terciles from largest to smallest are 115, 63, and 29. We motivate these tests by noting the possibility that the peer effects of retained students are more pronounced in smaller cohorts because exposure will be more intensive on a per-peer basis. However, we also acknowledge there are other possible sources of effect heterogeneity along this

[12] An alternative approach would be to add interactions with achievement terciles to the main model, but Feigenberg, Ost, and Qureshi (forthcoming) show our approach, which is mathematically equivalent to estimating a fully interacted model, is less likely to produce biased estimates.

dimension—e.g., differences in the schooling environments in larger and smaller schools may moderate the peer effects of retention. Regardless, we do not find consistent evidence of effect heterogeneity by school size. Inference is limited by the fact that our estimates are very imprecise in the bottom tercile, but the estimates for the largest and middle terciles exhibit no clear pattern of effect heterogeneity.

Finally, we test for effect heterogeneity based on students' demographic and socioeconomic attributes. Our data permit heterogeneity tests by gender, race-ethnicity, and students' FRL, ELL, and IEP designations. Along most demographic and socioeconomic dimensions there is not a strong theoretical reason to expect peer effect heterogeneity *ex ante*, and there are not strong or consistent patterns in the estimates to suggest effect heterogeneity. As such, we relegate most of these heterogeneity tests to the appendix (Appendix Tables A3 and A4). An exception, at least conditional on our main findings, is with regard to gender heterogeneity. That is, because boys are much more likely to have disciplinary incidents than girls, it is reasonable to hypothesize that our most pronounced finding—the peer effect on disciplinary outcomes among rising 4$^{th}$ graders—is driven by boys. Indeed, this is what our gender-heterogeneity analysis reveals, which we show in Table 8. The estimates in Table 8 also suggest girls rising into the 4$^{th}$ grade benefit in terms of ELA achievement from leaving behind a retained peer, though there is not a similar gendered effect in math. There is no evidence of gendered peer effects among rising third graders.[13,14]

[13] In results suppressed for brevity we also examine the persistence of the gendered peer effects. The disciplinary effect for boys fades out after the first year, which is consistent with the fade out of the main effect; the effect on ELA achievement for rising 4$^{th}$-grade girls persists through the fifth grade, then fades out by the sixth grade.

[14] Among the non-gender heterogeneity tests in the appendix, the strongest statistical evidence for heterogeneity is by IEP status among rising third graders. IEP students who are exposed to a newly retained peer perform better in math, suggestively better in ELA, and have fewer disciplinary incidents. *Ex post* we can articulate several potential explanations for these findings, but we do not believe there is good reason to expect these results *ex ante*, and choose not to highlight them at the risk of emphasizing false effect heterogeneity (von Hippel and Shuetze, 2025).

## 6 What if Peer Effects are Concentrated in Classrooms?

Our peer-effect estimates are small and most are not statistically significant. From the perspective of an individual student, we can rule out educationally meaningful peer effects from cohort-level exposure to retained students. This is true for both promoted students who lose the retained student as a peer, and students in the cohort below who gain the retained student as a peer. It is notable that we obtain these small effects in the context of the Indiana policy where the retention threshold is low. Relative to a policy with a higher retention threshold, the per-retention effects in Indiana should be larger because the average retained student is struggling more in school.

All of that said, a concern with this interpretation is that our estimates are averages across full cohorts and could mask larger effects for more proximal peers. For instance, suppose the cohort-wide effects are driven entirely by students who are in the same classroom as the retained peer, or in the case of rising fourth graders, students who *would have been* in the same classroom as the retained peer. If this were true, our estimates would include many untreated students and the impacts on affected students would be larger.

To explore the potential for larger, localized retained peer effects in more depth, we make *ad hoc* adjustments to our cohort-level estimates under the assumption that retained students only affect other students in the same classroom. Under this assumption, our cohort-level estimates can be viewed as intent-to-treat effects, which can be scaled up to recover classroom-level treatment effects by dividing by the inverse of the number of classrooms per cohort (or simply multiplying by the number of classrooms). This conversion assumes out-of-classroom peer effects in the same cohort are zero. It does not affect statistical significance; rather, it helps to convey the potential magnitude of peer effects if we assume they are concentrated within classrooms only.[15]

Table 9 shows point estimates from our preferred IV models, from Tables 4 and 5, after scaling them up to reflect classroom-level treatment effects. For rising fourth graders who lose a retained peer, the potential classroom-level effects on achievement implied by our point estimates remain small, suggesting increases of 0.01 to 0.03 standard deviations. Column (2) shows similarly sized decreases for rising third graders, from -0.01 to -0.04 standard deviations. For unexcused absences, the implied classroom-level effects are also small, at roughly four percent of the mean among promoted students (per Table 1). Our one statistically significant finding—for student discipline among rising 4th graders—scales up to a 1.9 percentage point reduction in the likelihood of a disciplinary incident if we assume the effect is concentrated among classroom-level peers. While this estimate is still small in an absolute sense, it is high relative to the baseline mean of about three percent (per Table 1). For rising third graders who gain the retained peer, the insignificant discipline estimate is of the same sign and similar in magnitude.

The results in Table 9 should not be overinterpreted because they are primarily based on insignificant cohort-level point estimates. Still, they provide a general sense of the potential scope for impact of retained peers. Again, these estimates are under the strong assumption that peer effects are zero outside of the retained student's classroom (or hypothetical classroom). If this assumption is violated, the estimates in Table 9 would be even smaller.

[15] For rising third graders, in principle we could directly estimate the classroom level effects, but this conversion is preferred because it avoids potential bias from the non-random placement of retained students into classrooms. For rising fourth graders, this conversion is the only way to calculate treatment effects at the classroom level because the (hypothetical) classroom placements of retained students are unobserved.

## 7 Conclusion

We study the peer effects of grade retention in the context of Indiana's third-grade retention policy. Our findings contribute to a thin literature on the peer effects of grade retention and make two novel contributions. First, we estimate the peer effects of grade retention comprehensively, in both the retained students' original and new cohorts. From the perspective of designing effective retention policies, it is important to understand the total impacts across all affected cohorts. Second, we are the first to estimate the peer effects of grade retention in a modern early-grade retention policy with a grade gate.

Our most pronounced finding is that rising fourth-grade students who leave behind a retained peer benefit from the retained peer's absence through a reduction in disciplinary incidents, concentrated among boys. There is also some evidence that rising fourth-grade girls have higher ELA achievement after one of their peers is retained. However, these effects fade out quickly, and we do not find consistent evidence of peer effects on other outcomes for promoted students.

We also find no evidence of peer effects among rising third graders. The use of the third grade as a promotional gate in Indiana likely increases the salience of the retention threat among these students when they gain a newly retained peer, more so than in the unstructured policy settings that have been studied previously (e.g., Bietenbeck, 2020; Gottfried, 2013a, 2013b). However, our findings give no indication that the threat effect leads to measurable improvements in the outcomes of rising third graders. We cannot entirely rule out the possibility of a positive threat effect, as it may exist but could be offset by the negative consequences of integrating retained peers into rising third-grade cohorts. However, our estimates imply that any positive peer effect operating through this channel, if present, is small.

Overall, we find Indiana's retention policy does not have educationally meaningful or persistent peer effects. It follows that the efficacy of the policy should be assessed based on how it affects retained students themselves.

**References**

Allen, C. S., Chen, Q., Willson, V. L., & Hughes, J. N. (2009). Quality of research design moderates effects of grade retention on achievement: A meta-analytic, multilevel analysis. *Educational Evaluation and Policy Analysis*, *31*(4), 480-499.

Anderson, G. E., Jimerson, S. R., & Whipple, A. D. (2005). Student ratings of stressful experiences at home and school: Loss of a parent and grade retention as superlative stressors. *Journal of Applied School Psychology*, *21*(1), 1-20.

Bietenbeck, J. (2020). The long-term impacts of low-achieving childhood peers: Evidence from Project STAR. *Journal of the European Economic Association*, 18(1), 392-426.

Burke, M.A., and Sass, T.R. (2013). Classroom peer effects and student achievement. *Journal of Labor Economics*, 31(1), 51-82.

Calonico, S., Cattaneo, M. D., & Farrell, M. H. (2020). Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs. *The Econometrics Journal*, *23*(2), 192-210.

Carrell, S.E., and Hoekstra, M. (2010). Externalities in the classroom: How children exposed to domestic violence affect everyone's kids. *American Economic Journal: Applied Economics*, 2(1), 211-28.

Cattaneo, M.D., Idrobo, N., and Titiunik, R. (2019). A practical introduction to regression discontinuity designs: Foundations. Working Paper. Retrieved 05.08.2024 at: https://arxiv.org/abs/1911.09511

Diaz, J., Grau, N., Reyes, T., and Rivera, J. (2021). The impact of grade retention on juvenile crime. *Economics of Education Review* 84, 1-16.

Epple, D., and Romano, R.E. (2011). Peer effects in education: A survey of the theory and evidence. *Handbook of Social Economics*, Vol. 1 (Editors Jess Benhabib, Alberto Bisin, and Matthew O. Jackson). Elsevier, Amsterdam, pp. 1053-1163.

Eren, O., Depew, B., & Barnes, S. (2017). Test-based promotion policies, dropping out, and juvenile crime. *Journal of Public Economics*, *153*, 9-31.

Eren, O., Lovenheim, M. F., & Mocan, H. N. (2022). The effect of grade retention on adult crime: Evidence from a test-based promotion policy. *Journal of Labor Economics*, *40*(2), 361-395.

Feigenberg, B., Ost, B., and Qureshi, J.A. (forthcoming). Omitted variables bias in interacted models: A cautionary tale. *Review of Economics and Statistics*.

Figlio, D. N. (2007). Boys named Sue: Disruptive children and their peers. *Education Finance and Policy*, 2(4), 376-394.

Figlio, D.N., Karbownik, K., and Ozek, U. (2023). Sibling spillovers may enhance the efficacy of targeted school policies. Working Paper No. 31406. Cambridge, MA: National Bureau of Economic Research.

Fletcher, J.M. (2009). The effects of inclusion on classmates of students with special needs: The case of serious emotional problems. *Education Finance and Policy*, 4(3), 278-99.

Gottfried, M. A. (2013a). The spillover effects of grade-retained classmates: Evidence from urban elementary schools. *American Journal of Education*, 119(3), 405-444.

Gottfried, M. A. (2013b). The spillover effects of grade-retained classmates: Evidence from urban elementary schools. *American Journal of Education*, 119(3), 405-444.

Greene, J. P., & Winters, M. A. (2007). Revisiting grade retention: An evaluation of Florida's test-based promotion policy. *Education Finance and Policy*, *2(4)*, 319-340.

Hanushek, E.A., Kain, J.F., Markman, J.M., and Rivkin, S.G. (2003). Does peer ability affect student achievement? *Journal of Applied Econometrics*, 18, 527-544.

Hill, A. (2014). The costs of failure: Negative externalities in high school course repetition. *Economics of Education Review*, 43, 91-105.

Hwang, N., and Koedel, C. (2025). Helping or hurting: The effects of retention in the third grade on student outcomes. *Educational Evaluation and Policy Analysis*, 47(1), 65-88.

Imberman, S.A., Kugler, A.D., and Sacerdote, B.I. (2012). Katrina's children: Evidence on the structure of peer effects from hurricane evacuees. *American Economic Review*, 102(5), 2048-82.

Jacob, B. A., & Lefgren, L. (2004). Remedial education and student achievement: A regression-discontinuity analysis. *Review of Economics and Statistics*, *86*(1), 226-244.

Jacob, B. A., & Lefgren, L. (2009). The effect of grade retention on high school completion. *American Economic Journal: Applied Economics 1*(3), 33-58.

Kristoffersen, J.H., Kraegpoth, M.V., Nielson, H.S., and Simonsen, M. (2015). Disruptive school peers and student outcomes. *Economics of Education Review*, 45, 1-13.

Lavy, V., Paserman, M.D., and Schlosser, A. (2012). Inside the black box of ability peer effects: Evidence from variation in the proportion of low achievers in the classroom. *The Economic Journal*, 122, 208-237.

Lee, D.S., McCrary, J., Moreira, M.J., and Porter, J. (2022). Valid t-ratio Inference for IV. *American Economic Review*, 112(10), 3260-3290.

Lefgren, L. (2004). Educational peer effects and the Chicago public schools. *Journal of Urban Economics*, 56, 169-91.

Manacorda, M. (2012). The cost of grade retention. *Review of Economics and Statistics*, *94*(2), 596-606.

Mariano, L.T., Martorell, P., and Berglund, T. (forthcoming). The effects of grade retention on high school outcomes: Evidence from New York City Schools. *Journal of Research on Educational Effectiveness*.

Martorell, P., & Mariano, L. T. (2018). The causal effects of grade retention on behavioral outcomes. *Journal of Research on Educational Effectiveness*, *11*(2), 192-216.

Ou, S. R., & Reynolds, A. J. (2010). Grade retention, postsecondary education, and public aid receipt. *Educational Evaluation and Policy Analysis*, *32*(1), 118-139.

Özek, U. (2015). Hold back to move forward? Early grade retention and student misbehavior. *Education Finance and Policy*, *10*(3), 350-377.

Özek, U., and Mariano, L.T. (2023). Think again: Is grade retention bad for kids? Policy Report. Thomas B. Fordham Institute.

Roderick, M., & Nagaoka, J. (2005). Retention under Chicago’s high-stakes testing program: Helpful, harmful, or harmless? *Educational Evaluation and Policy Analysis*, *27*(4), 309-340.

Sacerdote, B. (2011). Peer effects in education: How might they work, how big Are they and how much do we know thus far? *Handbook of the Economics of Education*, Vol. 3 (editors Eric A. Hanushek, Stephen Machin, and Ludger Woessmann). Elsevier, Amsterdam, pp. 249–277.

Schwerdt, G., West, M.R., and Winters, M.A. (2017). The effects of test-based retention on student outcomes over time: Regression discontinuity evidence from Florida. *Journal of Public Economics,* 152, 154-169.

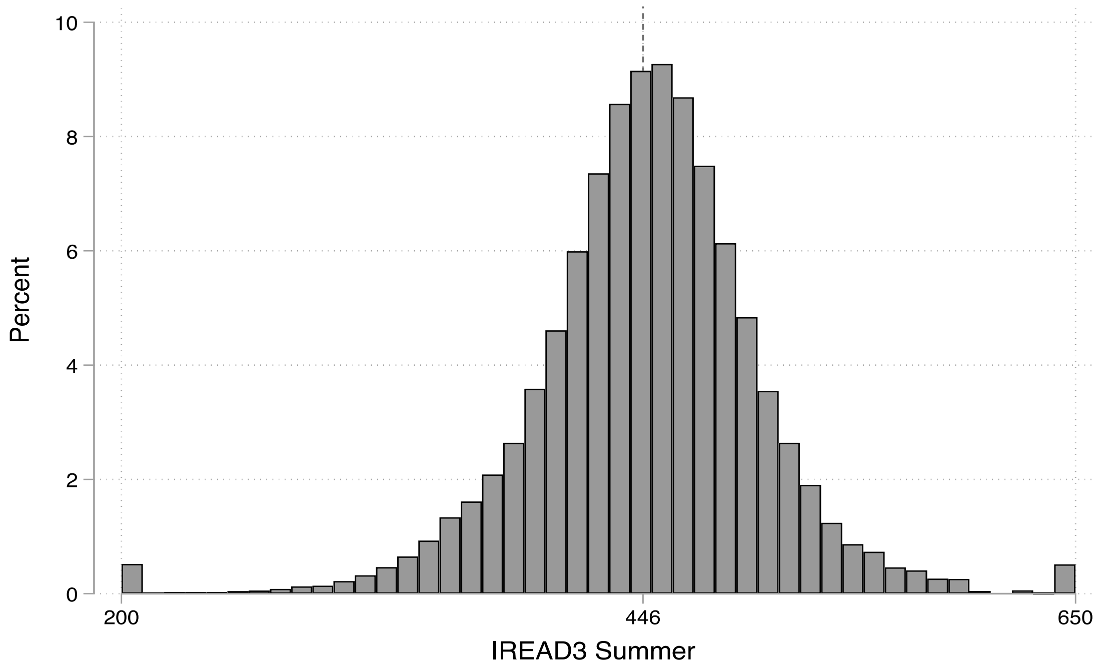
Sojourner, A. (2012). Identification of peer effects with missing peer data: Evidence from Project STAR. *The Economic Journal*, 123, 574–605.

von Hippel, P. and Schuetze, B.A. (2025). How not to fool ourselves about heterogeneity of treatment effects. EdWorking Paper No. 25-1116. Annenberg Institute at Brown University.

Westall, J., and Cummings, A. (2023). The effects of early literacy policies on student achievement. Working Paper. Education Policy Innovation Collaborative.

Winters, M. A., and Greene, J. P. (2012). The medium-run effects of Florida's test-based promotion policy. *Education Finance and Policy*, *7*(3), 305-330.

Figure 1. Density of the IREAD-3 Summer Test Around the Retention Cutoff in the Student Level Data.




Notes: This figure is taken from Hwang and Koedel (2025). Recall the sample is conditional on failing the spring test. We fail to reject the null hypothesis that the distribution is smooth through the cutoff (p-value 0.91).

Table 1. Descriptive Statistics.

| | Promoted (Analytic Sample) | | Retained | |
|---|---|---|---|---|
| <u>Student Characteristics</u> | Mean | SD | Mean | SD |
| Female | 0.49 | | 0.46 | |
| Asian | 0.02 | | 0.004 | |
| Black | 0.10 | | 0.35 | |
| Hispanic | 0.11 | | 0.12 | |
| White | 0.72 | | 0.46 | |
| Other race/ethnicity | 0.07 | | 0.06 | |
| Free/reduced-price lunch enrolled (FRL) | 0.50 | | 0.82 | |
| English language learner (ELL) | 0.08 | | 0.07 | |
| Individualized education program (IEP) | 0.13 | | 0.22 | |
| 3rd Grade ELA test (ISTEP+) | 0.05 | 0.97 | -1.42 | 0.72 |
| 3rd Grade Math test (ISTEP+) | 0.05 | 0.97 | -1.30 | 0.82 |
| 3rd Grade Discipline (at least one incident) | 0.03 | | 0.11 | |
| 3rd Grade Days Unexcused Absence | 1.72 | 3.20 | 3.61 | 5.09 |
| *N* (unique students) | 342,822 | | 7,348 | |
| | | | | |
| <u>Third Grade Cohort Characteristics</u> | All Third Graders | | | |
| Retained Peers (N) | 1.35 | 2.22 | -- | -- |
| Peers +/- 25 Bandwidth | 4.10 | 4.23 | -- | -- |
| Cohort Size (Enrollment) | 88.08 | 39.30 | -- | -- |
| N (unique third-grade cohorts) | 5158 | | -- | -- |

Notes: For students who are retained, third-grade outcomes are reported from the first third-grade year. Summary statistics for promoted students are for rising fourth graders. Note the average test scores for retained students reported in this table are lower than what is reported in Hwang and Koedel (2025). This is due to a reporting error in their paper (they reported the average of the second third-grade test for retained students, after retention, rather than the average of the initial third grade test). We confirmed this reporting error in their paper did not otherwise influence their analysis or findings.

Table 2. Tests of Covariate Balance.

| Variable | (1) Female | (2) FRL | (3) ELL | (4) IEP | (5) Black | (6) Hispanic | (7) Asian | (8) Other Race | (9) White | (10) Math (Grade 3) | (11) ELA (Grade 3) | (12) Discipline (Grade 3) | (13) Absence (Grade 3) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Peer Retention for Rising 4th graders | 0.002 | -0.002 | 0.001 | 0.001 | -0.001 | -0.002 | -0.000 | -0.003 | 0.005* | -0.003 | -0.007 | 0.000 | 0.025 |
| | (0.003) | (0.003) | (0.002) | (0.002) | (0.002) | (0.002) | (0.001) | (0.002) | (0.003) | (0.010) | (0.009) | (0.002) | (0.046) |
| *N* | 342822 | 342822 | 342822 | 342822 | 342822 | 342822 | 342822 | 342822 | 342822 | 342822 | 342822 | 342822 | 342822 |
| | Female | FRL | ELL | IEP | Black | Hispanic | Asian | Other Race | White | | | Discipline (Grade 2) | Absence (Grade 2) |
| Peer Retention for Rising 3rd graders | -0.002 | -0.000 | 0.006* | -0.002 | -0.002 | 0.003 | -0.000 | 0.003 | -0.005* | | | -0.001 | 0.000 |
| | (0.003) | (0.003) | (0.003) | (0.002) | (0.002) | (0.002) | (0.001) | (0.002) | (0.002) | | | (0.002) | (0.043) |
| *N* | 347468 | 347468 | 347468 | 347468 | 347468 | 347468 | 347468 | 347468 | 347468 | | | 347468 | 347468 |

Notes: ELL: English Language Learner. FRL: Enrolled in Free or Reduced-Price Lunch. IEP: Has Individual Education Program. ELA: English Language Arts. The coefficients in this table are from instrumental-variables models that estimate the "effect" of an additional retained peer on variables that should not be affected if our discontinuity-based instrument is exogenous. Thus, the expectation is that the coefficients are zero. All models control for school fixed effects, year fixed effects, school-by-grade cohort enrollment, and the number of peers who score within a bandwidth of 25 points around the test cutoff. Standard errors clustered at the school level are in parentheses.

* $p < 0.05$

Table 3. First-Stage Estimates for Rising 4th Graders and Rising 3rd graders for the Samples in the Year Immediately Following the Retention Event.

| | (1)<br>Rising 4th graders | (2)<br>Rising 3rd graders |
|---|---|---|
| Math | | |
| Low-Scoring Peers Within the Bandwidth | 0.310** | 0.305** |
| | (0.002) | (0.033) |
| F statistic | 77.90 | 86.26 |
| *N* | 342822 | 347468 |
| | | |
| ELA | | |
| Low-Scoring Peers Within the Bandwidth | 0.310** | 0.303** |
| | (0.002) | (0.033) |
| F statistic | 77.83 | 85.32 |
| *N* | 342592 | 346240 |
| | | |
| Discipline | | |
| Low-Scoring Peers Within the Bandwidth | 0.311** | 0.306** |
| | (0.002) | (0.033) |
| F statistic | 77.22 | 84.92 |
| *N* | 347329 | 352020 |
| | | |
| Days Absent | | |
| Low-Scoring Peers Within the Bandwidth | 0.310** | 0.306** |
| | (0.002) | (0.033) |
| F statistic | 87.70 | 84.87 |
| *N* | 347177 | 351916 |

Notes: ELA: English language arts. The first-stage estimates for rising fourth graders and rising third graders using the samples of students in later grades are similar to what we report here and suppressed for brevity. Standard errors clustered at the school level are in parentheses.
** $p < 0.01$

Table 4. The Effects of Peer Retention on Rising Fourth Graders, Who Lose a Peer, in the Fourth and Later Grades.

| | | | Grade 4 | | | Grade 5 | Grade 6 | Grade 7 |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Math | | | | | | | | |
| Peer Retention | -0.060** | -0.002 | 0.005* | 0.004 | 0.003 | -0.004 | -0.023** | -0.004 |
| | (0.004) | (0.002) | (0.002) | (0.008) | (0.008) | (0.008) | (0.008) | (0.010) |
| *N* | 342822 | 342822 | 342822 | 342822 | 342822 | 263915 | 190434 | 123381 |
| ELA | | | | | | | | |
| Peer Retention | -0.057** | 0.000 | 0.006** | 0.010 | 0.008 | 0.010 | -0.005 | 0.004 |
| | (0.004) | (0.001) | (0.001) | (0.006) | (0.006) | (0.007) | (0.007) | (0.008) |
| *N* | 342592 | 342592 | 342592 | 342592 | 342592 | 263792 | 190637 | 123965 |
| Discipline | | | | | | | | |
| Peer Retention | 0.007** | 0.001** | -0.001 | -0.007** | -0.007** | -0.003 | 0.001 | -0.004 |
| | (0.001) | (0.000) | (0.000) | (0.002) | (0.002) | (0.002) | (0.003) | (0.004) |
| *N* | 347329 | 347329 | 347329 | 347329 | 347329 | 267188 | 192770 | 125591 |
| Days Absent | | | | | | | | |
| Peer Retention | 0.205** | 0.042** | 0.010 | 0.026 | 0.027 | -0.028 | -0.076 | -0.012 |
| | (0.021) | (0.008) | (0.009) | (0.037) | (0.036) | (0.049) | (0.048) | (0.076) |
| *N* | 347177 | 347177 | 347177 | 347177 | 347177 | 267133 | 192380 | 125577 |
| Student Controls & Enrollment | | X | X | X | X | X | X | X |
| School and Year Fixed Effects | | | X | X | X | X | X | X |
| Bandwidth Control | | | | X | X | X | X | X |
| IV | | | | X | X | X | X | X |
| Average Achievement in Bandwidth | | | | | X | | | |

Notes: ELA: English language arts. Our samples shrink as we look forward into later grades because we can track fewer cohorts into these grades with our data panel. Standard errors clustered at the school level are in parentheses.
* $p < 0.05$, ** $p < 0.01$

Table 5. The Effects of Peer Retention on Rising Third Graders, Who Gain a Peer, in the Third and Later Grades

| | | | Grade 3 | | | Grade 4 | Grade 5 | Grade 6 |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Math | | | | | | | | |
| Peer Retention | -0.068** | -0.026** | 0.005* | -0.003 | -0.008 | -0.002 | -0.028* | -0.014 |
| | (0.004) | (0.004) | (0.002) | (0.009) | (0.009) | (0.009) | (0.013) | (0.012) |
| *N* | 347468 | 347468 | 347468 | 347468 | 347468 | 262929 | 188881 | 118972 |
| ELA | | | | | | | | |
| Peer Retention | -0.066** | -0.028** | 0.001 | -0.010 | -0.014* | -0.009 | -0.011 | 0.001 |
| | (0.004) | (0.002) | (0.002) | (0.007) | (0.007) | (0.008) | (0.009) | (0.011) |
| *N* | 346240 | 346240 | 346240 | 346240 | 346240 | 262516 | 188822 | 119281 |
| Discipline | | | | | | | | |
| Peer Retention | 0.006** | 0.002** | -0.001 | -0.004 | -0.003 | -0.001 | 0.006* | -0.007 |
| | (0.001) | (0.001) | (0.001) | (0.002) | (0.002) | (0.002) | (0.003) | (0.004) |
| *N* | 352020 | 352020 | 352020 | 352020 | 352020 | 265787 | 190660 | 120526 |
| Days Absent | | | | | | | | |
| Peer Retention | 0.232** | 0.068** | 0.017 | 0.023 | 0.026 | -0.039 | -0.034 | -0.112 |
| | (0.029) | (0.014) | (0.014) | (0.049) | (0.049) | (0.049) | (0.052) | (0.069) |
| *N* | 351916 | 351916 | 351916 | 351916 | 351916 | 265701 | 190637 | 120185 |
| Student Controls & Enrollment | | X | X | X | X | X | X | X |
| School and Year Fixed Effects | | | X | X | X | X | X | X |
| Bandwidth Control | | | | X | X | X | X | X |
| IV | | | | X | X | X | X | X |
| Average Achievement in Bandwidth | | | | | X | | | |

Notes: ELA: English language arts. Our samples shrink as we look forward into later grades because we can track fewer cohorts into these grades with our data panel. Standard errors clustered at the school level are in parentheses.
$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$

Table 6. Effect Heterogeneity by the Achievement Level of Peers.

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Rising 4th Graders in the 4th Grade | Math | ELA | Discipline | Days Absent |
| Achievement Tercile (Highest) | 0.003 | 0.014 | -0.005 | 0.029 |
| Peer Retention | (0.015) | (0.013) | (0.003) | (0.048) |
| *N* | 118905 | 118899 | 119436 | 119409 |
| | | | | |
| Achievement Tercile (Middle) | | | | |
| Peer Retention | -0.001 | 0.015 | -0.009** | 0.012 |
| | (0.009) | (0.008) | (0.003) | (0.045) |
| *N* | 119460 | 119418 | 120342 | 120295 |
| Achievement Tercile (Bottom) | | | | |
| Peer Retention | 0.008 | 0.004 | -0.007* | 0.032 |
| | (0.008) | (0.007) | (0.003) | (0.046) |
| *N* | 104457 | 104275 | 107551 | 107473 |
| | | | | |
| Rising 3rd Graders in the 3rd Grade | | | | |
| Achievement Tercile (Highest) | | | | |
| Peer Retention | -0.014 | -0.000 | 0.003 | -0.072 |
| | (0.012) | (0.011) | (0.002) | (0.052) |
| *N* | 119455 | 119305 | 119638 | 119611 |
| Achievement Tercile (Middle) | | | | |
| Peer Retention | 0.002 | -0.003 | -0.002 | 0.019 |
| | (0.006) | (0.006) | (0.002) | (0.047) |
| *N* | 116131 | 115856 | 116477 | 116447 |
| Achievement Tercile (Bottom) | | | | |
| Peer Retention | 0.005 | -0.010 | -0.007* | 0.059 |
| | (0.006) | (0.007) | (0.003) | (0.068) |
| *N* | 111882 | 111079 | 114800 | 114756 |
| | | | | |
| Student Controls & Enrollment | X | X | X | X |
| School and Year FE | X | X | X | X |
| BW Control | X | X | X | X |
| IV | X | X | X | X |

Notes: ELA: English language arts. Results are from our preferred instrumental variables specification, estimated on subsamples of students as indicated by the rows. Standard errors clustered at the school level are in parentheses.

Table 7. Effect Heterogeneity by School-Average Cohort Size.

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Rising 4th Graders in the 4th Grade | Math | ELA | Discipline | Days Absent |
| Enrollment Tercile (Highest) | | | | |
| Peer Retention | 0.000 | 0.007 | -0.010** | 0.024 |
| | (0.011) | (0.009) | (0.003) | (0.048) |
| *N* | 194056 | 193926 | 196283 | 196205 |
| Enrollment Tercile (Middle) | | | | |
| Peer Retention | 0.019 | 0.024 | -0.004 | 0.015 |
| | (0.014) | (0.013) | (0.004) | (0.072) |
| *N* | 110404 | 110300 | 112019 | 111969 |
| Enrollment Tercile (Bottom) | | | | |
| Peer Retention | -0.092 | -0.208 | -0.021 | -0.454 |
| | (0.138) | (0.168) | (0.040) | (0.591) |
| *N* | 38362 | 38366 | 39027 | 39003 |
| | | | | |
| Rising 3rd Graders in the 3rd Grade | | | | |
| Enrollment Tercile (Highest) | | | | |
| Peer Retention | -0.011 | -0.007 | -0.004 | -0.002 |
| | (0.012) | (0.010) | (0.003) | (0.061) |
| *N* | 200581 | 199968 | 202949 | 202894 |
| Enrollment Tercile (Middle) | | | | |
| Peer Retention | 0.018 | -0.019 | -0.001 | 0.136 |
| | (0.017) | (0.014) | (0.004) | (0.075) |
| *N* | 113483 | 112963 | 115139 | 115112 |
| Enrollment Tercile (Bottom) | | | | |
| Peer Retention | 0.069 | -0.005 | 0.018 | -0.554 |
| | (0.119) | (0.094) | (0.027) | (0.437) |
| *N* | 33404 | 33309 | 33932 | 33910 |
| | | | | |
| Student Controls & Enrollment | X | X | X | X |
| School and Year FE | X | X | X | X |
| BW Control | X | X | X | X |
| IV | X | X | X | X |

Notes: ELA: English language arts. Results are from our preferred instrumental variables specification, estimated on subsamples of students as indicated by the rows. Standard errors clustered at the school level are in parentheses.

Table 8. Effect Heterogeneity by Peer Gender.

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Rising 4th Graders in the 4th Grade | Math | ELA | Discipline | Days Absent |
| Female | | | | |
| Peer Retention | 0.009 | 0.021** | -0.003 | 0.018 |
| | (0.009) | (0.007) | (0.002) | (0.042) |
| *N* | 169245 | 169329 | 171205 | 171123 |
| Male | | | | |
| Peer Retention | -0.002 | -0.002 | -0.011** | 0.034 |
| | (0.008) | (0.007) | (0.003) | (0.042) |
| *N* | 173577 | 173263 | 176124 | 176054 |
| | | | | |
| Rising 3rd Graders in the 3rd Grade | | | | |
| Female | | | | |
| Peer Retention | -0.005 | -0.012 | -0.002 | 0.039 |
| | (0.011) | (0.009) | (0.002) | (0.051) |
| *N* | 170395 | 170183 | 172462 | 172410 |
| Male | | | | |
| Peer Retention | -0.000 | -0.008 | -0.005 | 0.010 |
| | (0.011) | (0.009) | (0.003) | (0.055) |
| *N* | 177073 | 176057 | 179558 | 179506 |
| | | | | |
| Student Controls & Enrollment | X | X | X | X |
| School and Year FE | X | X | X | X |
| BW Control | X | X | X | X |
| IV | X | X | X | X |

Notes: ELA: English language arts. Results are from our preferred instrumental variables specification, estimated on subsamples of students as indicated by the rows. Standard errors clustered at the school level are in parentheses.

Table 9. Implied Classroom-Level Peer Effects if We Assume Retained Students' Peer Effects Are Entirely Concentrated Within their Classrooms (or Potential Classrooms).

| | Rising 4th Graders in the 4th Grade | Rising 3rd Graders in the 3rd Grade |
|---|---|---|
| Math Achievement | 0.011 | -0.011 |
| ELA Achievement | 0.027 | -0.038 |
| Discipline | -0.019** | -0.015 |
| Unexcused Absence | 0.070 | 0.087 |

Notes: If we assume retained peers only affect other students in their own classroom, our cohort-wide estimates are averaged over many untreated students (i.e., students in the same original or new cohort as the retained peer, but not the same classroom). Here we estimate the classroom-level peer effects implied by our cohort-wide estimates under the assumption that the peer effects only happen within the same classroom by multiplying the point estimates from our preferred model by the average number of classrooms per grade in our sample ($4^{th}$ grade: 2.7; $3^{rd}$ grad: 3.8). This is equivalent to parameterizing the out-of-classroom peer effect within cohorts to zero. See discussion in the main text for more information. This conversion does not affect statistical significance—statistical significance indicators in this table are carried over from Tables 4 and 5. $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$

# Appendix
# Supplementary Tables

Appendix Table A1. The Effects of Peer Retention with Different Bandwidths (15, 35, and 45 test points)

| | Math | | | ELA | | | Discipline | | | Absence | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Rising Fourth Graders in the 4th Grade | | | | | | | | | | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
| | BW15 | BW35 | BW45 | BW15 | BW35 | BW45 | BW15 | BW35 | BW45 | BW15 | BW35 | BW45 |
| Peer Retention | 0.007 | 0.003 | 0.005 | 0.011 | 0.006 | 0.006 | -0.007** | -0.007** | -0.006** | 0.013 | 0.024 | 0.016 |
| | (0.009) | (0.007) | (0.007) | (0.008) | (0.006) | (0.005) | (0.003) | (0.002) | (0.002) | (0.041) | (0.036) | (0.032) |
| | 342822 | 342822 | 342822 | 342592 | 342592 | 342592 | 347329 | 347329 | 347329 | 347177 | 347177 | 347177 |
| Rising Third Graders in the 3rd Grade | | | | | | | | | | | | |
| | Math | | | ELA | | | Discipline | | | Absence | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
| | BW15 | BW35 | BW45 | BW15 | BW35 | BW45 | BW15 | BW35 | BW45 | BW15 | BW35 | BW45 |
| Peer Retention | 0.003 | -0.001 | -0.003 | -0.014 | -0.010 | -0.007 | -0.003 | -0.004* | -0.003* | 0.025 | 0.043 | 0.036 |
| | (0.011) | (0.009) | (0.008) | (0.009) | (0.007) | (0.006) | (0.002) | (0.002) | (0.002) | (0.051) | (0.046) | (0.042) |
| | 347468 | 347468 | 347468 | 346240 | 346240 | 346240 | 352020 | 352020 | 352020 | 351916 | 351916 | 351916 |

Notes: BW = bandwidth. ELA: English language arts. The results in the main text use a bandwidth of 25 test points. Standard errors clustered at the school level are in parentheses. * $p < 0.05$, ** $p < 0.01$

Appendix Table A2. Models that Measure Peer Exposure by the Proportion of Retained Peers.

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Rising 4th Graders in the 4th Grade | Math | ELA | Discipline | Days Absent |
| In the 4th Grade | | | | |
| Peer Retention ($T_{st} \in [0,1]$) | 0.350 | 0.793 | -0.443* | 3.098 |
| | (0.595) | (0.498) | (0.177) | (3.172) |
| *N* | 342822 | 342592 | 347329 | 347177 |
| | | | | |
| Rising 3rd Graders in the 3rd Grade | | | | |
| In the 3rd Grade | | | | |
| Peer Retention ($T_{st} \in [0,1]$) | 0.154 | -0.726 | -0.294 | 2.624 |
| | (0.735) | (0.600) | (0.183) | (4.007) |
| *N* | 347468 | 346240 | 352020 | 351916 |
| | | | | |
| Student Controls & Enrollment | X | X | X | X |
| School and Year FE | X | X | X | X |
| BW Control | X | X | X | X |
| IV | X | X | X | X |

Notes: These results are comparable to the results in Tables 4 and 5 in the main text, except here, retention is measured as a proportion of peers rather than the number of retained peers. These coefficients reflect the hypothetical effect of going from 0 to 100 percent retained peers. To make a rough comparison of these estimates to the estimates in the main text, note that a single retention corresponds to about 1.1 percent of the average school-by-grade cohort; thus, multiplying these estimates by 0.011 scales them to the same level as the per-retention estimates in the main text. Standard errors clustered at the school level are in parentheses. * $p < 0.05$, ** $p < 0.01$

Appendix Table A3. Peer Effect Heterogeneity for Rising 4th Graders in the 4th Grade Along Various Demographic and Socioeconomic Dimensions Not Covered in the Main Text

| | (1) Math | (2) ELA | (3) Discipline | (4) Days Absent |
|---|---|---|---|---|
| Black | | | | |
| Peer Retention | 0.006 | -0.001 | -0.009 | 0.077 |
| | (0.008) | (0.008) | (0.005) | (0.067) |
| *N* | 34314 | 34260 | 35101 | 35044 |
| Hispanic | | | | |
| Peer Retention | -0.004 | 0.014 | -0.007* | 0.079 |
| | (0.013) | (0.011) | (0.003) | (0.061) |
| *N* | 38838 | 38857 | 39414 | 39402 |
| White | | | | |
| Peer Retention | 0.001 | 0.014 | -0.006* | -0.047 |
| | (0.013) | (0.010) | (0.003) | (0.052) |
| *N* | 246443 | 246276 | 249289 | 249216 |
| FRL | | | | |
| Peer Retention | 0.005 | 0.006 | -0.008** | 0.024 |
| | (0.007) | (0.006) | (0.003) | (0.042) |
| *N* | 171486 | 171282 | 174693 | 174592 |
| Non-FRL | | | | |
| Peer Retention | -0.006 | 0.016 | -0.006* | 0.014 |
| | (0.016) | (0.013) | (0.002) | (0.051) |
| *N* | 171336 | 171310 | 172636 | 172585 |
| ELL | | | | |
| Peer Retention | -0.020 | 0.002 | -0.006 | 0.043 |
| | (0.016) | (0.011) | (0.004) | (0.062) |
| *N* | 27922 | 27945 | 28314 | 28308 |
| Non-ELL | | | | |
| Peer Retention | 0.007 | 0.011 | -0.007** | 0.023 |
| | (0.008) | (0.007) | (0.002) | (0.039) |
| *N* | 314900 | 314647 | 319015 | 318869 |
| IEP | | | | |
| Peer Retention | -0.002 | 0.001 | -0.004 | 0.032 |
| | (0.014) | (0.011) | (0.005) | (0.085) |
| *N* | 45360 | 45131 | 47159 | 47137 |
| Non-IEP | | | | |
| Peer Retention | 0.004 | 0.011 | -0.007** | 0.026 |
| | (0.008) | (0.007) | (0.002) | (0.036) |
| *N* | 297462 | 297461 | 300170 | 300040 |
| | | | | |
| Student Controls & Enrollment | X | X | X | X |
| School and Year FE | X | X | X | X |
| BW Control | X | X | X | X |
| IV | X | X | X | X |

Notes: ELL: English Language Learner. FRL: Enrolled in Free or Reduced Price Lunch. IEP: Has Individual Education Program. Standard errors clustered at the school level are in parentheses. * $p < 0.05$, ** $p < 0.01$

Appendix Table A4. Peer Effect Heterogeneity for Rising 3rd Graders in the 3rd Grade Along Various Demographic and Socioeconomic Dimensions Not Covered in the Main Text.

| | (1) Math | (2) ELA | (3) Discipline | (4) Absence |
|---|---|---|---|---|
| Black | | | | |
| Peer Retention | 0.008 | -0.010 | -0.008 | 0.209 |
| | (0.015) | (0.011) | (0.005) | (0.123) |
| *N* | 36954 | 36719 | 38089 | 38054 |
| Hispanic | | | | |
| Peer Retention | -0.002 | -0.028 | -0.003 | 0.044 |
| | (0.017) | (0.016) | (0.002) | (0.071) |
| *N* | 41231 | 41120 | 41818 | 41807 |
| White | | | | |
| Peer Retention | -0.011 | -0.000 | -0.001 | -0.124* |
| | (0.014) | (0.012) | (0.002) | (0.054) |
| *N* | 244782 | 243978 | 247290 | 247238 |
| FRL | | | | |
| Peer Retention | 0.007 | -0.006 | -0.005 | 0.050 |
| | (0.009) | (0.008) | (0.003) | (0.055) |
| *N* | 179241 | 178329 | 182603 | 182530 |
| Non-FRL | | | | |
| Peer Retention | -0.030 | -0.019 | -0.000 | -0.077 |
| | (0.019) | (0.015) | (0.002) | (0.058) |
| *N* | 168227 | 167911 | 169417 | 169386 |
| ELL | | | | |
| Peer Retention | -0.016 | -0.037 | -0.003 | 0.035 |
| | (0.019) | (0.020) | (0.002) | (0.078) |
| *N* | 28540 | 28458 | 28882 | 288875 |
| Non-ELL | | | | |
| Peer Retention | -0.005 | -0.010 | -0.004 | 0.025 |
| | (0.010) | (0.008) | (0.002) | (0.052) |
| *N* | 318928 | 317782 | 323138 | 323041 |
| IEP | | | | |
| Peer Retention | 0.046* | 0.013 | -0.014** | 0.036 |
| | (0.020) | (0.019) | (0.005) | (0.085) |
| *N* | 50163 | 49524 | 51194 | 51183 |
| Non-IEP | | | | |
| Peer Retention | -0.010 | -0.014 | -0.002 | 0.023 |
| | (0.010) | (0.008) | (0.002) | (0.049) |
| *N* | 297305 | 296716 | 300826 | 300733 |
| | | | | |
| Student Controls & Enrollment | X | X | X | X |
| School and Year FE | X | X | X | X |
| BW Control | X | X | X | X |
| IV | X | X | X | X |

Notes: ELL: English Language Learner. FRL: Enrolled in Free or Reduced Price Lunch. IEP: Has Individual Education Program. Standard errors clustered at the school level are in parentheses. * $p < 0.05$, ** $p < 0.01$