CALDER

National
Center for Analysis of Longitudinal Data in Education Research

TRACKING EVERY STUDENT'S LEARNING EVERY YEAR

Urban Institute

A program of research by the Urban Institute with Duke University, Stanford University, University of Florida, University of Missouri-Columbia, University of Texas at Dallas, and University of Washington

Classroom Peer Effects and Student Achievement

Mary A. Burke

and

Tim R. Sass

WORKING PAPER 18 • JUNE 2008

Classroom Peer Effects and Student Achievement

Mary A. Burke
Research Department
Federal Reserve Bank of Boston

Tim R. Sass
Department of Economics
Florida State University

June 2008

In this paper, we analyze the impact of classroom peers on individual student performance with a unique longitudinal data set covering all Florida public school students in grades 3–10 over a five-year period. Unlike many previous data sets used to study peer effects in education, our data allow us to identify each member of a given student's classroom peer group in elementary, middle and high school as well as the classroom teacher responsible for instruction. As a result, we can control for individual student fixed effects simultaneously with individual teacher fixed effects, thereby alleviating biases due to endogenous assignment of both peers and teachers, including some dynamic aspects of such assignments. Our estimation strategy, which focuses on the influence of peers' fixed (observed and unobserved) characteristics on individual test score gains, also alleviates potential biases due to measurement error of peer quality, simultaneity of peer outcomes, and mean reversion. Under linear-in-means specifications, estimated peer effects are small to nonexistent, but we find some sizable and significant peer effects within nonlinear models. For example, we find that peer effects depend on an individual student's own ability and on the ability level of the peers under consideration, results that suggest Pareto-improving redistributions of students across classrooms and/or schools. Estimated peer effects tend to be smaller when teacher fixed effects are included than when they are omitted, a result that suggests co-movement of peer and teacher quality within a student over time. We also find that peer effects tend to be stronger at the classroom level than the grade level.

We wish to thank the staff of the Florida Department of Education's K–20 Education Data Warehouse for their assistance in obtaining and interpreting the data used in this study. The views expressed is this paper are solely our own and do not necessarily reflect the opinions of the Florida Department of Education.

I. Introduction

The potential for peers to affect individual achievement is central to many important policy issues in elementary and secondary education, including the impacts of school choice programs, ability tracking within schools, "mainstreaming" of special education students, and racial and economic desegregation. Vouchers, charter schools and other school choice programs may benefit those who remain in traditional public schools by engendering competition that leads to improvements in school quality, but may also harm those left behind by diminishing the quality of their classmates (Epple and Romano 1998; Caucutt 2002). Grouping students in classrooms by ability can likewise have significant impacts on student achievement, depending on the magnitude of peer influences (Epple, Newlon, and Romano 2002). The effect of desegregation policies on achievement depends not only on potential spillovers from average ability, but on whether different peers exert different degrees of influence on individual outcomes (Angrist and Lang 2004; Cooley 2007; Fryer and Torelli 2005).

Despite the importance of these issues for American education policy, there are relatively few empirical studies of the magnitude and structure of peer effects on academic achievement in U.S. primary and secondary schools. A number of recent studies have attempted to estimate peer effects in the K–12 education context, yet most have been hampered by data limitations that constrain the scope of their analyses and the estimation techniques they are able to employ. With a unique panel data set encompassing all public school students in grades 3–10 in the state of Florida over the period 1999/00–2003/04, we have unprecedented resources with which to test for peer effects in the educational context. Unlike any previous study, we simultaneously control for the fixed inputs of students, teachers and schools in measuring peer influences on academic achievement. These controls sharply limit the scope for biases from endogenous selection of

peers and teachers and permit a sharper estimate of the influence of *classroom* peers (as opposed to grade-level-at-school peers), than previous studies. Further, unlike previous work, which focuses almost exclusively on peer effects in elementary school, our data allow us to compare the impact of peer influences on math and reading achievement in elementary, middle and high school.

In addition to exploiting an extremely rich data set, we also employ a new analytical technique, adapted from Arcidiacono et al. (2005), that alleviates a number of problems associated with using student performance to measure peer influence. Typically, past research uses contemporaneous or lagged peer outcomes to measure peer ability. This can lead to a number of related estimation problems, such as simultaneity bias, measurement error bias, and biases caused by regression to the mean. Because observed academic outcomes, whether current or lagged, constitute a noisy measure of a student's fixed inputs, measures of peer group influences based on such performance measures will be noisy and peer effects estimates may be biased downward. To better capture peer group characteristics, we estimate "peer fixed effects" simultaneously with individual fixed effects. The method has been shown to perform well even with a small number of observations per student. We extend the work of Arcidiacono et al. by estimating models which allow peer effects to operate through multiple moments of the distribution of the peer-group's fixed effects and in which the effects of peer-group ability depend on individual ability.

An alternative means of avoiding selection biases is to conduct a true experiment in which students and teachers are randomly assigned to classrooms. However, results from experimental data should not automatically be privileged, for reasons both theoretical and practical. First and foremost, a large-scale trial with random assignment of teachers and students

to classrooms is extremely difficult to conduct. While there are some interesting cases of large-scale random assignment at the college level (Sacerdote 2001; Carrell et al. 2008) and in foreign countries (Carman and Zhang 2008; Ding and Lehrer 2007; Duflo et al. 2008; Lai 2008), the legal and political hurdles to random assignment at the elementary and secondary level in the United States are daunting. As a result, there has been only one large-scale randomized trial in U.S. primary and secondary schools, Tennessee's Student/Teacher Achievement Ratio (STAR) experiment. Second, even in randomized trials, selective initial consent by schools to participate or selective individual attrition once the experiment has begun (as observed in the STAR experiment) can bias the results. Third, the magnitude and shape of peer influences may be different depending on whether peers are chosen deliberately—by the individual, her family, or school officials—or at random. Fourth, Arcidiacono et al. (2005) show via simulations that measured peer effects may be biased downward among randomly assigned classmates and that, counterintuitively, the presence of some degree of sorting on student ability may enable more accurate estimates of peer effects.

Based on our quasi-experimental approach, we find that peer effects are small, but statistically significant, when measured with linear-in-means models. We find generally larger and (both statistically and economically) significant peer effects in nonlinear models. In most specifications, omission of teacher effects leads to larger estimated peer effects, indicating that peer and teacher quality may co-vary over time within students and that student fixed effects may not be sufficient for alleviating "correlated effects" biases. Another advantage of controlling for teacher effects is that peer effects estimates are more precise. While we do not claim to identify the teacher effects themselves, controlling for teacher effects assists in the identification

of peer effects by controlling for the possibility that students are assigned to classrooms/teachers on the basis of transient rather than fixed factors.

In the nonlinear models, we find that the magnitude of peer effects depend on an individual student's own ability and on the ability of the peer group under consideration. Both results imply that there are opportunities for Pareto-improving redistributions of students across classrooms and/or schools. We also find that peer effects tend to be much stronger at the classroom level than the grade level—in most cases we find no significant peer effects at the grade-within-school level. This last fact agrees with recent findings by Carrell et al. (2008) that peer effects estimates can differ greatly depending on the accuracy with which the econometrician identifies the set of relevant peers.

II. Previous Literature

Measurements of peer effects at the classroom level have been scarce as a result of data and methodological limitations. Administrative data from Texas identifies only the school and grade level and not specific classroom assignments; hence studies using these data have been limited to grade-level peer effects (Hanushek et al. 2003; Hoxby 2000). Vigdor and Nechyba (forthcoming) and Cooley (2007) both employ statewide data from North Carolina. However, because the North Carolina data do not directly identify the teacher assignments for middle school and high school student records, Vigdor and Nechyba estimate classroom-level peer effects on 5th-grade reading and math achievement test gains, and Cooley estimates classroom-level peer effects—construed as effort spillovers rather than spillovers from fixed peer ability—on 4th and 5th grade reading achievement levels. Rather than employing fixed effects, Vigdor and Nechyba restrict the sample to classrooms satisfying an "apparent random assignment" condition. To isolate random variation in peer effort, Cooley exploits a change in school assessment policy that should

have increased the payoff to effort among lower-ability students. She includes teacher fixed effects and a proxy for unobserved reading ability, but does not include student fixed effects. In addition to estimating linear-in-means specifications, Cooley also uses quantile regression analysis to allow for differences in the impact of peers at different points of the achievement distribution.

Hoxby and Weingarth (2005) estimate classroom peer effects for 4th through 8th grade students from Wake County in North Carolina, using the sum of math and reading end-of-year test score levels as the outcome measure. To measure classroom-peer effects, they exploit a recent policy intervention in which some students were reassigned to different schools in a manner that was purportedly random conditional on students' fixed characteristics. They construct an instrumental variable for the lagged scores of current classroom peers using the initial-period scores and fixed characteristics of the randomly assigned segment of the current school-by-grade peer group. Student fixed effects and grade-level by year effects are accounted for, but school effects and teacher effects are omitted.[1] Multiple specifications of peer effects are estimated, including standard linear-in-means models as well as models in which peer effects are allowed to vary with the student's own ability and with the ability of the peers.

Zabel (2008) uses data from New York City public schools that indicate classroom assignments but not teacher identifiers. Classroom peer effects are estimated for 4th and 5th grade standardized test scores (in levels), but only school-level fixed effects are used. To avoid bias from nonrandom classroom assignment within schools, he takes two approaches: in one case, classroom peer characteristics are instrumented by grade-within-school peer characteristics, and

[1] Such omission need not imply failure of identification. As we discuss below, omission of teacher and/or school fixed effects does not universally result in inconsistent peer effects estimates.

in the second case tests are limited to schools with larger class sizes, within which there is less scope for classroom-level sorting.

Betts and Zau (2004) estimate classroom-level effects on standardized test-score gains in San Diego, controlling for student fixed effects and for several observed teacher characteristics, but they do not employ teacher fixed effects. They also limit their tests to elementary school students, on the grounds that only elementary students spend most of their time in a single classroom and therefore, presumably, are more susceptible to the influence of classroom peers than are students who move across classrooms throughout the day.

Figlio (2005) focuses on the effects of peer behavior on student outcomes. Employing data from a single large Florida school district, he estimates the impact of peer disruptive behavior on individual student behavior and test scores. He controls for student heterogeneity via student fixed effects, but does not include time-varying student covariates or teacher controls. He employs a novel identification strategy; the fraction of boys with female-sounding names in a classroom is used as an instrument for peer behavior. He finds that peer disruptive behavior is associated with both an increased likelihood that a student is suspended and a reduction in achievement test scores.

The current study contributes to the existing stock of peer effects research by providing reliable identification of classroom teachers across a broad range of schooling levels, estimating multiple levels of fixed effects, capturing spillovers from unobserved peer ability, and estimating nonlinear models that reveal heterogeneous peer effects with important policy implications. In addition, we use a large, representative data set that has not previously been employed in the estimation of peer effects.

III. Empirical Model and Identification Method

A. A Value-Added Model of Student Achievement

We begin by specifying a version of a cumulative achievement function with linear-in-means classroom-peer effects, as follows:[2]

$$A_{ijklmt} - A_{ijklm,t-1} = \Delta A_{ijklmt} = \vec{\boldsymbol{\alpha}}_1{}' \vec{\mathbf{X}}_{it} + \boldsymbol{\alpha}_2 u_i + \vec{\beta}_1{}' \vec{\bar{X}}_{\sim ikt} + \beta_2{}' \bar{u}_{\sim ikt} + \vec{\eta}' \vec{T}_{jt} + CS_{kt} + \delta_{jl} + \omega_m + \theta_t + \varepsilon_{it} \quad (1)$$

Equation (1) is a restricted, "value-added" form of the cumulative achievement function specified by Boardman and Murnane (1979) and by Todd and Wolpin (2003),[3] in which we relate the achievement gain, ΔA_{ijklmt}, for individual *i* with teacher *j* in classroom *k* at school *l* in grade level *m* between time *t-1* and time *t,* to the following inputs: a vector, $\vec{\mathbf{X}}_{it}$, of observed (fixed and time-varying) characteristics of individual *i*; a composite of fixed unobserved individual characteristics, u_i, (such as the fixed portion of parental inputs and the student's innate learning potential); the average, $\vec{\bar{X}}_{\sim ikt}$, of the observed (fixed) characteristics of individual *i*'s classroom peers at time *t*; the average, $\bar{u}_{\sim ikt}$, of the unobserved fixed characteristics of current classmates; the observed (time-varying) teacher characteristics, $\vec{T}_{jt}$; class size in classroom *k* at time *t*, CS_{kt}; the effect of the fixed (observed and unobserved) characteristics of teacher *j* at school *l*, δ_{jl}; the fixed effects, ω_m and θ_t respectively, of being in grade level *m* and of the current calendar year *t*, and a time-varying individual disturbance, ε_{it}.

---

[2] We discuss and estimate non-linear peer effects specifications below.

[3] The derivation of the linear education production function in equation (1) from a less restrictive model can be found in Todd and Wolpin (2003) and Sass (2006).

The cumulative achievement specification in equation (1) suits the nature of the outcome measure we observe, which is the Florida Comprehensive Assessment Test-Norm Referenced Test (FCAT-NRT). The test is "vertically scaled," which means that gains from any initial value on the scale are intended to be fully comparable to each other.[4] The model assumes that the cumulative achievement function does not vary with a student's age, although we relax this assumption by estimating separate models for elementary, middle, and high school observations. The model also assumes that schooling inputs applied at any point in time have an immediate and permanent impact on cumulative achievement—in effect, prior learning does not decay or depreciate over time. As a result of these (admittedly strong) assumptions, once-lagged individual achievement serves as a sufficient statistic for all prior schooling inputs and it drops out of the right-hand side of the gain equation. If the no-decay assumption is relaxed, the once-lagged individual score should enter the right-hand-side of the equation, in which case OLS estimation is inconsistent.[5]

To facilitate estimation of simultaneous peer effects, additional assumptions are necessary. Assuming for now that $\vec{\mathbf{X}}_{it}$ contains no time-varying factors, the fixed component of the individual gain score can be written as $\gamma_i = \vec{\boldsymbol{\alpha}}_1{}' \vec{\mathbf{X}}_i + \boldsymbol{\alpha}_2 u_i$. If we assume, in addition, that the relationship between the marginal effect of any given mean peer characteristic is the same multiple of the marginal effect of the characteristic at the individual level, that is, letting

[4] It has been argued that vertical scaling cannot guarantee true comparability of gains (nor of achievement levels) across grade levels (Schafer and Twing 2006). Our schooling-level-specific estimations assume only comparability of gains within a schooling level (e.g., elementary), not across all grade levels.

[5] Of course, if the lagged score ought to enter the gain equation but does not, OLS will be inconsistent due to omitted-variable bias. Most previous studies of peer effects using standardized test scores for elementary and middle school students adopt an equally restrictive specification of the cumulative achievement function. Betts and Zau (2004) relax the no-learning-decay assumption and include once-lagged achievement on the right-hand-side of the gain equation.

$\vec{\beta}_1 = \lambda\vec{\alpha}_1$ and $\beta_2 = \lambda\alpha_2$, we can express the combined impact of average peer observed and unobserved characteristics as $\lambda\bar{\gamma}_{\sim ikt}$, where $\bar{\gamma}_{\sim ikt}$ refers to the average fixed effect for each individual in *i*'s peer group at time *t*, other than herself. This assumption enables us to bundle all of the peer characteristics into a single regressor that represents the mean of the fixed (gain) effects of the individual's current classroom peers. Time-varying individual characteristics can be added back into the model but these are not included in the peer variable. Incorporating these assumptions, the linear-in-means estimation model becomes:

$$A_{ijklmt} - A_{ijklm,t-1} = \Delta A_{ijklmt} = \gamma_i + \lambda\bar{\gamma}_{\sim ikt} + \vec{\alpha}'\vec{X}_{it} + \vec{\eta}'\vec{T}_{jt} + CS_{kt} + \delta_{jl} + \omega_m + \mu_t + \varepsilon_{it} \quad (2)$$

In this model, the individual fixed effect represents a fixed achievement gain or amount of learning per period,.[6] This idiosyncratic learning rate represents the per-period effect on cumulative achievement of the bundle of fixed factors associated with the student. Here we have in mind factors such as the student's innate capacity for learning and the flow of familial monitoring and support. For shorthand we will refer to this effect as student "ability" or "quality." To the extent that family inputs may vary over time, the deviations are embedded in ε_{it} and assumed to be random—specifically, mean zero and i.i.d.—conditional on the vector of regressors.[7] However, we effectively allow for systematic variation in the contributions of

[6] Some specifications of the value-added model assume that the innate ability endowment contributes only to initial achievement and not to ongoing gains, while the family input is modeled as a flow that contributes to gains. However, if there is student-level heterogeneity in gains and if family inputs and ability endowments are not observed it will be impossible to separate the contribution to achievement gains of these different factors. This specification requires only that the combination of student-level unobservables contribute a fixed amount to the expected achievement gain in each period.

[7] Evidence of systematic responses in parental inputs to changes in schooling inputs reveal mixed results. Bonesrønning (2004) finds that class size has a negative effect on parental effort in Norway, suggesting that school

unobserved student-level inputs across schooling levels by estimating separate models for elementary school, middle school, and high-school outcomes.

B. Modeling and Measuring Peer Effects

In light of evidence that teacher quality matters a great deal for student achievement and yet is not strongly linked to observed teacher characteristics (Rockoff 2004; Rivkin et al. 2005; Kane et al. 2006), and evidence that teacher assignments are nonrandom within schools (Oakes 1990; Argys, Rees, and Brewer 1996; Vigdor and Nechyba forthcoming; Feng 2005; Clotfelter et al. 2006), controlling for unobserved teacher inputs would appear to be crucial when measuring classroom-level peer effects. While previous studies have accounted for by-student average teacher quality with student fixed effects, and in some cases for average teacher quality at the school-by-grade level, such controls are likely to be insufficient at the classroom level. For example, if matching of students to teachers with respect to fixed abilities is neither perfectly random nor perfectly deterministic, the average (fixed) ability of classroom peers in a given year will be a better predictor of teacher quality in that year than will be the individual's own ability. If, for example, better teachers are matched on average with better students but there is within-classroom variation in student ability, peer effects estimates will be biased upward when teacher inputs are omitted, even in a model with student fixed effects.[8] Furthermore, observed teacher inputs, such as experience, will constitute inadequate controls if most of the variation in teacher effectiveness derives from unobserved factors.

---

and home inputs are complements. In contrast, Houtenville and Conway (forthcoming) find that parental effort is negatively correlated with school-level per pupil expenditures on instructional personnel, implying that school resources and parental effort are substitutes.

[8] We verify this using simulated data in which teacher ability is positively correlated with the classroom-average ability of her students.

To control both for unobserved student heterogeneity and for unobserved teacher heterogeneity, we employ models with student fixed effects and teacher-school spell effects, plus grade and year controls. We identify peer effects using within-student variation in the distribution of classroom peer quality, isolating the portion of this variation that is not predicted by the teacher–school pair, the grade level, or the school year. This removes the possibility of confounding the effects of within-student peer variation with the effects of within-student teacher (or school, grade level, or year) variation. If teacher identity and peer inputs are perfectly collinear, peer and teacher effects are not separately identified. In such cases our method—which de-means outcomes with respect to the teacher—will yield no identifying variation in the peer variable and therefore will detect no peer effects. Our results indicate that collinearity between the teacher and peer variables is not strong enough to undermine the identification.

Peer effects in our specification represent spillovers from the current peer group's average *fixed (gain) effect*, which we take as a proxy for average "ability" or "quality" among the peer group. In the linear specification, and assuming the coefficient λ in equation (2) is strictly positive, the model says that the greater the average innate learning rate of one's current classroom peers, the greater the individual's achievement gain in the current period, all else equal. The supposition underlying this model is that innate characteristics, aptitudes, motivation levels, and fixed habits, as manifested in students' idiosyncratic learning rates, constitute the main channels by which school peers influence each others' outcomes. For example, students may learn directly from peers based on their high aptitude levels and knowledge of a subject; they may benefit from having well-behaved peers who create a classroom atmosphere that is conducive to learning, or they may free-ride on classmates' questions or superior note-taking skills.

While much of the previous literature takes a similar view, emphasizing spillovers from permanent peer ability rather than from transient, simultaneously determined behavior or outcomes, most of the existing studies measure peer ability on the basis of lagged test scores or various instruments for (current or lagged) test scores, measures that are likely to capture true peer quality with considerable error. [9] Such measurement error will result in downward biases on the estimated peer effects, ignoring other sources of bias. By contrast, our individual fixed effects capture the contribution of both observed and unobserved factors to the idiosyncratic learning rate. A peer variable based on these fixed effects is likely to offer a more accurate gauge of the permanent component of peer ability or quality and so reduce the potential for measurement error bias.

Another advantage of using peer fixed effects is that we avoid the risk of bias caused by regression to the mean, a bias that may affect coefficient estimates on lagged mean peer test scores when the student's own lagged score is omitted from the regression.[10] Because our peer variable represents an average of time-invariant quantities, it does not manifest any one-time shocks to peers' outcomes and will not be subject to this source of bias.

One potential disadvantage of the peer fixed effects is that, because we bundle observed and unobserved characteristics into a fixed peer effect, we will be unable to isolate the effects of race, gender and other fixed observed characteristics. However, recent evidence (Hoxby and Weingarth 2005; Cooley 2007) suggests that race and gender effects serve mainly as proxies for

[9] One exception is Cooley (2007), who emphasizes endogenous effects and uses a control function approach with an exogenous utility shifter in order to avoid simultaneity bias.

[10] As Betts and Zau (2004) explain, if the members of a student's peer group don't change much over time, regression to the mean will cause the individual's current test-score gain to be negatively correlated with her own lagged score as well as with the lagged score of her current peer group; if the student's lagged score is omitted, the estimated coefficient on the lagged mean peer score will be biased downward.

ability, indicating that policy should focus on finding the optimal ability mix rather than the optimal racial mix.

In addition to learning externalities operating through fixed peer characteristics— termed "exogenous effects" in the social interactions literature—there may be spillovers of voluntary behavior across students or "endogenous effects." For example, behaviors may be (at least temporarily) contagious in that a student may adjust her effort level upward in the current period when surrounded by peers with high effort levels. We do not, as in Cooley (2007), model an achievement function in which students choose effort levels simultaneously with a preference for conformity. However, we cannot rule out the possibility that fixed peer characteristics will appear to matter because they proxy for contagious behaviors, such as good study habits or attentiveness in class, and not (only) because peer conduct yields direct benefits. If we (rightly) want to attribute such endogenous effects to peer influence, we can view the empirical cumulative achievement function as a reduced form equation in which the peer variable captures the effect of innovations to individual inputs caused by the peer influence together with any "passive" peer effects. However, because fixed characteristics measure current effort with error, the endogenous-effects component of the coefficient (if positive in fact) will be biased downward. The positive tradeoff is that we don't face simultaneity bias, a risk that arises when using current or lagged peer test scores to proxy for effort or ability.

We think that spillover effects from peers' current *outcomes* (test scores) are in the current context unlikely: even if a student were to seek to match her own gain-score to the mean peer gain-score, observing the gain-score to target would be difficult because the outcomes we observe are achieved simultaneously in a single testing event per year, and because students would have to observe peers' past scores as well in order to calculate gains. (Conformity effects

on test score levels in each time period could induce some conformity effects on test score gains but only imperfectly and the same basic critique applies.) However, as in the case of effort spillovers, we can't reject the possibility that endogenous effects of test scores are bundled with exogenous effects of peer quality: for example, a student may achieve more when surrounded by "better" peers because she learns from them, and/or because better peers have better outcomes and she wishes to match those outcomes.

For illustrative purposes, we have described a model in which peer effects operate linearly through average peer characteristics. Although the linear-in-means model has been the most common specification in the education peer effects literature, recent evidence suggests that the model is misspecified, leading to biased estimates (Hoxby and Weingarth 2005). This is an important development, because only if peer effects are nonlinear can policy interventions result in global welfare (achievement) gains—in the linear-in-means setting, policy effects are all zero-sum. In light of this evidence, we estimate two nonlinear specifications. In the first, the influence of the mean peer effect is allowed to depend on a quintile ranking—lowest 20, middle 60, highest 20 percent—of the individual student's ability (based on her fixed effect) relative to the entire estimation sample. In the second, students are affected by the proportion of classroom peers in each ranking group (e.g., the higher the percentage of high-ranking peers, the greater one's current gains) and these effects in turn may depend on the student's own rank. For example, low-ability students may benefit more from an increase in the proportion of high-ranked peers than would high-ability students. Again, all peer variables are based on the peer fixed effects rather than noisier measures of peer ability. Consistent with previous work in this direction, we find that nonlinear models indicate a rich set of peer effects that cannot be detected in linear-in-means estimation.

C. Controlling for Nonrandom Selection into Peer Groups

So far we have addressed concerns about measurement error of peer quality, bundling of endogenous and exogenous effects, simultaneity of outcomes, mean reversion, and model misspecification. A more basic concern entails the endogeneity of the classroom peer group. With nonrandom peer selection there is concern for whether peer influences of any sort can be distinguished from spurious or correlated effects. Correlated effects arise if individuals in a group are more similar to each other, on average, than to individuals outside the group, or if the group is exposed to a common influence that varies across groups. The constant component of selection is taken care of with individual student fixed effects; identification uses only within-student variation in peer group quality. However, we must also take care that variation in the peer group does not proxy for variation in another relevant factor such as the grade level, the time period, the school, or the teacher.[11]

Referring to equation (2), recall that δ_{jl} is the fixed effect of a given teacher-school combination. The teacher-school "spell" fixed effect allows the combined effect to be nonadditive—for example, some teachers may make more efficient use of a school's resources than others or the same teacher may perform differently at different schools.[12] Assume that δ_{jl} is non-zero conditional on $\vec{T}_{jt}$ and has a non-zero variance both within and across schools. If, on average, higher ability students are matched with higher quality teachers and yet there is some

[11] In models involving endogenous effects, an omitted correlated effect can bias estimation even under perfect random assignment because the correlated factor promotes similarity of outcomes within groups regardless of how the peer group is selected.

[12] The implementation of the spell effects is described in Section III. As explicated in Andrews et al. (2006), the method does not separately identify school and teacher contributions to achievement gains. However, we also estimate models with student and school fixed effects only (rather than student and teacher-school effects) to isolate the impact of the teacher controls on the peer effects estimates.

randomness in teacher assignments, then mean peer ability, $\bar{\gamma}_{\sim ikt}$, will be correlated with the teacher-school input, δ_{jl}, even after conditioning on individual ability, and measured peer effects will be biased upward when teacher quality is not controlled for.[13]

Another concern is that students may be assigned to teachers on the basis of prior shocks to the achievement level that are observed by the school principal but not by the econometrician. If assignments are made on this basis and if shocks to individual gains are serially correlated, teacher quality will be correlated with the error term and estimated teacher effects will be biased. Rothstein (2008) produces evidence of such dynamic sorting for a single cohort of elementary students in North Carolina. He finds that future teachers appear to influence current student achievement gains. However, dynamic student-teacher matching of this sort does not induce any particular correlation between the current error term and our peer variable, because by construction the errors are orthogonal to fixed ability. The students in a given classroom will have similar lagged errors and, with serial correlation, similar current errors, but not—in expectation—similar fixed abilities. In such a setting, the peer variable and the teacher variable no longer co-vary, and omission of teacher effects will not introduce bias in the estimates of peer effects. However, if teacher inputs matter, the precision of peer effects estimates may be reduced considerably when teacher controls are omitted. We have verified these statements by running regressions on simulated data with the appropriate correlation properties.

[13]Evidence suggests that good teachers get "plum" assignments within a school, and this is consistent with our empirical findings. If there is perfect sorting (that is, a fixed, one-to-one map from student type to teacher type), there is no within-student variation in peer quality nor in teacher quality, and student fixed effects will sweep out both teacher and peer effects. If students are perfectly randomly assigned, teacher type will vary within a student but this variation will be orthogonal to variation in the peer group.

Even with no unobserved heterogeneity in teacher inputs, including teacher fixed effects may assist the estimation. A common identifying assumption in the literature is that variation in unobserved student/family inputs over time must be orthogonal to variation in peer group quality. For example, if parents at some point decide to exert greater effort to help their child achieve, they might try to secure a better peer group relative to the previous year, in addition to spending more time helping the student complete homework assignments. Alternatively, parents could adjust inputs—in either direction—in response to observing an improvement in the child's peer group quality.[14] In either case, the peer variable will proxy for unobserved parental inputs and peer effects may be biased in either direction. Teacher controls will mitigate the problem if teacher identity serves as a better proxy for the unobserved input than does the peer variable. Since parents are likely to have greater control over their child's teacher than over the specific classmates she gets, then even if parents choose teachers merely as a proxy for the peer group, this better-proxy condition may hold. If teacher heterogeneity matters, however, teacher effects are not a suitable proxy variable and peer effects estimation with teacher effects may be inconsistent. By estimating models both with and without controls for unobserved teacher heterogeneity (the alternative model still controls for school-level effects), we can analyze results in light of alternative assumptions about the role of teachers.

[14] However, if parents exert extra effort in order to help their child "keep up with the Joneses," this could get classified as a type of peer effect-by-proxy.

IV. Data, Sample Selection and Computational Issues

A. Data

In the present study we make use of a unique panel data set of school administrative records from Florida.[15] The data cover five school years, 1999/00 through 2003/04, and include all public-school students in the state of Florida. Achievement test scores are available for both math and reading in each of grades 3–10, for each of two different achievement tests. One of these tests is the "Sunshine State Standards" Florida Comprehensive Achievement Test (FCAT-SSS), a criterion-based exam designed to test for the skills that students are expected to master at each grade level. The other test is the FCAT Norm-Referenced Test (FCAT-NRT), a version of the Stanford-9 achievement test used throughout the country. We use the FCAT-NRT scores and not the FCAT-SSS scores because only the former are readily comparable across grade levels and students: the FCAT-NRT (like the Stanford-9) scores are "vertically" scaled, such that a one-point increase from one place on the scale should, in theory, represent an equivalent achievement gain to a one-point increase from anywhere else on the scale.

B. Sample Selection

To permit a flexible education production function, we divide the sample into three groups: (1) elementary school observations, used to estimate the model of test score gains for the 4th and 5th grades; (2) middle school data, used to estimate the model for the 6th, 7th, and 8th grades; and (3) high school data, used to estimate the model for the 9th and 10th grades.[16] The drawback of estimating separate models is that we limit the number of gain-score observations per student to

[15] A more detailed description of the data is provided in Sass (2006).

[16] Note that 5th grade scores are used to calculate 6th grade gain-scores, and similarly for 8th grade scores and 9th grade gains.

two in the cases of elementary and high school, and to three per student for middle school.[17] In a small number of cases, students are observed more than twice (or, for middle school, more than three times) because they repeated a grade one or more times.[18] Within each level of schooling, we observe four cohorts, covering the four academic years beginning with 2000/01 and ending with 2003/04. Descriptive statistics within each schooling-level sample are given in table 1.

In addition to linking students and teachers to specific classrooms, our data indicate the (average) proportion of time each student spends in each classroom. Although primary school students typically receive academic instruction from a single teacher in a "self-contained" classroom, this is far from universal. During the periods we observe, in addition to being enrolled in a self-contained class, five percent of elementary school students were enrolled in a separate math course, four percent in a separate reading course, four percent in a separate language arts course, and nearly 13 percent in either a gifted-student or special-education course. We restrict our analysis to students who receive instruction in the relevant subject area (math or reading/language arts) in just a single classroom. At the elementary level, this means that we exclude students enrolled in separate math or reading classes, even if they spend most of their time in the all-purpose classroom. We also exclude elementary students who spend less than one hour per day in the all-purpose class, even if not enrolled in a separate math or reading class—for

[17]Since we are not differencing out student effects, our estimation method will assign fixed effects to students with just a single gain observation ("singletons"); the fixed effect just equals the gain score and the student contributes no identifying variation. In the following analysis we omit such singleton students. While it may seem innocuous to omit these observations, in doing so we also omit them from the peer groups of others. If, for example, the omitted students exert less influence on their peers than do the included students, our peer effects estimates will be biased upwards. On the other hand, including such observations puts downward pressure on estimated peer effects, because among such students any peer influences will be incorrectly attributed to the individual effect. We have run most of the models including singletons, and the peer effects are generally smaller, indicating that the latter bias likely dominates.

[18]Our estimation models include repeater-by-grade indicators to allow for differential achievement gains of students who repeat a grade.

example, students who spend most of their time in the special-education classroom. These exclusions allow us to avoid the problem of determining the proper math or reading peer group, and the proper teacher, for students with nonconventional schedules. [19]

At the middle and high-school levels, we drop students enrolled in more than one course in the subject area pertaining to the given test score (math or reading/language arts). To avoid atypical classroom settings and jointly taught classes, we consider only courses with 10–50 students and with only one "primary instructor" of record. Finally, we eliminate charter schools from the analysis since they may have different curricular emphases, and because student-peer and student-teacher interactions may differ in fundamental ways from those in traditional public schools.

Previous work (Bifulco and Ladd 2006; Sass 2006; and others) has shown that student performance suffers in the first year following a move to a new school. In light of this evidence, we include three measures of student mobility among the set of regressors: the number of schools attended in the current year, and indicators of "structural" and "nonstructural" moves by the student. A structural move is defined as a move in which at least 30 percent of a student's cohort in the same grade at the initial school makes the same move. This variable captures the effects of normal transitions from elementary to middle school and from middle to high school, as well as the impact of significant school rezonings. Correspondingly, a nonstructural move is defined as any change in school attendance between the end of the preceding school year and the current school year that does not satisfy the structural-move condition. This variable captures the

[19] Previous studies lack data on students' complete course enrollments and so cannot exclude on such detailed criteria. Hanushek et al. (2003), remove special-education students altogether, while other studies include all students regardless of special-education or multiple-course status.

impact of moves due to events such as family relocations and parents exercising school choice options.

Time-varying teacher attributes are captured by a set of three dummy variables representing varying experience levels: zero years of experience (first-year teachers), one year of experience, and two to four years of experience. Teachers with five or more years of experience are the omitted category.[20] In addition to the time-varying teacher and student factors, all of the remaining regressors represent fixed effects, which are either estimated directly or accounted for using de-meaned variables, as explained below.

C. Computational Issues

Estimation of the achievement function in (2) is computationally challenging since it includes multiple levels of fixed effects. Combining teacher and school effects into teacher-school spell effects simplifies the estimation considerably, but even with this simplification we must estimate fixed effects for over 200,000 students, plus two or three grade levels and four calendar years within each schooling-level model. Standard fixed effects methods eliminate one effect by de-meaning the data with respect to the variable of interest. Additional effects must then be explicitly modeled with dummy variable regressors. After de-meaning the data by the teacher-school combination, we would be faced with simultaneous estimation of more than 200,000 dummy variables, on average, in any given model.

[20] Most longitudinal studies of student achievement find that the marginal effect of additional teacher experience approaches zero after five years of experience. See, for example, Rockoff (2004), Rivkin et al. (2005), Kane et al. (2006).

To estimate the multiple levels of fixed effects, we adopt an extension of the iterative fixed effects estimator recently proposed by Arcidiacono, et al (2005). Taking deviations from the teacher-school spell means, the achievement equation becomes:

$$\Delta A_{ijklmt} - \overline{\Delta A}_{jl} = (\gamma_i - \bar{\gamma}_{jl}) + \lambda(\bar{\gamma}_{\sim ikt} - \bar{\gamma}_{jl}) + \vec{\alpha}'(\vec{X}_{it} - \overline{X}_{jl}) + \vec{\eta}'(\vec{T}_{jt} - \overline{T}_{jl}) + (\omega_m - \bar{\omega}_{jl}) + (\mu_t - \bar{\mu}_{jl}) + \varepsilon_{it} \quad (4)$$

where $\bar{\gamma}_{jk}$ refers to the mean fixed effect of all students (including student *i*) encountered in the set of observations involving teacher *j* at school *l*—call this set of observations "group *jl*";[21] $\overline{X}_{jl}$ denotes the mean (vector) of the time-varying student characteristics within group *jl*;[22] $\overline{T}_{jl}$ denotes the mean of the teacher experience dummy vector across group *jl*—all observations contributing to $\overline{T}_{jl}$ pertain to the same teacher and the mean is automatically weighted by the proportion of students taught at each level of experience; $\bar{\omega}_{jl}$ and $\bar{\mu}_{jl}$ denote the group *jl* means of the grade level and calendar year dummies, respectively. We assume that the error terms within each group *jl* average out to zero. Subtracting the de-meaned individual effect from both sides and collecting terms yields:

$$\Delta A_{ijklmt} - \overline{\Delta A}_{jl} - \gamma_i = \lambda\bar{\gamma}_{\sim ikt} + \delta\bar{\gamma}_{jl} + \vec{\alpha}'(\vec{X}_{it} - \overline{X}_{jl}) + \vec{\eta}'(\vec{T}_{jt} - \overline{T}_{jl}) + (\omega_m - \bar{\omega}_{jl}) + (\mu_t - \bar{\mu}_{jl}) + \varepsilon_{it} \quad (5)$$

In the above, $\delta \equiv (-1 - \lambda)$. Note that if $\gamma_{\sim it}$ equals $\bar{\gamma}_{jk}$ for a given observation—i.e. if the current average peer type equals the average student type for the teacher-school affiliation—the observation contributes no identifying variation. If this is true universally in the data, the teacher

[21] It is useful to refer to this as a group of observations rather than a group of students, in order to avoid confusion in the calculation of group-level means.

[22] Students observed in multiple time periods with the same teacher-school group enter as two different observations and, in such cases, varying values of the time-varying characteristics for a single student enter the calculation of the group mean.

indicator and the peer variable are perfectly collinear and the peer effects and teacher effects are not identified. Using the teacher/school-demeaning method, estimated peer effects will be zero because peer effects will be swept out with the teacher/school-group mean outcome. Such extreme sorting is empirically unlikely, however, and under such conditions our method will yield conservative estimates of peer effects.[23]

Equation (5) is estimated by ordinary least squares (OLS), using initial guesses for the individual fixed effects, γ_i and $\bar{\gamma}_{jk}$. This produces coefficient estimates which are then used to calculate predicted outcomes and corresponding residuals for each individual. The individual fixed effects estimates are then updated by taking the mean residual for each individual. The parameters are re-estimated using the updated fixed effects, and the process is iterated until the coefficient estimates converge. Standard errors are obtained by bootstrapping.

This method yields results that are only approximately correct, however, because the updating of fixed effects based on the mean residuals is an approximation of the value of the fixed effect that minimizes the sum of squared errors within each iteration. Arcidiacono et al. (2007) provide an exact solution, but we are unable to estimate the mathematically exact model successfully with the Florida data.[24] We can, however, estimate the model under both methods using simulated data. In doing so we find that both methods produce fairly precise estimates

---

[23] Furthermore, the mean peer variable will also be perfectly collinear with the student dummy, because the only way to achieve equality between $\gamma_{\sim it}$ and $\bar{\gamma}_{jk}$ is to have perfectly homogeneous classrooms. If teacher effects are omitted, collinearity between the individual effects and the peer variables remains, and peer effects are still not identified.

[24] The value of a student's fixed effect influences the residuals among her own observations and the residuals for all observations in which she is a member of the peer group. The approximate method sets the value of a given individual's fixed effect taking into account the impact on the residuals of the observations for that individual alone, while the exact method also factors in the impact on the residuals of the observations in which the individual enters the peer group. The difference across the methods in the estimated fixed effects, and therefore in the estimated peer effects, will be greater the stronger are peer effects and the smaller is the average class size.

which are close to the true parameter value under a broad range of conditions. As demonstrated in the Appendix, only when sorting into classrooms on student ability is very strong does the approximate method produce estimates that are significantly different than those obtained with the exact method. As sorting gets either very strong or very weak, adding more noise to the data tends to result in biased estimates regardless of the method used.

In the Florida data we find that classroom-level sorting on student ability (measured by the ratio of average classroom variance in estimated student fixed effects to the variance in estimated student fixed effects across all students) is moderate to low, lying in the range of 0.5 to 0.8 (a value of 1.0 indicates no sorting). For these levels of student sorting, our simulation results indicate that the approximate method produces peer effect estimates that are close in magnitude and never statistically significantly different from those produced by the exact method. Given the moderate amount of sorting in the data, the results are quite robust to noise.

Based on the evidence from our data simulation exercise, we are confident that our results would not be improved significantly using the exact method. Furthermore, when we observe biases in the results on the simulated data, the approximate method tends to underestimate true peer effects whereas the exact method sometimes overestimates peer effects.[25] Thus our estimates of peer effects will be conservative.

[25] It can be shown that the bias under the approximate method relative to the exact method is increasing in the true magnitude of peer effects. For this reason we applied a relatively large peer effect in our simulations. The chosen coefficient on mean peer ability was 0.15, which is greater than most empirical estimates of linear-in-means coefficients and three times as great as the estimate we get in the Florida data using the approximate method.

V. Results

A. Mean Peer Effects

We first discuss results under linear-in-means specifications, in which the peer variable is the mean ``ability" (as measured by our fixed effects) of current (classroom or grade-level) peers, not including the student herself. Table 2 reports coefficient estimates for the covariates of interest under our preferred model specification, in which we account for multiple levels of fixed effects, including teacher-school spell effects. We find positive and highly significant peer effects within every level of schooling and for both reading and math. The magnitude of this effect, however, is generally quite small: for elementary school mathematics, for every one-point increase in the mean peer fixed effect the individual experiences an increase of .044 points in her current gain score. Evaluated at the representative peer group within this sample (with a mean fixed effect of .877), the realized effect would be .0386 points. This is equivalent to 0.0015 of a standard deviation in the achievement gain or about one-fourth the impact of reducing class size by one student. The coefficient is smallest, at .015, for elementary reading, and greatest, at .069, for middle-school reading. Counter to a standard presumption in the literature, effects are not systematically smaller in middle school or high school than elementary school, despite the fact that students experience multiple peer groups during the day in the higher grades. We suspect this finding relies on an accurate identification of classroom peers for the given subject.

Notice that the signs on most of the time-varying regressors are as we would expect: achievement gains decline with the number of schools attended in a year (but results are significant only for high school math and elementary school reading). Nonstructural moves between years are associated with greater achievement gains, perhaps because of parental self-selection into optimal learning environments for their children. Larger class size has a uniformly

negative impact on outcomes, and the effect is significant in elementary school for both math and reading, and for middle-school reading. Notice also that within-teacher variation in experience has little significant effect on outcomes, consistent with findings of Rivkin et al. (2005) and Harris and Sass (2008).

Table 3 gives results for a model that is similar to that reported in table 2, but in which data from elementary and middle school are pooled and in which we restrict the minimum observations per student to three. For math achievement, the estimated peer effect agrees strongly with the effects estimated under the separate elementary and middle-school models, which also closely resemble each other. The remaining coefficient estimates are qualitatively similar, in terms of direction and significance, between the combined model and the separate models, with the exception of the effect of number of schools attended during the year, which becomes negative and significant in the pooled model. This latter result may simply reflect greater variation in the number of schools attended per year when students are observed over a longer time period. For reading achievement the peer effect becomes very small and insignificant, although the remaining effects appear qualitatively robust.

Table 4 shows how different model specifications influence the peer effects estimates.[26] The peer effects coefficients in the first row correspond to those reported in table 2, from our preferred specification. The coefficients in the second row come from a model in which we do not control for unobserved teacher effects; we use school fixed effects rather than teacher-school spell effects, but the models are otherwise identical. The third row reports estimated peer effects when the peer group is defined as all others in the same grade level (within the same school and

[26]Unless otherwise specified, all models reported in this table include the same set of controls as the baseline model presented in Table 2.

year), but where the specification is otherwise identical to the preferred one. (In this case we need only control for average teacher quality at the school-by-grade level, doing so through the combination of school level and grade level fixed effects.)

The first pattern to note is that estimated peer effects are generally larger and less precise in the absence of controls for unobserved teacher inputs. Considering the middle-school math results, the estimated peer effect without the teacher controls, 0.228, is more than 5 times the size of the estimated effect with the controls. For elementary math, the point estimate is also greater when teacher controls are omitted, but the coefficient is not significant in that case. For reading achievement, the differences are less stark, but at the elementary level the estimated peer effect is significantly greater when teacher controls are omitted. The results suggest that there may be a significant positive correlation between peer ability and teacher quality, even after controlling for individual ability, and that such a correlation could distort estimates of peer effects in non-experimental data when unobserved teacher inputs are not taken into account.

The second important finding is that linear-in-means peer effects among grade-level-at-school peers are always insignificant, with point estimates close to zero. Taking these results at face value, a natural interpretation is that the classroom setting facilitates learning spillovers in a way that non-classroom interactions do not. Taking a skeptical view, however, one might argue that we are also more likely to find spurious peer effects at the classroom level than at the grade level due to classroom-level sorting. In the discussion below we consider potential sources of residual bias and whether these might be stronger at the classroom level than the grade-within-school level.

B. Nonlinear Peer Effects

As a first step in relaxing the linear-in-means specification, we allow the peer effect to depend on the mean and the standard deviation of peer ability, again measured by fixed effects estimated within the model. As seen in table 5, greater dispersion in peer ability is associated with a significant, negative effect on math achievement gains for both middle school and high school students. Otherwise, no significant effects of the standard deviation are found. One interpretation of these results would be that it is difficult to effectively teach student groups with diverse math ability (although not so for diverse verbal ability). The effects of mean peer ability in the current model are largely similar to those obtained from the linear-in-means model. Where the dispersion effects are significant, the results imply that imposing a mean-preserving spread of classroom ability will reduce average classroom achievement gains.

While we find no significant impact of ability variance at the elementary level, Vigdor and Nechyba found that ability dispersion had a positive impact on test scores (in levels) among 5th graders in their North Carolina sample. However, Duflo et al. (2008) found that, among students in Kenyan primary schools randomly assigned to institute a tracking policy, test score gains on a combined math/literacy exam were greater than they were among students in the untreated control group of schools, though only the math score results were individually significant at conventional levels. In the same experiment, however, they found no spillover effects of mean peer ability. Being the best student in a class of relatively low-achieving students or being the worst student among a class of relatively high-ability students made no difference. Duflo et al. argue their results imply that students benefit from classroom homogeneity because the teacher can better tailor her instruction to students' needs.

In the second nonlinear specification, we allow peer effects to depend on the student's own ability—defined by the ranking of her fixed effect within the sample population.[27] For a given distribution of student fixed effects within the sample, a student is designated as a "low" type if her fixed effect falls within the bottom quintile of the population distribution, as a "middle" type if her effect lies between the 20th and 80th percentiles, and as a "high" type if she falls in the top quintile. (The iterative model updates the rankings each time the fixed effects values are updated.) We therefore include three peer variables: "Lowest Ability Quintile × Mean Peer FE," "Middle 3 Ability Quintiles × Mean Peer FE," and "Highest Ability Quintile × Mean Peer FE," where the type variables are binary indicators. As in the linear-in-means model, identification of the peer effects relies on variation in peer quality within a student over time as well as on variation in the student quality distribution across different sections taught by the same teacher.[28]

Table 6 reports the type-specific peer effect coefficients for each of the six subject-by-schooling level models. Among the elementary-level results, all effects are highly significant except those pertaining to reading outcomes among high-ranked students, and the significant effects are all much larger than the estimated effects from the linear-in-means models. These results imply that the average treatment effect (for either math or reading, taken across student ranks) is significantly greater than that estimated under the corresponding linear-in-means specification. This discrepancy is made possible by the fact that individual fixed effects, as well

[27] Due to the computational costs of estimating the non-linear models, we estimate only our preferred specification (including teacher effects, no singleton students, classroom level effects). We assume that the impact of changes in specification would be similar qualitatively to the impact on the results in the linear models.

[28] The mean value by teacher of a given peer variable—for example, "Lowest Ability Quintile × Mean Peer FE" is the average value of the mean peer fixed effect variable among all low-type students taught by the given teacher, weighted by the proportion of all of the teacher's students that were low types. Recall that teacher groups are specific to a single school.

as the peer variables, are estimated anew within the context of each specific model. The linear-in-means model, by disallowing type-specific effects, likely attributes a greater portion of the outcome variation to individual effects as opposed to peer effects, since the (omitted) interaction variables are correlated with individual type.

The elementary school results also indicate that the lowest-ranked students appear to receive the greatest benefits from having higher-quality peers, but middle-ranked students also receive sizable benefits. For example, low-ranked elementary students will experience a .82 point (0.03 standard deviation) boost to their math gain score for every 1 point increase in the mean peer fixed-effect variable, whereas high-ranked students will receive only a .10 point increase in the math gain score under the same marginal treatment. These results provide a strong argument in favor of distributing top students relatively evenly across classrooms at the elementary level rather than isolating them from other students. Put differently, if the objective is to maximize total learning gains it would appear preferable to have evenly mixed groups rather than ability-tracked groups.

At the middle school level, estimated treatment effects are smaller than at the elementary level, but again the average treatment effects are larger than those estimated under the linear models. Results do not differ much between reading and math—in both cases, middle-ability students experience the greatest benefits from a peer quality improvement. Based just on the point estimates, the highest-ranked students appear to experience larger peer effects than the lowest-ranked students, and effects on math scores for low-ranked students are only marginally significant (the p-value is 0.109). The findings argue for moving the best students to "middling" classrooms rather than to the weakest classrooms, and also argue against strict tracking.

At the high school level, there are fewer significant effects but the point estimates are close to those for middle school in most cases. As in the middle school results, effects are strongest for middle-ranked students. Unlike middle school, however, the estimates suggest a negative effect on the best math students of having higher-quality peers; results are weakest for high-school reading, with no significant effects found for either low- or high-ranking types. Unlike the linear-in-means case, we find an attenuation of the rank-specific peer effects between elementary school and the upper grades. If we put more faith in the nonlinear model than the linear model, we should conclude that the stakes for classroom peer assignments are greater in elementary school than in either middle school or high school, although they are not insignificant for the latter cases.

The two other existing studies that allow for nonlinear peer interactions at the classroom level, Cooley (2007) and Hoxby and Weingarth (2005) also find for elementary school students that low and middle ability students benefit more from an improvement in peer quality than do higher-ability students. However, Hanushek et al. (2003), using school-by-grade level data did not find many significant differences in measured peer effects across the achievement distribution.

To allow for an even more complex set of peer influences, we estimate a two-way interaction model, similar to that of Hoxby and Weingarth (2005). In this model, each student type (low-rank or bottom 20 percent of the sample-wide fixed effects distribution, mid-rank or middle 3/5th, high-rank or top 20 percent) is subject to peer effects from three different moments of the peer distribution: the proportions of low-ranked, mid-ranked, and top-ranked peers. To estimate these effects, we construct six peer variables, each the product of a binary type indicator and the proportion of peers of a given rank: for example, “Individual in Lowest Quintile

× Fraction of Peers in Lowest Quintile," "Individual in Lowest Quintile × Fraction of Peers in Highest Quintile," and similarly for mid-ranked and high-ranked individuals. Due to collinearity of the proportions within any student observation, we omit the effect of the proportion of mid-ranked peers on each individual type. Therefore, the marginal effect (on a given individual type) of an increase in the proportion of high-ranked peers (alternatively, low-ranked peers) represents the net effect of increasing the high-ranked (low-ranked) proportion and reducing the mid-ranked proportion, since the low-ranked (high-ranked) proportion and class size are being held constant.

As seen in table 7, all of the effects are highly significant and many are of a large magnitude. Consider, for example, the effect of the fraction of lowest-ability-quintile peers on lowest-ability-quintile individuals for elementary-school math scores. The coefficient estimate means that an (additive) increase of one-tenth of a unit in the fraction of lowest-quintile peers (and a corresponding decrease in the fraction of peers in the middle three quintiles) will raise the math test-score gain of the lowest-ability students by approximately 2 points (0.08 of a standard deviation in achievement gains). While this result seems to point in favor of ability tracking, low ability students get an even greater boost from an increase in the fraction of peers in the top ability quintile. Across schooling levels and disciplines, low ability students benefit about twice as much from an increase in the share of top-quality peers as they do from an increase in the share of low ability peers. Effects are strongest at the elementary level, for both math and reading achievement, but effects are not weaker in high school than they are in middle school. Such students apparently perform less well the greater the share of middling students they are grouped with.

Students in the middle three quintiles benefit from having a higher share of high ability peers but suffer losses as the share of low ability peers increases. In most cases these respective

gains and losses are of roughly equal magnitude, although in a few cases the gains appear slightly smaller. In high school math, for example, middle-ability students would prefer to replace a low-ability student with a middle-ability student than to replace a middle-ability student with a high-ability student. Of course, replacing a low-ability student with a high-ability student would dominate either of these options. As in the case of the lowest-ability students, effects are strongest at the elementary level, but roughly equal between middle school and high school.

Students in the highest ability quintile appear to benefit most from having peers of middle ability rather than peers of either high ability or low ability, but the losses are greatest as the share of low ability peers increases. Again the effects are greatest at the elementary school level, although at the high school level we observe a relatively large negative impact on math achievement gains as high ability students get more low-ability peers.

C. Policy experiments

Table 8 shows the impact of three different classroom assignment experiments. In each case, the initial classroom ability distribution is assumed to be representative of the aggregate rankings, with 20 percent of students in the lowest-ability quintile, 60 percent in the middle three quintiles, and 20 percent in the highest quintile. In the first reassignment, the class becomes heavily weighted toward low ability students, and the new respective shares are 60 percent, 30 percent, and 10 percent. The table shows the impact on the students remaining in the same classroom, by ability level. The lowest-ability students are made better off, but these gains are more than offset by the losses to middle and high ability students. In the second experiment, the class becomes dominated by high ability students, with respective shares of 10, 30, and 60 percent. In this case, low ability students benefit by a large margin, middle ability students benefit modestly, and high

ability students are made somewhat worse off. In the third experiment the distribution becomes shifted toward the middle, with only 5 percent each in the lowest and highest ability quintiles. The net effects are all close to zero, although they are positive in some cases and negative in others.

These results do not represent general equilibrium effects—that is, they do not consider the impact on the students who were "exported" from a given classroom. In addition, students imported from other classrooms would experience different impacts if their initial assignment differed from the initial assignment assumed in the experiment. While it appears that net benefits accrue in the second experiment, not all low- and middle- ability students can be assigned to classrooms dominated by high ability students.

In table 9 we consider the impact of a hypothetical school choice program which leads to the exit of 2.5 percent of students from each classroom, where all of these students were in the highest ability quintile. (Again we assume that the initial ability distribution was representative of the aggregate distribution.) The net effect is negative but quite small, and the highest-ability students experience very small gains. Again, any gains or losses experienced by the exiting students are not considered. The findings suggest that the effects of school choice programs on those "left behind" are likely to be small.

VI. Summary and Conclusions

This paper adds to a growing list of studies that use matched panel data in direct tests for peer effects in academic achievement. As in earlier studies, the panel data facilitate the identification of peer effects on academic achievement by enabling some degree of control for endogenous variation in peer groups. Unlike many earlier studies, we are able to place students within

classroom groups with specific teachers, and we observe each teacher with more than one group of students. Accordingly, ours is the first study to control simultaneously for unobserved heterogeneity in both student ability and in teacher effectiveness, among other unobserved effects, and the first to estimate classroom-level peer effects at the elementary, middle, and high-school levels for the same school system and to compare these to grade-level effects. While not the first to do so, we add further value by adopting an innovative computational technique which aims both to facilitate fixed-effects estimation and to minimize measurement error in peer ability, and we estimate nonlinear peer effects models that allow for non-zero-sum policy implications.

We find significant peer effects only at the classroom level and not at the general grade level, a result that emphasizes the importance of identifying the salient peer group. We also find that estimated peer effects are generally weaker when we control for unobserved inputs at the teacher-school level. This result indicates that teacher ability may vary systematically with peer ability conditional on individual student ability. Such co-movement is plausible in the context of student-teacher matching policies that result in a positive but imperfect correlation between students' and teachers' fixed abilities. These findings suggest that accessing random within-student variation in peer ability will not guarantee unbiased estimates of peer effects when unobserved teacher effects are not also accounted for.

We find that peer effects are not "one-size-fits-all," but rather exhibit striking differences across students of different abilities and across different segments of the peer ability distribution. For example, the weakest students appear to experience the biggest positive impact from having higher quality peers. At the same time, however, such benefit appears to derive specifically from having peers in the highest quintile of the ability distribution. High ability students appear to experience the weakest spillovers from mean peer ability, but nonetheless may suffer sharp

losses due to an increase in the share of peers of very low ability. The sizable effects observed in the nonlinear models are obscured in the linear-in-means models, within which we find only very modest, but positive, spillovers from mean peer ability. Furthermore, comparisons of effects between math and reading scores, and across different schooling levels, also depend on whether linear or nonlinear models are employed.

Considering the more nuanced results of the nonlinear models, the policy recommendations are not clear cut. For example, while low-ability students appear to benefit significantly from having top-quality peers, those peers will experience reductions in achievement gains from mixing with students of very low ability, reductions that may fully offset the weaker students' gains. On the other hand, policies that mix middle and high ability students with each other are likely to strictly dominate those that segregate the top students in a separate track. While parents may prefer strict tracking, our results indicate that the highest-ability students actually benefit from mixing with students of middling ability. We also find that any negative peer effects from school choice programs are likely to be small. A choice program that attracted 2.5 percent students, all of them from the top ability quintile, would have only very small negative effects on the learning gains of lower ability student who remain behind.

References

Andrews, Martyn, Thorsten Schank, and Richard Upward. 2006. "Practical Fixed Effects Estimation Methods for the Three-Way Error Components Model." 6 *Stata Journal*, 461-481.

Angrist, Joshua D. and Kevin Lang. 2004. "Does School Integration Generate Peer Effects? Evidence from Boston's Metco Program." 94 *American Economic Review* 1613-1634.

Arcidiacono, Peter, Gigi Foster, Natalie Goodpaster, and Josh Kinsler. 2005. "Estimating Spillovers in the Classroom with Panel Data." unpublished manuscript.

———. 2007. "Estimating Spillovers Using Panel Data, with an Application to the Classroom." unpublished manuscript.

Argys, Laura M., Daniel I. Rees, and Dominic J. Brewer. 1996. "Detracking America's Schools: Equity at Zero Cost?" 15 *Journal of Policy Analysis and Management* 623-645.

Betts, Julian R., and Andrew Zau. 2004. "Peer Groups and Academic Achievement: Panel Evidence from Administrative Data." unpublished manuscript.

Bifulco, Robert, and Helen F. Ladd. 2006. "The Impacts of Charter Schools on Student Achievement: Evidence from North Carolina." 1 *Education Finance and Policy* 50-90.

Boardman, Anthony E., and Richard J. Murnane. 1979. "Using Panel Data to Improve Estimates of the Determinants of Educational Achievement." 52 *Sociology of Education* 113-121.

Bonesr●nning,[[AU: symbol ok?]] Hans. 2004. "The Determinants of Parental Effort in Education Production: Do Parents Respond to Changes in Class Size?" 23 *Economics of Education Review* 1-9.

Carman, Katherine, and Lei Zhang. 2008. "Classroom Peer Effects and Academic Achievement: Evidence from a Chinese Middle School." unpublished manuscript.

Carrell, Scott E., Richard L. Fullerton, and James E. West. 2008. "Does Your Cohort Matter? Measuring Peer Effects in College Achievement." unpublished manuscript.

Caucutt, Elizabeth. 2002. "Educational Vouchers When there are Peer Group Effects -- Size Matters." 43 *International Economic Review*, 195-222.

Clotfelter, Charles T., Helen F. Ladd and Jacob L. Vigdor. 2006. "Teacher-Student Matching and the Assessment of Teacher Effectiveness." 41 *Journal of Human Resources* 778-820.

Cooley, Jane. 2007. "Desegregation and the Achievement Gap: Do Diverse Peers Help?" unpublished manuscript.

Ding, Weili, and Steven F. Lehrer. [[AU: Year?]]"Do Peers Affect Student Achievement in China's Secondary Schools?" 89 *Review of Economics and Statistics*, 300-312.

Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2008. "Peer Effects and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya." unpublished manuscript.

Epple, Dennis, Elizabeth Newlon, and Richard Romano. 2002. "Ability Tracking, School Competition and the Distribution of Educational Benefits." 83 *Journal of Public Economics*, 1-48.

Epple, Dennis, and Richard E. Romano. 1998. "Competition between Private and Public Schools, Vouchers, and Peer-Group Effects." 88 *American Economic Review* 33-62.

Feng, Li. 2005. "Hire Today, Gone Tomorrow: The Determinants of Attrition among Public School Teachers." Unpublished manuscript.

Figlio, David N. 2005. "Boys Named Sue: Disruptive Children and Their Peers." NBER working paper #11277. Cambridge, MA: NBER.

Fryer, Roland G., and Paul Torelli. 2005. "An Empirical Analysis of 'Acting White.'" NBER working paper #11334. Cambridge, MA: NBER.

Hanushek, Eric A., John F. Kain, Jacob M. Markham, and Steven G. Rivkin. 2003. "Does Peer Ability Affect Student Achievement?" 18 *Journal of Applied Econometrics* 527-544.

Harris, Douglas N., and Tim R. Sass. 2008. "Teacher Training, Teacher Quality and Student Achievement." unpublished manuscript.

Houtenville, Andrew J., and Karen S. Conway. Forthcoming. "Parental Effort, School Resources and Student Achievement." *Journal of Human Resources*.

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

Hoxby, Caroline M., and Gretchen Weingarth. 2005. "Taking Race Out of the Equation: School Reassignment and the Structure of Peer Effects." unpublished manuscript.

Kane, Thomas J., Jonah E. Rockoff, and Douglas O. Staiger. 2006. "What Does Certification Tell Us About Teacher Effectiveness? Evidence from New York City." NBER working Paper #12155. Cambridge, MA: NBER.

Lai, Fang. 2008. "How Do Classroom Peers Affect Student Outcomes? Evidence from a Natural Experiment in Beijing's Middle Schools." unpublished manuscript.

Oakes, J. 1990. *Multiplying Inequalities: The Effects of Race, Social Class, and Tracking on Opportunities to Learn Mathematics and Science.* Thousand Oaks, CA: RAND.

Rivkin, Steven G., Eric A. Hanushek, and John F. Kain. 2005. "Teachers, Schools and Academic Achievement." 73 *Econometrica*, 417-458.

Rockoff, Jonah E. 2004. "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data." 94 *American Economic Review*, 247-252.

Rothstein, Jesse. 2008. "On the Identification of Teacher Quality: Fixed Effects, Tracking and Causal Attribution." unpublished manuscript.

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

Sass, Tim R. 2006. "Charter Schools and Student Achievement in Florida." 1 *Education Finance and Policy* 91-122.

Schafer, William D., and Jon S. Twing. 2006. "Growth Scales and Pathways." In *Longitudinal and Value-Added Models of Student Performance,* edited by R. W. Lissitz ([[AU: pages?]]). [[AU: city?]]: JAM Press.

Todd, Petra E., and Kenneth I. Wolpin. 2003. "On the Specification and Estimation of the Production Function for Cognitive Achievement." 113 *Economic Journal* F3-F33.

Vigdor, Jacob L., and Thomas S. Nechyba. Forthcoming. "Peer Effects in North Carolina Public Schools." In *Schools and the Equal Opportunity Problem*, edited by P.E. Peterson and L. Wößmann ([[AU: pages?]]. Cambridge, MA: MIT Press.

Zabel, Jeffrey E. 2008. "The Impact of Peer Effects on Student Outcomes in New York City Public Schools." 3 *Education Finance and Policy* 197-249.

Table 1.
Mean Values for Florida Public School Students, 1999/2000–2003/2004

	Math			Reading		
	Elementary (Grades 3-5)	Middle (Grades 6-8)	High School (Grades 9-10)	Elementary (Grades 3-5)	Middle (Grades 6-8)	High School (Grades9-10)
Achievement Gain	20.246	14.398	11.772	16.656	15.965	-2.616
Std. Dev. of Achiev. Gain	25.561	23.846	25.637	26.313	25.540	25.306
Number of Schools Attended	1.040	1.038	1.025	1.040	1.034	1.027
"Structural" Mover	0.011	0.227	0.315	0.011	0.193	0.403
"Nonstructural" Mover	0.118	0.157	0.162	0.117	0.141	0.192
Class Size	25.797	27.322	27.931	25.803	26.764	27.795
Teacher Experience	10.601	9.882	11.217	10.611	9.685	10.476
Mean Peer Discipline Incid.$_{t-1}$	0.087	0.452	0.508	0.087	0.430	0.565

Table 2.
Estimates of the Determinants of Math and Reading Achievement Gains in Florida, 1999/2000–2003/2004

	Math			Reading		
	Elementary (Grades 4-5)	Middle (Grades 6-8)	High School (Grades 9-10)	Elementary (Grades 4-5)	Middle (Grades 6-8)	High School (Grades9-10)
Mean Peer Fixed Effect	0.0437**	0.0426**	0.0577**	0.0147**	0.0688**	0.0444**
	(0.0079)	(0.0153)	(0.0106)	(0.0036)	(0.0123)	(0.0131)
Number of Schools Attended	-0.3417	-0.7271	-1.0996*	-1.0135**	-0.1699	-0.5348
	(0.4189)	(0.4717)	(0.5492)	(0.3390)	(0.3788)	(0.6349)
Structural Mover	-1.3524	-0.5066	1.6544**	0.1704	-0.3706	-0.2692
	(1.0721)	(0.3755)	(0.3825)	(1.1266)	(0.3281)	(0.5335)
Nonstructural Mover	0.6945**	0.0135	2.3373**	0.7627**	0.4940	0.0609
	(0.2484)	(0.2873)	(0.3338)	(0.2424)	(0.2776)	(0.5409)
Class Size	-0.1633**	-0.0276	-0.0192	-0.0950**	-0.0524**	-0.0291
	(0.0293)	(0.0141)	(0.0141)	(0.0336)	(0.0127)	(0.0194)
Teacher with 0 Years of Experience	-1.3884	-0.5084	-0.9803	-1.3472	-0.1337	0.3947
	(1.2883)	(1.0217)	(1.3099)	(1.0615)	(1.0481)	(1.7568)
Teacher with 1-2 Years of Experience	-0.4849	0.7043	-0.1907	0.5584	0.2751	1.0727
	(1.0377)	(0.9035)	(1.1180)	(0.8400)	(0.8690)	(1.4678)
Teacher with 3-4 Years of Experience	0.1368	0.8509	0.1465	0.6451	-0.2075	0.8317
	(0.9527)	(0.6812)	(0.8885)	(0.7824)	(0.6975)	(1.3659)
Teacher with 5-9 Years of Experience	-0.2999	1.0246*	0.1462	0.7859	-0.2151	-0.2048
	(0.6265)	(0.4700)	(0.6751)	(0.7237)	(0.5828)	(1.0451)
Number of Students	263,241	204,668	202,882	263,882	268,097	154,487
Number of Observations	534,430	446,878	445,456	535,769	599,284	311,056

Models also include year, grade level, and repeater-by-grade indicators. Bootstrapped standard errors are in parentheses. * indicates significance at the .05 level and ** indicates significance at the .01 level in a two-tailed test.

Table 3.
Estimates of the Determinants of Math and Reading Achievement Gains in Florida, 1999/2000–2003/2004
(Minimum of 3 Observations per Student)

	Math	Reading
	Elementary/Middle (Grades 4-8)	Elementary/Middle (Grades 4-8)
Mean Peer Fixed Effect	0.0444**	0.0078
	(0.0119)	(0.0137)
Number of Schools Attended	-0.8534**	-1.0307**
	(0.3297)	(0.3203)
Structural Mover	0.3820	0.0662
	(0.3057)	(0.2748)
Nonstructural Mover	0.6373*	0.6537**
	(0.2505)	(0.2187)
Class Size	-0.0572**	-0.0424**
	(0.0129)	(0.0125)
Teacher with 0 Years of Experience	-0.4069	-0.1375
	(0.8879)	(1.0519)
Teacher with 1-2 Years of Experience	0.5706	0.3266
	(0.7541)	(0.8788)
Teacher with 3-4 Years of Experience	0.1740	0.4233
	(0.5981)	(0.7653)
Teacher with 5-9 Years of Experience	0.0204	0.6858
	(0.4514)	(0.4717)
Number of Students	159,664	189,711
Number of Observations	508,763	609,758

Models also include year, grade level, and repeater-by-grade indicators. Bootstrapped standard errors are in parentheses. * indicates significance at the .05 level and ** indicates significance at the .01 level in a two-tailed test.

Table 4.
Comparison of Estimated Effects of Mean Peer Fixed Effects on Math and Reading Achievement Gains in Florida From Models with Varying Peer Group Levels and Varying Teacher Controls, 1999/2000–2003/2004

	Math			Reading		
Peer Group/ Teacher Controls	Elementary (Grades 4-5)	Middle (Grades 6-8)	High School (Grades 9-10)	Elementary (Grades 4-5)	Middle (Grades 6-8)	High School (Grades9-10)
Classroom Peers/ With Teacher FE	0.0437**	0.0426**	0.0577**	0.0147**	0.0688**	0.0444**
	(0.0079)	(0.0153)	(0.0106)	(0.0036)	(0.0123)	(0.0131)
Classroom Peers/ No Teacher FE	0.1401	0.2280**	0.0256*	0.0723*	0.0903**	0.0678**
	(0.0993)	(0.0412)	(0.0118)	(0.0364)	(0.0207)	(0.0205)
Grade Level Peers/ With Teacher FE	-0.0021	-0.0010	0.0152	-0.0009	-0.0004	0.0025
	(0.0015)	(0.0014)	(0.0137)	(0.0009)	(0.0007)	(0.0017)
Number of Students	263,241	204,668	202,882	263,882	268,097	154,487
Number of Observations	534,430	446,878	445,456	535,769	599,284	311,056

Models include number of schools attended, structural and nonstructural mover indicators, class size, teacher experience indicators, and year, grade level, and repeater-by-grade indicators. Bootstrapped standard errors are in parentheses. * indicates significance at the .05 level and ** indicates significance at the .01 level in a two-tailed test.

Table 5.
Estimates of the Effects of Mean Classroom Peer Fixed Effects and the Effects of the Standard Deviation in Classroom Peer Effects on Math and Reading Achievement Gains in Florida, 1999/2000–2003/2004

	Math			Reading		
	Elementary (Grades 4-5)	Middle (Grades 6-8)	High School (Grades 9-10)	Elementary (Grades 4-5)	Middle (Grades 6-8)	High School (Grades 9-10)
Mean Peer Fixed Effect	0.0157*	0.0424*	0.0679**	0.0140**	0.0184	0.0473**
	(0.0069)	(0.0191)	(0.0095)	(0.0053)	(0.0151)	(0.0115)
Standard Deviation of Peer Fixed Effects	-0.0085	-0.0315*	-0.0491**	-0.0075	-0.0239	0.0062
	(0.0077)	(0.0139)	(0.0109)	(0.0081)	(0.0168)	(0.0093)
Number of Students	263,241	204,668	202,882	263,882	268,097	154,487
Number of Observations	534,430	446,878	445,456	535,769	599,284	311,056

Models include number of schools attended, structural and nonstructural mover indicators, class size, teacher experience indicators, and year, grade level, and repeater-by-grade indicators. Bootstrapped standard errors are in parentheses. * indicates significance at the .05 level and ** indicates significance at the .01 level in a two-tailed test.

Table 6.
Estimates of the Effect of Mean Classroom Peer Fixed Effects by Own Ability Level on Math and Reading Achievement Gains in Florida, 1999/2000–2003/2004

	Math			Reading		
	Elementary (Grades 4-5)	Middle (Grades 6-8)	High School (Grades 9-10)	Elementary (Grades 4-5)	Middle (Grades 6-8)	High School (Grades9-10)
Lowest Ability Quintile × Mean Peer Fixed Effect	0.8207** (0.0309)	0.1052 (0.0656)	0.0670 (0.0915)	0.7703** (0.0378)	0.0796** (0.0335)	0.1011 (0.0957)
Middle 3 Ability Quintiles × Mean Peer Fixed Effect	0.6081** (0.0239)	0.2138** (0.0186)	0.2121** (0.0211)	0.5043** (0.0238)	0.2038** (0.0164)	0.1922** (0.0203)
Highest Ability Quintile × Mean Peer Fixed Effect	0.1005** (0.0018)	0.1423** (0.0294)	-0.0752** (0.0170)	-0.0108 (0.0309)	0.0994** (0.0222)	0.1004 (0.0809)
Number of Students	263,241	204,668	202,882	263,882	268,097	154,487
Number of Observations	534,430	446,878	445,456	535,769	599,284	311,056

Models include number of schools attended, structural and nonstructural mover indicators, class size, teacher experience indicators, and year, grade level, and repeater-by-grade indicators. Bootstrapped standard errors are in parentheses. * indicates significance at the .05 level and ** indicates significance at the .01 level in a two-tailed test.

Table 7.
Estimates of the Effects of Peer Ability Level by Own Ability Level on Math and Reading Achievement Gains in Florida, 1999/2000–2003/2004

	Math			Reading		
	Elementary (Grades 4-5)	Middle (Grades 6-8)	High School (Grades 9-10)	Elementary (Grades 4-5)	Middle (Grades 6-8)	High School (Grades9-10)
Lowest Quintile × Fraction of Peers in Lowest Quintile	20.540** (1.791)	5.302** (0.972)	5.894** (0.948)	19.901** (2.166)	6.513** (1.116)	7.595** (1.708)
Lowest Quintile × Fraction of Peers in Highest Quintile	40.556** (1.517)	10.369** (1.374)	13.423** (1.221)	38.931** (1.499)	11.262** (1.188)	12.034** (1.704)
Mid. 3 Quintiles × Fraction of Peers in Lowest Quintile	-14.363** (1.406)	-4.114** (0.610)	-4.948** (0.665)	-14.916** (1.215)	-3.443** (0.6465)	-2.288* (1.000)
Mid. 3 Quintiles × Fraction of Peers in Highest Quintile	14.901** (1.128)	3.207** (0.663)	2.606** (0.614)	11.367** (1.283)	3.891** (0.631)	2.393* (0.970)
Highest Quintile × Fraction of Peers in Lowest Quintile	-39.542** (1.686)	-11.035** (1.056)	-20.494** (1.265)	-42.471** (1.516)	-8.304** (1.298)	-12.998** (1.599)
Highest Quintile × Fraction of Peers in Highest Quintile	-18.352** (2.201)	-3.914** (0.9700)	-9.669** (1.092)	-28.880** (1.701)	-3.035* (1.134)	-7.248** (1.727)
Number of Students	263,241	204,668	202,882	263,882	268,097	154,487
Number of Observations	534,430	446,878	445,456	535,769	599,284	311,056

Models include number of schools attended, structural and nonstructural mover indicators, class size, teacher experience indicators, and year, grade level, and repeater-by-grade indicators. Bootstrapped standard errors are in parentheses. * indicates significance at the .05 level and ** indicates significance at the .01 level in a two-tailed test.

Table 8.
Estimated Effects of Alternative Classroom Assignments on Student Math and Reading Achievement in Florida by Student Ability Ranking, 1999/2000–2003/2004

	Math			Reading		
	Elementary (Grades 4-5)	Middle (Grades 6-8)	High School (Grades 9-10)	Elementary (Grades 4-5)	Middle (Grades 6-8)	High School (Grades9-10)
Change from (20 pct. in lowest quintile, 60 pct in middle 3 quintiles and 20 pct. in top quintile) to (60 pct. in lowest quintile, 30 pct. in middle 3 quintiles and 10 pct. in highest quintile)						
Lowest Quintile	4.160	1.084	1.015	4.067	1.479	1.835
Middle 3 Quintiles	-7.235	-1.966	-2.240	-7.103	-1.766	-1.155
Highest Quintile	-13.982	-4.023	-7.231	-14.100	-3.018	-4.474
Change from (20 pct. in lowest quintile, 60 pct in middle 3 quintiles and 20 pct. in top quintile) to (10 pct. in lowest quintile, 30 pct. in middle 3 quintiles and 60 pct. in highest quintile)						
Lowest Quintile	16.170	4.124	5.533	15.485	4.328	4.498
Middle 3 Quintiles	7.397	1.694	1.537	6.038	1.901	1.186
Highest Quintile	-3.387	-0.462	-1.818	-7.305	-0.384	-1.599
Change from (20 pct. in lowest quintile, 60 pct in middle 3 quintiles and 20 pct. in top quintile) to (5 pct. in lowest quintile, 90 pct. in middle 3 quintiles and 5 pct. in highest quintile)						
Lowest Quintile	-9.164	-2.351	-2.898	-8.825	-2.666	-2.944
Middle 3 Quintiles	-0.081	0.136	0.351	0.532	-0.067	-0.016
Highest Quintile	8.684	2.242	4.524	10.703	1.701	3.037

Table 9.
Policy Simulation: Estimated Effect of School Choice Program That Removes 2.5 Percent of Students, All from the Highest Quintile

	Math			Reading		
	Elementary (Grades 4-5)	Middle (Grades 6-8)	High School (Grades 9-10)	Elementary (Grades 4-5)	Middle (Grades 6-8)	High School (Grades9-10)
Lowest Quintile	-0.749	-0.191	-0.252	-0.718	-0.204	-0.215
Middle 3 Quintiles	-0.385	-0.088	-0.079	-0.313	-0.099	-0.062
Highest Quintile	0.188	0.027	0.101	0.394	0.022	0.087

Appendix A
Comparison of Estimation Methods Using Simulated Data

The following table compares the performance of two different iterative estimation methods for estimating peer effects using simulated data. Method 1 is that which we use to estimate peer effects in the Florida data; Method 2 is its mathematically exact cousin. The methods are adapted, respectively, from Arcidiacono et al. (2005) and Arcidiacono et al. (2007). In the data, each student is randomly assigned a permanent ability value from the same normal distribution, a value that represents her fixed contribution to the gain score. Teachers are also assigned permanent ability values randomly from a normal distribution. Students are grouped into classrooms and assigned to a teacher according to rules that vary in the degree of randomness with respect to student and/or teacher ability. The data properties that we allow to vary are the following:

(1) Degree of student selection: the number in this column refers to the ratio of the average within-classroom variance of student ability to the global variance of student ability. A number close to 1 indicates near-random assignment of students to classrooms, while a number close to zero indicates a high degree of selection of students into classrooms by ability level. Simulated values range from a low near 0.25 to high values very close to 1. In the Florida data, based on our estimated student fixed effects, these values range from a low of 0.49, for elementary school math, to a high of 0.78, for middle school reading. The remaining values were 0.59, 0.67, 0.69, and 0.72.

(2) Degree of teacher selection: the number in this column refers to the correlation coefficient between classroom-average student ability and the ability of the teacher assigned to that classroom. A number close to zero indicates that teachers are assigned to classrooms randomly (and this can be done regardless of the degree of student selection), and larger numbers indicate that higher-ability teachers tend to get paired with student groups with high average ability. The more random is student classroom assignment, the harder it is to produce a high degree of teacher selection.

(3) Noise level: the standard deviation of the time-varying idiosyncratic shock applied to the student gain scores.

The properties that are constant across estimations are the following:

(1) The magnitude of peer effects, as indicated by the coefficient on mean peer ability. This number, denoted γ, is set at 0.15 universally.

(2) The number of observations per student, which is set at 2. The two observations of a given student are associated with different grade levels, different years, and different teachers, each of which contributes a fixed effect to the gain score. We construct two cohorts of students, such that students can be in one of two different grade levels in each time period.

(3) The number of observations per teacher, which is set at 2.

(4) Tolerance set at .001. This means that the iterative process stops when the estimated peer effect changes by less than this absolute amount relative to the previous iteration's estimate.

(5) Standard errors obtained by bootstrapping; number of bootstrap replications set at 50.

Table A1.
Comparison of Estimation Methods Using Simulated Data
(True Peer Effect = 0.15)

Student Selection	Teacher Selection	Std. Dev. of Shocks	Method 1 Estimated Peer effect	Method 1 (Standard Error)	Method 2 Estimated Peer Effect	Method 2 (Standard Error)	
0.9948	-0.0262	0.123	0.1521	(0.0095)	0.1496	(0.0084)	
0.9750	0.0362	0.423	0.1828	(0.0262)	0.1940	(0.0331)	
0.9869	-0.0222	1.563	0.0927	(0.0617)	0.1843	(0.1418)	
0.7454	0.0650	0.123	0.1518	(0.0047)	0.1509	(0.0049)	
0.7447	0.5953	0.123	0.1469	(0.0053)	0.1444	(0.0051)	
0.7459	0.0423	0.423	0.1489	(0.0138)	0.1505	(0.0174)	
0.7491	0.6322	0.423	0.1712	(0.0200)	0.1721	(0.0181)	
0.7509	0.0623	1.563	0.1121	(0.0455)	0.1344	(0.0645)	
0.7451	0.5900	1.563	0.1310	(0.0531)	0.1651	(0.0949)	
0.4977	0.0137	0.123	0.1391	(0.0046)	0.1413	(0.0049)	
0.4965	0.6209	0.123	0.1386	(0.0052)	0.1398	(0.0057)	
0.4924	-0.0077	0.423	0.1368	(0.0169)	0.1402	(0.0199)	
0.4939	0.6619	0.423	0.1558	(0.0176)	0.1592	(0.0179)	
0.4970	-0.0309	1.563	0.0548	(0.0455)	0.0670	(0.0595)	
0.4951	0.6373	1.563	0.1121	(0.0562)	0.1456	(0.0921)	
0.2509	-0.0446	0.123	0.1171	(0.0085)	0.1217	(0.0089)	
0.2511	0.6724	0.123	-0.3207	(0.0062)	0.1165	(0.0073)	*
0.2495	0.1071	0.423	-0.0659	(0.0661)	0.0973	(0.0358)	*
0.2487	0.6301	0.423	-0.3087	(0.0081)	0.1197	(0.0326)	*
0.2486	0.0258	1.563	0.0946	(0.0754)	0.1709	(0.1755)	
0.2494	0.6533	1.563	0.1605	(0.0748)	0.3116	(0.1495)	

Notes: tolerance for estimation = 0.001; bootstrapped standard errors with 50 repetitions.
*signifies that the two coefficient estimates are significantly different from each other

CALDER