

Draft for Presentation at the American Economic Association Annual Meeting
January, 2010

Comments Appreciated

The Influence of Peers on Student Proficiency in Elementary School

Marcus A. Winters
Manhattan Institute for Policy Research
mwinters@manhattan-institute.org
(Phone) 646-839-3354

Abstract

I use an instrumental variables approach to study the influence of the quality of peers on the academic proficiency of elementary school students. The instrument takes advantage of exogenous across-cohort variation in the quality of peers in Florida elementary schools that resulted from the implementation of a test-based promotion policy for third grade students. I find significant peer effects in both math and reading. Peer effects are relatively linear in the student's prior math achievement and are non-linear in prior reading achievement such that the highest performing reading students are most greatly affected by the quality of their peers. The results suggest that sorting policies can create a more equitable distribution of learning but such policies are not Pareto-improving.

I) Introduction

Social scientists across several disciplines have an interest in understanding the influence of peers on an individual's behavior and outcomes. Economists in particular are interested in peer effects because it is likely they are not completely internalized. From a public policy perspective, if peer effects are non-linear across individuals of varying characteristics then it might be possible to adopt sorting policies that increase overall productivity, and potentially even Pareto-efficiency.

The potential impact of peer effects looms large in education policy. If peer effects exist, then the individual educational benefits of any effective reform would be multiplied by the additional influence that the affected student has on his future peers. Peer effects were a primary justification for previous forced desegregation policies and now underlie a new generation of policies integrating students by income levels, such as one underway in Wake County, North Carolina. Further, important theoretical models for the impact of residential sorting (Nechyba 2000; Ferreyra 2007) and for the implications of school choice policies (Epple and Romano 1998) on student achievement rely on the existence of peer effects to drive their results.

However, these literatures currently lack the reliable estimates of the size of peer effects in education that they need in order to accurately calibrate this key part of their models (Nechyba 2006). The primary difficulty with estimating the influence of peer effects with which prior empirical work has grappled, particularly when evaluating early grades, is the fact that individuals sort themselves into peer groups. Recent attempts to take advantage of exogenous variation in peer quality are plausible but somewhat wanting.

This paper utilizes an IV approach to estimate peer effects on student math and reading proficiency in the fourth and fifth grades. I identify the causal impact of peer effects using exogenous, across-cohort variation in the quality of peers that resulted from the implementation of a test-based promotion policy for third grade students.

Beginning with the third-grade class of 2002-03, third grade students in Florida were required to reach a minimal benchmark on the state's reading test in order to be default promoted to the fourth grade. Students who scored below this threshold could obtain an exemption to the policy and be promoted, but the default was changed so that they would be retained without additional intervention. Prior to this policy, the vast majority of low-performing students were "socially promoted" to the next grade regardless of their academic proficiency. Adoption of the

test-based promotion policy substantially changed the probability that a low-performing student would be retained.

One result of this policy was that students in cohorts subject to it entered later grades with fewer low-proficiency students in their classes than did students in prior cohorts whose low-proficiency classmates were mostly promoted along with them. That is, the policy produced variation in the proportion of students in a cohort who were low-performing that was in large part determined by the year in which the student happened to have entered the third grade. I take advantage of this exogenous variation in peer quality to identify causal estimates of the influence of peers on student math and reading achievement in the fourth and fifth grades.

I find evidence that the quality of peers has a significant influence on student academic proficiency. Importantly, the magnitude of peer effects differs according to prior student achievement, though students across the ability spectrum are influenced by peers. The structure of peer effects is such that sorting can be used to create a more equitable distribution of student learning and in the case of reading could potentially increase overall learning, but sorting itself is not Pareto-improving -- any learning gains made by low performing students due to such policies come at the expense of learning losses for higher performing students.

The remainder of this paper follows in six additional parts. Section II discusses methodological issues with measuring peer effects and prior attempts to solve them. Section III describes the instrument utilized in this paper. Section IV describes the data and sample and Section V lays out the formal model for estimation. Section VI reports the results of the analysis and Section VII provides some discussion of how they should be interpreted.

II) Conceptual Issues with Measuring Peer Effects

Measuring the influence of peers on student learning has proven a daunting task. Two issues face the researcher intending to measure peer effects: accounting for omitted variables that are very likely correlated with peer quality, and for the fact that a student is both influenced by and influences his peers. As does most prior empirical research, I focus on solving the former problem, which is thought to be the most important for estimating peer effects (see for example Hanushek and, Kain, Markman and Rivin 2002).

Consider a simple education production function where the academic achievement of student i assigned to teacher j within school s at the end of year t (Y_{ijst}) is a linear function of

student characteristics (X), combined school and teacher attributes (S), and the quality of classroom peers, which is represented as the average prior year's test score of the other students in i 's grade within the school during year t ($\bar{Y}_{-ijst-1}$). Formally:

$$(1) \quad Y_{ijst} = c + \psi Y_{ijst-1} + \theta X_{ijst} + \lambda S_{ijst} + \alpha \bar{Y}_{-ijst-1} + \varepsilon_{ijst}$$

The model appears relatively straightforward. However, identification is problematic. The inclusion of the student's prior year's test score is meant to account both for the student's prior schooling and for unobserved factors about the student that are important to his learning, though it is an imperfect proxy. Some of these unobserved factors are likely captured by the error term, while those that are correlated with other regressors in the model will bias the estimates of their coefficients.

The omitted variable problem discussed above is common in education policy research. However, the problem is exacerbated and must be addressed when measuring peer effects. As pointed out by Hanushek and, Kain, Markman and Rivin (2002), because they live in the same neighborhoods and have attended the same schools, members of peer groups likely have had similar historical experiences. In fact, $\bar{Y}_{-ijst-1}$ could be thought of as a "strong proxy" for omitted historical variables. Thus, omitted or poorly measured historical factors that produced both student i 's achievement and his peers achievement will be correlated, and this correlation will tend to bias estimation.

Prior researchers have tried a variety of approaches to handle the endogeneity problem produced by likely omitted historical variables. Some analyses of peer quality in higher education have utilized circumstances where individuals were randomly assigned to peer groups. Sacerdote (2001) and Zimmerman (2003) evaluate the influence of peers on performance in college using as an instrument the random assignment of roommates, while in a similar procedure Carrel, Fullerton, and West (2008) look at random assignment of peers within the United States Air Force Academy and Lyle (2007) uses random assignment of cadets to companies at West Point.

International studies in Kenya (Duflo, Dupas, and Kremer 2008) and in China (Carman and Zhang 2008) have also utilized instances where students are randomly assigned to peer groups in K-12 schools. There are two U.S. studies evaluating the influence of peers on student performance in K-12 grades that take advantage of random assignment, but neither is able to do

so directly. Both of these studies utilize data from the Tennessee STAR experiment, under which students were randomly assigned to smaller or larger classes in kindergarten through the third grade, and then returned to regular classes in the fourth grade. Boozer and Cacciola (2001) use as an instrument the percentage of a student's peers who were previously exposed to the class size treatment, and thus would have benefited with greater proficiency. Whitmore (2005) evaluates student performance in the randomly assigned grades and uses as her instrument for peer quality the percentage of students in the randomly assigned classroom who were female.

Another class of studies has utilized instances when individuals are randomly assigned to residential neighborhoods to study the impact of peers on outcomes (Hanratty et al. 1998; Katz et al. 2001; Ludwig, Duncan and Hirschfeld 2001; Ludwig, Duncan and Pinkston, 2000; Ludwig, Ladd and Duncan 2001; Leventhal and Brooks-Gunn, 2001; Rosenbaum 1991). However, of these studies of neighborhood peers only Ludwig, Ladd and Duncan (2001) evaluates their impact on measured academic outcomes in elementary and secondary schools.

When random assignment is lacking, other studies have looked for natural variation in peer assignments. One such strategy has been to use idiosyncratic variation in the composition of peer groups across cohorts (Hoxby 2000; Vigdor and Nechyba 2004). Other recent studies have looked for variation in peer quality induced by public policy. Angrist and Lang (2004) evaluate the impact of an inter-district bussing program on suburban students in schools that accept urban minority students, taking advantage of the fact that assignment of urban students is related to the number of classroom openings expected within the school. Lai (2007) takes advantage of a classroom division rule in the sorting of students in China.

Cooley (2009) utilizes the implementation of a test-based promotion policy for fifth grade students in North Carolina to instrument peer quality. However, while the policy utilized in that paper is similar to that taken advantage of in the present paper, the way that the policy is used to instrument for peer quality is quite different. While the current paper focuses on differences in the quality of peers in years after students were subject to test-based promotion, Cooley (2009) assumes that peer quality increases within the gateway grade by increasing the effort level of low-performing students who want to avoid retention.

Some prior work has also relied heavily on the use of powerful controls to directly capture the selection process. Hanushek, Kain, Markman and Rivin (2002) and Vigdor and Nechyba (2004) rely on within-school variation in peers by controlling for school fixed-effects

while Burke and Sass (2008) link students to classrooms and utilize a teacher-school fixed effect. Ding and Lehrer (2007) utilize the fact that sorting of students to schools in China is more directly related to observed characteristics rather than individual sorting choices as is the case in the U.S. system.

A reasonable criticism of this prior work evaluating peers within U.S. elementary and secondary schools, particularly in early grades, is that while it has identified some significant and likely random differences in peers, these differences due to the instrument utilized are often minor -- for example, if the cohort happens to have a few more female students in it than did the last cohort. Such minor differences in peer quality due to exogenous factors could make strong identification difficult.

(III) Instrumental Variable Approach

Lacking random assignment, the most common procedure to identify models facing endogeneity issues such as those facing the measurement of peer effects is to utilize an instrumental variable in a two-stage least squares (2SLS) framework. In the context of equation (1), such an instrument, Z , must satisfy two conditions. First, the instrument must be uncorrelated with ε -- $\text{Cov}(Z, \varepsilon) = 0$ -- and secondly it must be partially correlated with $\bar{Y}_{-ijst-1}$ after the influence of all other regressors in (1) -- X , and S -- have been netted out. In the current context, a suitable instrument would be something that is related to the average test scores of a student's peers at the beginning of the academic year and yet is otherwise unrelated to the student's own learning during that year.

The implementation of a test-based promotion policy for third-grade students in Florida provides a massive natural experiment that can be used to identify peer effects in such an IV setting. Beginning with the third grade class of 2002-03, Florida students were required to reach a minimal benchmark on the state's mandated reading test -- the Florida Comprehensive Assessment Test (FCAT) -- in order to be default promoted to the fourth grade. Students who failed to reach this benchmark could obtain a promotion through one of six exemptions.¹ During

¹ A student could obtain an exemption if he: was a limited-English-proficiency student with less than two years of instruction in English; had a disability whose IEP indicated that testing was inappropriate for him; scored above the 51st percentile on another standardized reading test (far above the standard set on the original reading assessment); was disabled and received intensive remediation in reading; demonstrated proficiency through a student portfolio; or had been retained twice previously.

the years evaluated in this paper, the third grade was the only "gateway" grade from which students were required to test out.

Adoption of the test-based promotion policy substantially increased the percentage of third grade students who were retained at the end of the year. Table 1 shows retention percentages for third grade students by year. In each of the two years prior to the adoption of the test-based promotion policy, less than 3 percent of third grade students were retained at the end of each of those years. This jumped to nearly 11 percent of students being retained in the third grade class of 2002-03, the first class required to pass a reading test in order to be default promoted, and nearly 8 percent of third graders in the second class subject to the policy being retained.

[TABLE 1 ABOUT HERE]

Other recent studies have used the exogenous increase in the probability of retention for low-performing students due to this and similar policies to study the impact of retention on student achievement (Greene and Winters 2007; Jacob and Lefgren 2004; Roderick and Nagaoka 2005). However, while the adoption of test-based promotion certainly influenced students who were retained due to the policy, the change in the probability that low-performing students were retained also had an important impact on the quality of the peers surrounding students in later grades.

The requirement that students meet a minimal benchmark in order to be default-promoted severely reduced the number of low-proficiency students sharing classrooms with higher proficiency students in later grades for the effected cohorts. Low-proficiency students who were retained at the end of the third grade exited their cohort and moved into the next third grade cohort. The result was that students subject to the test-based promotion policy had on average higher quality peers in later grades than did students in the cohorts prior to the introduction of the policy. This increase in peer quality was primarily produced by the exogenous fact that these students happened to have entered the third grade in the fall of 2002 rather than the fall of 2001.

Figures 1-4 illustrate the differences in peer quality before and after adoption of the test-based promotion policy. The figures provides kernel density estimates for the distribution of the average reading and math test scores of school-grade peers in the fourth and fifth grade for those students who entered the third grade in each class from 2001 to 2004 and were not retained in any of the observed years. The fourth and fifth grade peers for students who were in the third

grade in the fall of 2000 or 2001 -- and thus were not subject to the policy -- are quite similar to one another. However, beginning with the third grade class of 2002-03 -- the first two cohorts required to test into the fourth grade -- there is a large shift to the right in the distribution of peer math and reading scores.

[FIGURES 1-4 ABOUT HERE]

Table 2 quantifies these differences. To put the differences in these scores into context, using summary statistics on the 2003-04 administration of the third grade assessments indicates that students who entered the third grade cohort in 2002-03 had fourth grade peers with an average test score that was 0.30 standard deviations higher in reading and 0.16 standard deviations higher in math than did the prior cohort.² The larger difference in reading scores between cohorts subject to the policy is expected since it is the results of the reading exam that triggers the retention decision.

[TABLE 2 ABOUT HERE]

I argue that at least a substantial part of the difference in the average quality of peers for students across these cohorts is due to the test-based promotion policy. Further, I argue that this policy change itself is exogenous to the later academic proficiency of individual students after they exit the fourth grade. That is, whether or not the student was in a cohort subject to the policy is related to the quality of his peers in later grades but is otherwise unrelated to his own academic growth in those grades. This is consistent with the classical definition of an instrumental variable for quality of peers.

While the structural break shown in the figures strongly suggests that adoption of the policy is good candidate to be used to instrument peer quality, there are a few concerns that should be considered.

Some might worry that being subjected to the policy in the third grade might actually be related to the student's learning in later grades, whether or not he was actually retained by it. In particular, it is likely that low-achieving students that were subject to the policy were motivated to work harder while in the third grade in order to avoid retention. Schools may have also focused more resources in the third grade in order to help students meet the minimal test score requirement. Such factors would influence a student's proficiency as he enters the fourth grade.

² For students in the dataset, the 2003-04 administration of the third grade tests, had a math test a mean of 1367 with a standard deviation of 300 and the reading exam had a mean of 1317 and a standard deviation of 390.

I address this issue by adopting a value-added model that controls for prior student proficiency such as that specified in (1). Controlling for the student's prior test score would only be insufficient for this task if the student's third grade score was not a true measure of his proficiency (for example, if the student cheated in order to be exempt from the policy) or if the relationship between prior student proficiency and his learning in a given year differed according to the process by which he achieved his prior proficiency. While there are likely some instances of cheating, I assume that it is not systematic and thus that student third grade test scores are accurate measures of proficiency. A relationship that would produce the latter concern is conceivable, though unlikely. To my knowledge, no prior work suggests that the relationship between prior proficiency and current learning depends on how that prior proficiency was achieved, and no prior empirical research in education of which I am aware attempts to account for such bias.

A more realistic worry about the proposed instrument is that the substantial increase in the number of students retained in their cohort might have also been related to a reduction in average class sizes in later grades. If teachers were not redistributed to different grades within schools, then the reduction in the number of students entering the fourth grade in 2003-04 and fifth grade in 2004-05 due to the policy would lead to a decrease in class sizes for this cohort that could severely confound our estimates. Fortunately, this does not appear to have been the case. A comparison of average class sizes in the fourth and fifth grades across cohorts finds no meaningful differences. The average fourth grade class size was 21.5 for students who entered the third grade in 2001-02 and 21.7 for students who entered the third grade in 2002-03. The average fifth grade class size was 23.7 for students who entered the third grade in 2001-02 and 23.6 for students who entered the third grade in 2002-03. I further account for such an effect by incorporating a control for the size of the student's class. The inclusion or removal of this control has no impact on the coefficients related to peer quality.

We might additionally worry that teachers were non-randomly moved from the fourth to the third grade. In particular, school systems may have moved better teachers into the third grade in order to reduce the number of students who fail the promotion benchmark in that gateway grade. I account for such potential bias by incorporating a teacher fixed-effect in the preferred models to account for teacher quality directly. To the extent that such bias remains in models that do not use a teacher fixed-effect, we should expect that it would tend to understate the

importance of peers since peer quality was greater for students from the third grade class of 2002-03, who would have been deprived of better teachers due to such non-random sorting.

IV) Data and Sample

I utilize an administrative dataset containing math and reading test scores and observable demographic information for the universe of test-taking elementary school students in the state of Florida for each year from 2001-02 through 2004-05. The dataset also contains student, school, teacher, and grade identifiers.³ Test scores are measured according to a vertically aligned scale. This means that the test score is developed in such a way that identical scores should represent identical levels of proficiency no matter the grade level of the test.

I restrict the models for estimation to only include students who were not retained in any of the observed years. The reason for this exclusion is that students retained according to the policy do not have a natural comparison group in the cohorts that were not subjected to the policy. However, the scores of these students are used to calculate the average incoming test scores of a student's peers.

Table 3 illustrates the cohort samples used for estimation. The years included in the dataset limit the available sample to students who were in the fourth and/or fifth grade in years 2001-02 through 2004-05. In order to properly identify grade and year effects during the estimation, I incorporate the observed fifth grade performance of the cohort who entered the third grade in 2000-01 and the observed fourth grade performance of students in the cohort who entered the third grade in 2004-05. I do not include fourth grade observed performance of students in the cohort that entered in 2000-01 because it would have been the only observed student performance that year, leading a fixed effect accounting for that year to be unidentified. A similar problem is faced for the fifth grade performance of the 2004-05 cohort, in addition to the fact that the dataset does not include the necessary 2005-06 test scores.

[TABLE 3 ABOUT HERE]

Descriptive statistics for variables relevant to estimation are reported in Table 4.

[TABLE 4 ABOUT HERE]

³ 1,289 student observations in the dataset were not matched to a particular individual teacher, and thus are excluded from analyses that use a teacher fixed-effect or that define peers at the classroom level. This explains the drop in the number of observations shown in the tables.

V) Estimation Method

The 2SLS procedure follows thusly. A first-stage (instrumenting) OLS regression uses observable student characteristics (X) -- including cubic functions of the student's prior math and reading test scores -- fixed effects for the teacher (δ) and the year (κ), an indicator for whether the student is in the fourth or fifth grade ($Grade$), the percentage of a student's peers who have a series of observable characteristics (M) and an indicator variable for whether the student entered the third-grade after implementation of the test-based retention policy ($Post_Policy$) to explain the average incoming math or reading scores of student i 's grade level peers in the fourth and fifth grades. I also allow the relationship between being subject to the policy and peer quality to differ by grade by incorporating an interaction between $Post_Policy$ and $Grade$.

$$(2) \quad \begin{aligned} \bar{Y}_{-ijst-1} = & b + \phi X_{ijst} + \lambda M_{-ijgst} + \delta_j + \kappa_t + \nu Grade_{ijst} \\ & + \gamma Post_Policy_i + \zeta (Grade_{isjt} * Post_Policy_i) + \mu_{ijst} + \mu_s \end{aligned}$$

Where b , ϕ , λ , ν , γ , and ζ are parameters to be estimated, μ_{ijst} is a stochastic term and μ_s reminds us that error is clustered by school.

Reduced form estimates of (2) are used to predict the average test score of a student's peers that year, \hat{Y} , which is then used as an independent variable in a second stage OLS regression explaining student achievement in the fourth and fifth grades.

$$(3) \quad Y_{ijst} = g + \rho X_{ijst} + \eta M_{-ijst} + \chi_j + o_t + \omega Grade_{isjt} + \beta \hat{Y}_{-ijst} + \tau_{ijst} + \tau_s$$

Where g , ρ , η , ω , and β are parameters to be estimated, χ is a teacher-school fixed effect, o is a year fixed-effect, and τ is a stochastic term clustered by school.

If we find that $\gamma \neq 0$ and/or that $\zeta \neq 0$, and if we believe that the assumption that $Cov(Cohort, \tau) = 0$ holds, then the OLS estimator of β is a consistent estimate of the impact of average peer quality on student academic proficiency.

I further modify (3) in search of mediators of peer effects. In particular, estimates are obtained using interactions to evaluate whether peers have differential influence on students by gender or race/ethnicity, and also whether peer effects differ according to prior student proficiency.

VI) Results

Table 5 reports the results of the instrumenting equations in math and reading. Each of the models has strong predictive power, as indicated by the very high R-squared values. The table only reports coefficient estimates and t-statistics for the instrumental variables -- the indicator for whether the student was in a cohort that was subject to the test-based promotion policy and an interaction between this variable and the student's current grade level. In each model the coefficients are positive, substantial, and highly statistically significant, which is important for strong identification of the second stage.

[TABLE 5 ABOUT HERE]

The results of the linear-in-means estimations of the second stage equations are reported in Table 6. The preferred estimates are those that incorporate fixed-effects at the teacher level. I find significant peer effects in both math and reading. In reading the model finds that a one point increase in the quality of a student's peers leads to an average of a 0.113 point increase in his learning that year. The effect is larger in math. A one point increase in the average math test score of a student's peers leads to a 0.331 point increase in his own math test score at the end of the year.

In both math and reading, the estimated influence of peers grows when we incorporate a teacher fixed effect. This finding is consistent with what we would expect if more able teachers were assigned to the third grade due to the policy.

[TABLE 6 ABOUT HERE]

In order to better understand the structure of peer effects, and thus their implications for sorting policies, I now look for modifiers of peer effects based on observed student characteristics. Table 7 reports the results of regressions that incorporate interactions between observed student characteristics and the instrumented measurement of peer quality. In both math and reading, the coefficient estimates are quite similar whether the model uses a school or a teacher fixed effect.

[TABLE 7 ABOUT HERE]

Of particular interest are the coefficient estimates for the variables interacting the instrumented average peer quality with an indicator of the quartile of the student's prior test score. These measures show whether the influence of peers differs by the student's own

proficiency level, which is necessary for sorting policies to have anything other than distributional effects.

Interestingly, these results differ substantially across subjects. Though the differences in the influence of peer quality by the student's beginning math proficiency quartile are statistically significant, they are not particularly large, suggesting that peer effects in math are relatively linear across the distribution of prior student proficiency. In reading, however, while students in each quartile of the achievement distribution benefit from having more advanced peers, students in the top quartile benefit much more than do lower achieving students. For instance, according to the model that incorporates a teacher fixed-effect (Column VI), all else equal, a student in the bottom quartile will see a 0.273 point increase in his reading proficiency for a one point increase in the average quality of his peers, while a student in the top quartile will see his reading proficiency increase by 0.496 points for a one point increase in the quality of his reading peers.

VII) Interpretation and Conclusion

I find evidence that the quality of a student's peers has a substantial influence on his learning. Students across the prior ability distribution are influenced by their peers in both math and reading. However, the structure of the influence of peers differs substantially by subject.

Peer effects do not appear to take a structure that would allow for sorting policies that are Pareto-improving. Since peer effects are important for students across the ability distribution, learning gains made for low-achieving students by placing more high-achieving students in their classroom come at the expense learning losses for those high-achieving students. Thus, sorting policies can be used to make the distribution of learning more equitable, but it cannot do so without harming some students. Policymakers intending to use sorting policies must weigh the gains made by some students against the losses of others.

References

- Angrist, Joshua. D., and Kevin Lang. 2004. "Does School Integration Generate Peer Effects? Evidence from Boston's Metco Program." *American Economic Review*, 94: 1613-1634.
- Boozer, Michael A. and Stephen E. Cacciola. 2001. "Inside the 'Black Box' of Project STAR: Estimation of Peer Effects Using Experimental Data." Unpublished Manuscript.
http://www.econ.yale.edu/growth_pdf/cdp832.pdf
- Burke, Mary A. and Tim Sass. 2008. "Classroom Peer Effects and Student Achievement." National Center for Analysis of Longitudinal Data in Education Research, Working Paper 18.
- Carman, Katherine and Lei Zhang. 2008. "Classroom Peer Effects and Academic Achievement: Evidence from a Chinese Middle School." Unpublished manuscript.
http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1157549
- Carrell, Scott. E., Richard. L. Fullerton, and James. E. West. 2008. "Does Your Cohort Matter? Measuring Peer Effects in College Achievement." Unpublished manuscript.
<http://www.econ.ucdavis.edu/faculty/scarrell/peer2.pdf>
- Cooley, Jane. 2009. "Desegregation and the Achievement Gap: Do Diverse Peers Help?" Unpublished manuscript. <http://www.ssc.wisc.edu/~jcooley/CooleyDeseg.pdf>
- Ding, Weili. and Steven Lehrer. 2007. "Do Peers Affect Student Achievement in China's Secondary Schools?" *The Review of Economics and Statistics*, 89(2): 300-312.
- Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2008. "Peer Effects and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya." NBER Working Paper # 14475.
- Epple, Dennis and Romano, Richard. 1998. "Competition Between Private and Public Schools, Vouchers and Peer Group Effects" *American Economic Review*, 88(1): 33-62.

Ferreya, Maria M. 2007. "Estimating the Effects of Private School Vouchers in Multidistrict Economies." *American Economic Review*, 97(3): 789-817.

Greene, Jay. P. and Marcus A. Winters. 2007. "Revisiting Grade Retention: An Evaluation of Florida's Test-Based Promotion Policy." *Education Finance and Policy*, 2(4): 319-340.

Hanratty, Maria, Sara McLanahan, and Becky Pettit. 1998. "The Impact of the Los Angeles Moving to Opportunity Program on Residential Mobility, Neighborhood Characteristics, and Early Child and Parent Outcomes." Bendheim-Thoman Center for Research on Child Wellbeing Working Paper #98-18.

Hanushek, Eric, John. F. Kain, Jacob M. Markman, Steven G. Rivkin. 2003. "Does Peer Ability Affect Student Achievement?" *Journal of Applied Econometrics*, 18(5): 527-544.

Hoxby, Caroline. 2000. "Peer Effects in the Classroom: Learning from Gender and Race Variation." NBER Working Paper #7867.

Jacob, Brian, and Lars Lefgren. 2004. "Remedial Education and Student Achievement: A Regression-Discontinuity Analysis." *The Review of Economics and Statistics* 86(1): 226-44

Katz, Lawrence F., Jeffrey R. Kling, and Jeffrey B. Liebman. 2001. "Moving to Opportunity in Boston: Early Results of a Randomized Mobility Experiment." *Quarterly Journal of Economics*, 116: 607-654.

Lai, Fang. 2008. "How Do Classroom Peers Affect Student Outcomes? Evidence from a Natural Experiment in Beijing's Middle Schools." Unpublished manuscript.
http://www.aeaweb.org/annual_mtg_papers/2008/2008_447.pdf

Leventhal, T. and J. Brooks-Guss (2001). "Moving to Opportunity: What About the Kids?" Unpublished Manuscript.

Ludwig, Jens, Greg J. Duncan, and Paul. Hirschfield. 2001. "Urban Poverty and Juvenile Crime: Evidence from a Randomized Housing-Mobility Experiment." *Quarterly Journal of Economics*, 116: 655-680.

Ludwig, Jens, Greg J. Duncan, and Joshua C. Pinkston. 2000. "Neighborhood Effects on Economic Self-Sufficiency: Evidence from a Randomized Housing-Mobility Experiment." Unpublished Manuscript. http://www.nber.org/mtopublic/baltimore/mto_balt_employment.pdf

Ludwig, Jens, Helen Ladd, and Greg Duncan. 2001. "The Effects of Urban Poverty on Educational Outcomes: Evidence from a Randomized Experiment." Unpublished Manuscript.

Lyle, David S. 2007. "Estimating and Interpreting Peer and Role Model Effects From Randomly Assigned Social Groups at West Point." *The Review of Economics and Statistics*, 89(2): 289-299.

Nechyba, Thomas J. 2000. "Mobility, Targeting and Private School Vouchers." *American Economic Review*, 90: 130-46.

Nechyba, Thomas J. 2006. "Income and Peer Quality Sorting in Private and Public Schools." in Eric Hanushek and Finis Welch eds. *Handbook of the Economics of Education*, Volume 2. North Holland Press.

Roderick, Melissa, and Jenny Nagaoka. 2005. "Retention Under Chicago's High-Stakes Testing Program: Helpful, Harmful, or Harmless?" *Educational Evaluation and Policy Analysis* 27(4): 309-40.

Rosenbaum, James. 1991. "Black Pioneers -- Do Their Moves to the Suburbs Increase Economic Opportunity for Mothers and Children?" *Housing Policy Debate* 2: 1179-1213.

Sacerdote, Bruce. 2001. "Peer Effects with Random Assignment: Results for Dartmouth Roommates." *Quarterly Journal of Economics*, 116: 681-704.

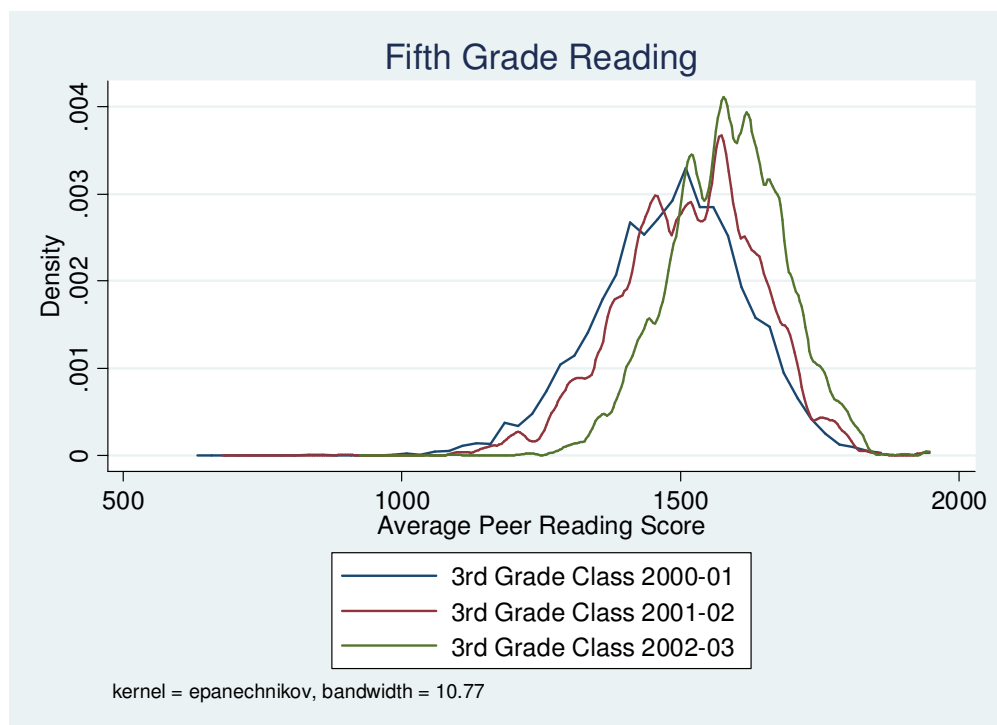
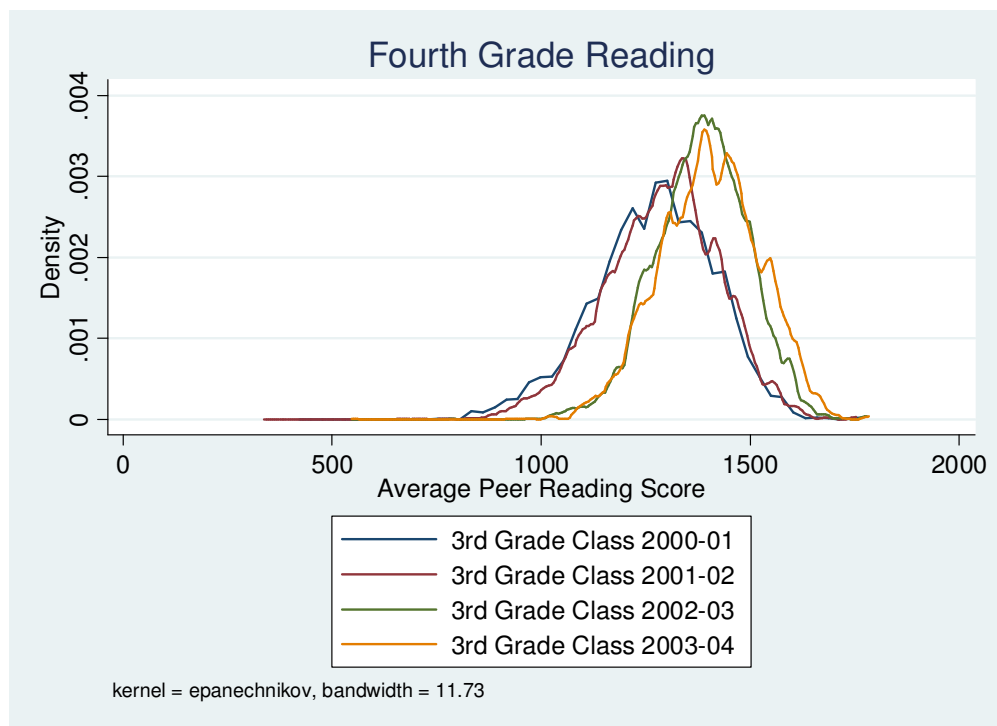
Vigdor, Jacob and Thomas J. Nechyba. 2006. "Peer Effects in North Carolina Public Schools." In P. Peterson and L. Wossmann (Eds.), *Schools and the Equal Opportunity Problem*. MIT Press, Cambridge.

Whitmore, Diane. 2005. "Resource and Peer Impacts on Girls' Academic Achievement: Evidence from a Randomized Experiment." *American Economic Review*, 95: 199-203.

Zimmerman, David J. 2003. "Peer Effects in Academic Outcomes: Evidence from a Natural Experiment." *The Review of Economics and Statistics*, 85(1): 9-23.

Figures 1 - 4

Kernel Density Estimates of Grade-Level Peers by Later Grade



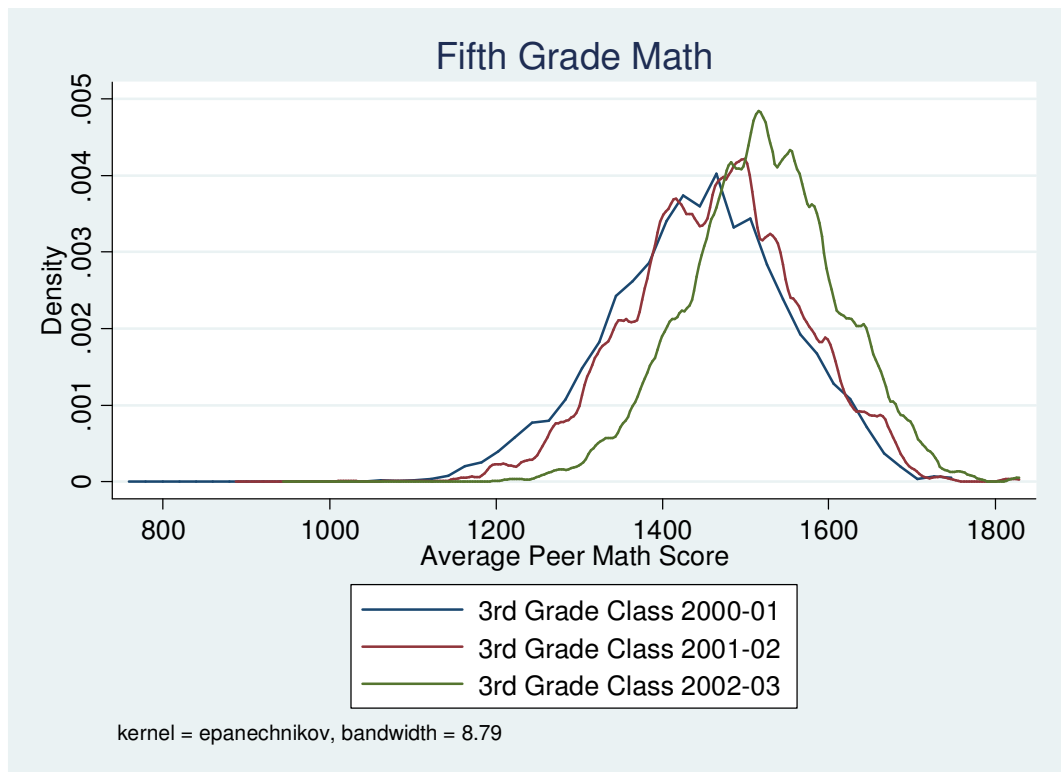
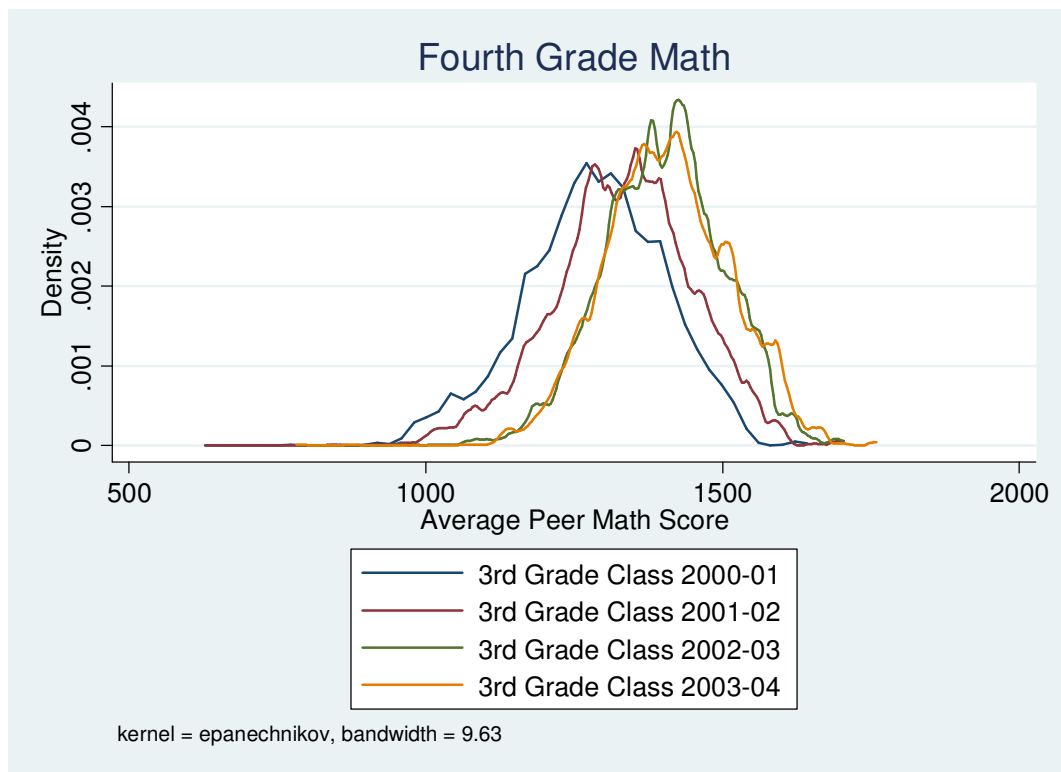


Table 1
Third Grade Retentions by Year

	Began Third Grade	Retained End of Year	Percent Retained
2000-01	180,997	4,764	2.6%
2001-02	186,972	5,246	2.8%
2002-03	186,892	19,828	10.6%
2003-04	203,722	15,452	7.6%

Note: Test-based promotion policy enacted 2002-03.

Table 2
Average Entering Test Scores of Fourth and Fifth Grade Peers by Third Grade Cohort

		Entered Third Grade			
		2000-01	2001-02	2002-03	2003-04
Fourth Grade	Reading	1264	1288	1378	1403
	Math	1281	1330	1393	1406
Fifth Grade	Reading	1480	1517	1582	
	Math	1441	1463	1520	

Table 3
Sample Cohort by Grade and Year

Observed Year	Grade 4	Grade 5
2002-03	Cohort 2002	Cohort 2001
2003-04	Cohort 2003	Cohort 2002
2004-05	Cohort 2004	Cohort 2003

Table 4
Descriptive Statistics for Variables Included in Regressions

	Mean	Std. Dev.
Prior Math Score	1446.380	272.118
Prior Reading Score	1470.395	344.595
Grade 4	0.496	0.500
Grade 5	0.504	0.500
Male	0.501	0.500
White	0.537	0.499
Asian	0.020	0.142
African-American	0.217	0.412
Hispanic	0.204	0.403
American Indian	0.003	0.057
Multiple Race	0.030	0.172
Eligible for Free Lunch	0.481	0.500
Not Limited English Proficient	0.848	0.359
Disabled	0.142	0.349
Entered Third Grade Post Policy	0.481	0.500
Peer Percent Male	0.510	0.052
Peer Percent African-American	0.222	0.239
Peer Percent Hispanic	0.218	0.238
Peer Percent American Indian	0.003	0.007
Peer Percent Asian	0.021	0.024
Peer Percent Eligible for Free Lunch	0.496	0.253
Peer Percent Disabled	0.155	0.069

Table 5
Results of First Stage (Instrumenting) Equations in Math and Reading

	Reading		Math	
	(I)	(II)	(III)	(IV)
Entered Third Grade Post Policy	13.55*** [6.807]	7.545*** [3.618]	25.05*** [13.76]	20.40*** [10.88]
Entered Third Grade Post Policy * Grade 4	28.31*** [14.93]	30.67*** [15.25]	7.584*** [4.448]	9.606*** [5.390]
School Fixed-Effect	√		√	
Teacher Fixed Effect		√		√
Number of Schools	1,847	1,847	1,847	1,847
Number of Teachers		30,910		30,910
Observations	857,475	855,736	857,475	855,736
R-squared	0.939	0.952	0.913	0.935

Dependent variable in columns (I) and (II) is the average reading test score for the student's grade-school peers, and in columns (III) and (IV) is the average math test score for the student's grade-school peers. Regressions also control for year of observation, grade level, average class size, cubic functions for the student's prior year's math and reading scores, gender, race/ethnicity, free or reduced priced lunch status, Limited English Proficiency status, and whether the student has an Individual Education Plan.

Regressions also account for percentage of the student's grade level peers by gender, race/ethnicity, free or reduced-priced lunch status, and that have an Individual Education Plan. Students who were retained at any time during the observed years are excluded from the analysis. T-statistics using standard errors clustered by school are reported in the brackets.

Table 6
Linear in Means Regressions in Math and Reading

	Reading (I)	(II)	Math (III)	(IV)
Average Peer Score (Instrumented)	0.0472 [1.389]	0.113*** [2.938]	0.282*** [7.558]	0.331*** [7.896]
School Fixed-Effect	√		√	
Teacher Fixed-Effect		√		√
Number of Schools	1,847	1,847	1,847	1,847
Number of Teachers		30,889		30,887
Observations	855,388	853,674	855,344	853,626
R-squared	0.668	0.697	0.71	0.752

Dependent variable in all regressions is the student's reading score at the end of the observed school year. Regressions also control for year of observation, grade level, average class size, cubic functions for the student's prior year's math and reading scores, gender, race/ethnicity, free or reduced priced lunch status, Limited English Proficiency status, and whether the student has an Individual Education Plan. Regressions also account for percentage of the student's grade level peers by gender, race/ethnicity, free or reduced-priced lunch status, and that have an Individual Education Plan. Students who were retained at any time during the observed years are excluded from the analysis. T-statistics using standard errors clustered by school are reported in the brackets.

Table 7
 Modifiers of Peer Effects by Observed Student Characteristics

	Reading (V)	(VI)	Math (VII)	(VIII)
Average Peer Reading Score (Instrumented)	0.487*** [13.00]	0.496*** [10.98]	0.294*** [7.598]	0.280*** [6.449]
Peer * Bottom 25th Percentile Reading Score	-0.217*** [-26.54]	-0.223*** [-29.25]	-0.0351*** [-4.113]	-0.0281*** [-3.870]
Peer * 26th - 50th Percentile Reading Score	-0.255*** [-40.14]	-0.238*** [-39.32]	-0.0730*** [-12.78]	-0.0560*** [-10.76]
Peer * 51st - 75th Percentile Reading Score	-0.186*** [-33.87]	-0.171*** [-32.62]	-0.0503*** [-10.87]	-0.0375*** [-8.864]
Peer * Male	0.0220*** [7.297]	0.0111*** [3.665]	-0.00465 [-1.498]	-0.0196*** [-6.664]
Peer * Asian	-0.0804*** [-7.722]	-0.0695*** [-7.058]	-0.000199 [-0.0188]	0.00948 [0.951]
Peer * African-American	-0.0453*** [-7.827]	-0.0215*** [-4.266]	0.00567 [0.870]	0.0123** [2.455]
Peer * Hispanic	-0.0409*** [-6.803]	-0.0179*** [-3.291]	0.0218*** [3.284]	0.0278*** [5.857]
Peer * Indian	-0.0186 [-0.763]	-0.0204 [-0.838]	0.0173 [0.687]	0.0178 [0.721]
Peer * Multiple Race	-0.0174** [-2.043]	-0.0142* [-1.681]	0.00459 [0.555]	0.0132* [1.702]
Peer * Fourth Grade	-0.101*** [-12.96]	-0.0816*** [-7.591]	0.0177* [1.830]	0.0590*** [5.419]
School Fixed-Effect	√		√	
Teacher Fixed-Effect		√		√
Number of Schools	1,847	1,847	1,847	1,847
Number of Teachers		30,889		30,887
Observations	855,388	853,674	855,344	853,626
R-squared	0.671	0.698	0.71	0.752

Dependent variable in all regressions is the student's reading score at the end of the observed school year. Regressions also control for year of observation, grade level, average class size, cubic functions for the student's prior year's math and reading scores, gender, race/ethnicity, free or reduced priced lunch status, Limited English Proficiency status, and whether the student has an Individual Education Plan. Regressions also account for percentage of the student's grade level peers by gender, race/ethnicity, free or reduced-priced lunch status, and that have an Individual Education Plan. Students who were retained at any time during the observed years are excluded from the analysis. T-statistics using standard errors clustered by school are reported in the brackets.