

Peer Effects on Student Achievement: An Instrumental Variable Approach Using School Transition Data*

Hongliang Zhang[†]

Department of Economics

The Chinese University of Hong Kong

March 2010

Abstract

In this paper, we examine peer effects on student achievement in middle school using a multi-cohort longitudinal data set from China. We use cohort-to-cohort variation in peer composition within schools to identify the impact of schoolmates. Our methods innovate on previous peer effects studies that employ the value-added models with lagged achievement. We show that estimates from the typical value-added specifications are biased if a student's old and new peer groups have members in common, and propose an empirical strategy to use the lagged achievement of new peers to instrument for that of all peers. Our within-school IV estimate of the linear-in-means peer effects coefficient is positive but insignificant. While we cannot reject the null hypothesis of no peer effects, we also cannot reject relatively large peer effects that have been found in the previous literature. Estimates of a heterogeneous peer effects model, however, show evidence in support of the single-crossing property: an increase in average peer ability benefits high-achieving students more than low-achieving students.

*I am grateful to Joshua Angrist, Esther Duflo, and Karen R. Polenske for their advice and support. I thank David Autor, Abhijit Banerjee, Amy Finkelstein, Rongzhu Ke, Tao Jin, Weifeng Li, and seminar participants at MIT and the Chinese University of Hong Kong for valuable comments and suggestions. I gratefully acknowledge financial support from the MIT Shultz Fund and the MIT Center for International Studies Summer Study Grant. All remaining errors are my own.

[†]E-mail: hongliang@cuhk.edu.hk/ Address: Department of Economics, Chinese University of Hong Kong, Shatin, Hong Kong SAR.

1 Introduction

The effects of peer groups on students' academic performance play a prominent role in various education policy debates. Many current education interventions – for example, school choice, ability tracking, and affirmative action – have the potential to influence student outcomes through their impacts on peer composition. Understanding the structure and the magnitude of peer effects is therefore a critical ingredient in evaluating these policies. However, despite the importance of peer influences for education policies, empirical research has not yet reached a consensus on the existence and the nature of peer effects. While some studies show large positive effects of peer quality on academic achievement (e.g., McEwan, 2003; Kang, 2007), others find small or insignificant effects (e.g., Angrist and Lang, 2004; Lefgren, 2004).

The lack of consensus on peer influences reflects various challenges confronted by empirical research on peer effects (Manski, 1993; Moffitt, 2001; Brock and Durlauf, 2001). The first challenge is to isolate peer effects from “correlated effects” due to the correlation between peer composition and the omitted individual or institutional characteristics that can affect student outcomes. The second challenge, known as the “reflection problem,” arises from the reciprocal nature of peer interactions: a student influences her peers and is also influenced by her peers, which causes a classical simultaneity problem of econometrics. These two challenges have engaged much of the attention of the peer effects literature and have been treated intensively. The past decade has seen the development of a variety of empirical strategies to identify exogenous sources of variation in peer characteristics to deal with the endogeneity problem. These recent studies have exploited within-school (grade) variation (e.g., Hoxby, 2000; Hanusheck et al., 2003; McEwan, 2003; Vigdor and Nechyba, 2004, 2006; Lavy and Schlosser, 2007; Ammermueller and Pischke, 2009; Gould, Lavy, and Paserman, forthcoming), within-student variation (e.g., Betts and Zau, 2004; Lavy, Silva, and Weinhardt, 2009), subgroup reassignment (e.g., Angrist and Lang, 2004; Hoxby and Weingarth, 2005), instrumental variables (IV) (e.g., Kang, 2007; Zabel, 2008), and random assignment (e.g., Sacerdote, 2001; Zimmerman, 2003; Duflo, Dupas, and Kremer, 2007, 2008). Identifying the structural parameters under the simultaneity problem, however, has proved difficult or impossible without imposing severe restrictions on the econometric model.¹ The empirical peer

¹Necessary conditions for identification of the structural parameters can be found in Brock and Durlauf (2001). Two principle methods for identification are either to introduce some type of nonlinearity into the model (e.g., suppose

effects literature has often resorted to estimating the exogenous relationship between individual outcomes and predetermined measures of peer composition to circumvent the reflection problem (Nechyba, 2006). While some studies have focused on the relationship between exogenous peer characteristics (such as race, gender, and family background) and individual outcomes (e.g., Hoxby, 2000; Angrist and Lavy, 2004; Ammermueller and Pischke, 2009; Lavy and Schlosser, 2007; Gould, Lavy, and Paserman, forthcoming), other studies have benefited from panel data to include lagged outcome measures rather than contemporary values (e.g., Hanusheck et al., 2003; Ding and Lehrer, 2003; Vigdor and Nechyba, 2004, 2006; Lavy, Silva, and Weinhardt, 2009).²

In this paper, we examine peer effects on students' math scores using a matched panel data set from China. The data set consists of 7,435 students from three successive cohorts of all the 15 middle schools in a school district and tracks their academic histories from finishing primary school (grades 1-6) to completing middle school (grades 7-9). By taking advantage of the panel data set, we address the reflection problem by focusing our interest on the exogenous relationship between predetermined peer characteristics – gender and lagged achievement in particular – and individual outcomes. We base our identification on within-school, between-cohort variation in peer composition, thereby controlling for omitted variables due to unobserved school characteristics and student sorting across schools. In terms of the identification strategy, the papers closest to ours are Hoxby (2000) and Gould, Lavy, and Paserman (forthcoming), both of which use comparisons in adjacent cohorts' peer composition within schools. Our identification strategy is also similar in spirit to studies that assume random classroom assignment within schools and use comparisons across classrooms for the same cohort (grade) in the same school (e.g., McEwan, 2003; Vigdor and Nechayba, 2004, 2006; Kang, 2007; Ammermueller and Pischke, 2009).

A surprising result for our data is that the within-school estimate of the coefficient of the average lagged peer test score is negative and significant. We argue that the unexpected negative sign of the estimated peer coefficient is explained by correlation in measurement errors between the individual- and the peer-level regressors. In our within-school estimation, we simultaneously control for lagged individual test score and the average lagged peer test score. These two variables,

individual behavior varies with other moments of group behavior) or to assume there exists one individual variable whose group-level average has no direct influence on individual outcomes or vice versa.

²Hanushek et al. (2003) provide a thorough discussion of how this does and does not fully address the reflection problem.

however, are subject to transitory common shocks due to the continuing presence of former peers in a student's current peer group. In this paper, we refer to transitory common shocks as group-specific contextual, or environmental, influences that have only transitory effects on students' observed outcomes, i.e., these influences affect the observed test scores of all students in a group, but not their abilities. For example, if a teacher happens to cover in the classroom some materials that for random reasons are tested in the exam, the test scores of all students in the class will be inflated for this particular exam. The presence of such transitory common shocks will lead to a positive correlation in measurement errors between the individual- and the peer-level lagged test score variables. We derive formally that such a positive correlation in measurement errors will lead to a negative bias in the estimated peer coefficient. In our context, this negative bias due to transitory common shocks on lagged test scores dominates the within-school estimator, making the point estimate negative and significant. The source of this bias is the presence of a student's former peers in her current peer group. The longitudinal structure of our data allows us to track the primary school origins of a student's peers and to distinguish between new peers and old peers. In our sample, on average three-quarters of a student's middle school peers are new peers. Our way to address this transitory-common-shock problem is to use the lagged test score measures of new peers to instrument for the corresponding lagged test score measures of all peers. Under the assumption that transitory shocks in lagged test scores are uncorrelated for students from different primary schools, the measurement error in the lagged test scores of new peers is expected to be unrelated to the measurement error in lagged individual test score. Hence, the transitory-common-shock problem can be circumvented by using the lagged test score measures of new peers as instruments. The existence of transitory common shocks has been well documented in the school accountability literature, in which it leads a regression-to-the-mean problem in school or teacher assessment (e.g., Kane and Staiger, 2002; Betts and Dannenberg, 2002). The potential effect of transitory common shocks on lagged individual and peer test scores, however, has largely been ignored in the peer effects literature. This paper clarifies the econometric problem of transitory common shocks on lagged test scores and makes an important methodological contribution to the existing peer effects literature by proposing an IV strategy to correct this problem.

In our linear-in-means peer effects models, we examine the effect of peer gender mix and average lagged peer test score on a student's 9th-grade math score in a school fixed-effect framework. As

discussed earlier, we instrument the average lagged peer test score with the average lagged test score of new peers to circumvent the transitory-common-shock problem. Unlike some previous studies that find positive spillover effects of girls on math scores (Hoxby, 2000; Whitmore, 2003; Lavy and Schlosser, 2009), we find no evidence that peer gender composition has an impact on students' 9th-grade math scores. Our within-school IV estimate of the linear-in-means peer effects coefficient is positive but insignificant. While we cannot reject the null hypothesis of no peer effects, we also cannot reject relatively large peer effects that have been found in the previous literature. Estimates of a heterogeneous peer effects model, however, show evidence in support of the single-crossing property: an increase in average peer ability benefits high-achieving students more than low-achieving students.

The remainder of this paper proceeds as follows: Section 2 provides background and describes the data; Section 3 discusses the econometric problem and presents the empirical strategy; Section 4 presents the empirical results on peer effects on student test scores; and Section 5 provides some concluding remarks.

2 Empirical Framework

2.1 A Simple Model of Achievement

As in many peer effects studies, we start from a simple value-added education production function in which a student's own achievement depends on her own predetermined (academic) ability and the average predetermined (academic) ability of her peers:

$$Y_{ics} = \beta A_{ics} + \lambda \bar{A}_{(-i)cs} + \phi_s + \kappa_{cs} + v_{ics} \quad (1)$$

where Y_{ics} is the current achievement for student i of cohort c in school s ; A_{ics} is the ability of student i predetermined at the entry of school s , which captures the cumulative effects of prior family, neighborhood, and school inputs as well as generic ability; $\bar{A}_{(-i)cs}$ is the average predetermined ability of student i 's peers; ϕ_s represents the school-specific common-shock effects, arising from school-level unobserved common contextual, or environmental, influences that affect the outcomes of all students in that school; κ_{cs} represents variation in the common-shock effects across cohorts

within schools and has a zero-mean for each school; and v_{ics} is an individual-level stochastic error term that has a zero-mean within each cohort in each school. For simplicity, the model is set up by assuming no cohort-to-cohort evolution in student ability A_{ics} or achievement Y_{ics} . In practice, such cohort-to-cohort evolution can be easily controlled by including a cohort fixed effect.

It is worth noting that λ in equation (1) is a reduced-form parameter, which captures the combination of exogenous effects and endogenous effects in the language of Manski (1993). Exogenous effects arise when peer ability is a direct input to the achievement production process. Endogenous effects arise when peer ability affects student achievement indirectly through its correlation with contemporary peer behavior, the effects of which are not modeled directly here due to the "reflection problem" mentioned earlier. In this paper, we make no attempt to distinguish between these two types of peer influences, but instead focus on separating the reduced-form effects of peer ability from other confounding nonsocial influences. To focus attention to this identification question, we shall assume for a moment that A_{ics} and $\bar{A}_{(-i)cs}$ are both observable and defer the discussion of their measurement problem to the next subsection.

The two common-shock effects ϕ_s and κ_{cs} reflect correlated effects, which arise when individuals in the same group are subject to common influences that are not modeled directly. These correlated effects can bias the coefficient λ if they are correlated with peer ability $\bar{A}_{(-i)cs}$. Random assignment of students to groups, where a group refers to a cohort in a school in our context, can solve this problem because randomization breaks the potential link between peer ability ($\bar{A}_{(-i)cs}$) and the common shock effects (ϕ_s and κ_{cs}). However, true random assignment rarely exists outside experimental settings (Sacerdote, 2001; Zimmerman, 2003; Duflo, Dupas, and Kremer, 2007, 2008). In practice, parents choose a school based on its quality and the composition of its peers, and schools also have some discretion in choosing students for admission. Hence, peer ability $\bar{A}_{(-i)cs}$ will be systematically correlated with common shock effects ϕ_s at the school level, causing the OLS estimator of λ to be biased. A possible way to account for such school-level correlated effects is to use a within-school regression that exploits variation in peer composition across adjacent cohorts within the same school. As shown in Appendix A, the within-school transformation of the achievement function takes the following form:

$$y_{ics} = \beta a_{ics} + \lambda \bar{a}_{(-i)cs} + \kappa_{cs} + v_{ics}$$

where y_{ics} , a_{ics} , and $\bar{a}_{(-i)cs}$ are the deviations of Y_{ics} , A_{ics} and $\bar{A}_{(-i)cs}$ from their school means, respectively.

The basic idea behind the within-school regression is to examine whether, for students from adjacent cohorts in the same school, those who have more favorable peers in their cohort score higher conditional on their own abilities. The identification assumption of the within-school model is that the within-school, between-cohort variation in peer ability, $\bar{a}_{(-i)cs}$, is uncorrelated with the within-school, between-cohort variation in common shock effects, κ_{cs} . This identification strategy is the same as Hoxby (2000), Gibbons and Telhaj (2008), and Gould, Lavy, and Paserman (forthcoming), all of which use comparisons in adjacent cohorts' peer composition within schools. It is also similar in spirit to studies that assume random classroom assignment within schools and use comparisons across classrooms for the same cohort (grade) in the same school (e.g., McEwan, 2003; Vigdor and Nechayba, 2004, 2006; Kang, 2007; Ammermueller and Pischke, 2009; Carman and Zhang, 2009). Under this identification assumption, the common-shock component, κ_{cs} , and the individual stochastic component, v_{ics} , can be subsumed into a general error term ϵ_{ics} such that $\epsilon_{ics} = \kappa_{cs} + v_{ics}$. Consequently, the within-school regression estimates the following equation:

$$y_{ics} = \beta a_{ics} + \lambda \bar{a}_{(-i)cs} + \epsilon_{ics} \quad (2)$$

The estimation of the reduced-form λ in equation (2) is not confounded by correlated effects because $cov(\bar{a}_{(-i)cs}, \epsilon_{ics}) = 0$ under the identification assumption. However, equation (2) still cannot be estimated directly because the de-meaned ability variables a_{ics} and $\bar{a}_{(-i)cs}$ are not observed directly. In the empirical peer effects literature, lagged achievement is often as a proxy for the unobserved ability. We also follow this approach and present in the next subsection a statistical model in which latent ability variables are proxied by measures of lagged achievement.

2.2 The Statistical Model

Let x_{ics} denote the deviation of student i 's lagged achievement from its school mean, and w_{ics} denote the deviation of the average lagged achievement of student i 's peers from its school mean.

Then, x_{ics} and w_{ics} can be related to a_i and $\bar{a}_{(-i)g}$, respectively, as follows:

$$x_{ics} = a_{ics} + v_{ics} \quad (3a)$$

$$w_{ics} = \bar{a}_{(-i)cs} + u_{ics} \quad (3b)$$

where v_{ics} is a stochastic error term that has a zero mean within each school and is uncorrelated with ϵ_{ics} in equation (2), and $u_{ics} = \bar{v}_{(-i)cs}$ (see Appendix A for further details on the within-school transformation). Substituting equations (3a) and (3b) into equation (2) yields

$$y_{ics} = \beta x_{ics} + \lambda w_{ics} + \psi_{ics} \quad (4)$$

where $\psi_{ics} = \epsilon_{ics} - \beta v_{ics} - \lambda u_{ics}$. Note that x_{ics} (w_{ics}) and ψ_{ics} are correlated because they both v_{ics} (u_{ics}).

Assumptions (A1)-(A4):

(A1) $cov(a_{ics}, \kappa_{cs}) = 0$;

(A2) $cov(\bar{a}_{(-i)cs}, \kappa_{cs}) = 0$;

(A3) $cov(a_{ics}, \bar{a}_{(-i)cs}) = 0$;

(A4) $cov(v_{ics}, u_{ics}) = 0$.

Assumptions (A1) and (A2) indicate that the cohort-to-cohort innovation in the common-shock component (for example, teacher quality), κ_{cs} , is random and uncorrelated with own ability and peer ability. Assumption (A3) implies no student sorting across cohorts within the same school. Although students may choose schools based on teacher quality and peer composition, these assumptions hold as long as students do not choose schools on the basis of expected innovation in teacher quality and peer composition in a given year. It is this line of argument that we will follow in our within-school strategy. Assumption (A4) requires that measurement errors in the lagged individual and peer achievement are uncorrelated, i.e. $cov(v_{ics}, u_{ics}) = 0$, a condition that would hold if the error terms of the lagged individual achievement are i.i.d. within each cohort in each school. Assumptions (A1) to (A4) are the same set of assumptions made by Ammermueller and Pischke (2009) in their estimation of peer effects in European primary schools. While we rely on variation across cohorts within the same school for identification, they use comparisons across classes within

the same grade in the same school. In their paper, they do not observe students' lagged achievement but use parents' reports of number of books at home as a peer quality measure. They argue that classes are formed roughly randomly in European primary schools and that measurement errors in number of books at homes are uncorrelated within classes. Thus, assumptions (A1) to (A4) hold in their context.

Proposition 1 *Under assumptions (A1) to (A4), the within-school estimates of β and λ in equation (4) converge as follows:*

$$\begin{aligned} p \lim \widehat{\beta}_W &= \frac{\sigma_a^2}{\sigma_a^2 + \sigma_v^2} \beta \\ p \lim \widehat{\lambda}_W &= \frac{\sigma_a^2}{\sigma_a^2 + \sigma_u^2} \lambda \end{aligned}$$

where σ_a^2 , σ_v^2 , σ_a^2 , and σ_u^2 denote the variances of a_{ics} , v_{ics} , $\bar{a}_{(-i)cs}$, and u_{ics} , respectively.

The proof of Proposition 1 is provided in Ammermueller and Pischke (2009) and can also be found in our Appendix B as a special case of Proposition 2. The plims of $\widehat{\beta}_W$ and $\widehat{\lambda}_W$ are both subject to attenuation biases in the classical errors-in-variables (EIV) problem because x_{ics} and w_{ics} are both correlated with the error term ψ_{ig} in equation (4). Specifically, the plims of $\widehat{\beta}_W$ and $\widehat{\lambda}_W$ are regression coefficients in the following model:

$$y_{ics} = \beta^* x_{ics} + \lambda^* w_{ics} + \mu_{ics} \tag{5}$$

where $\beta^* = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_v^2} \beta$, $\lambda^* = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_u^2} \lambda$, and $\mu_{ics} = (\beta - \beta^*)x_{ics} + (\lambda - \lambda^*)w_{ics} + (\epsilon_{ics} - \beta v_{ics} - \lambda u_{ics})$. It can be shown that the error term μ_{ics} in equation (5) is orthogonal to both x_{ics} and w_{ics} . Thus, although the structural coefficients β and λ in the value-added achievement model cannot be estimated directly due to the unobservable nature of individual and peer abilities, the statistical coefficients β^* and λ^* in equation (5) can be estimated consistently.

Ammermueller and Pischke argue that, if there exists another set of independent measures of the same individual and peer variables x'_{ics} and w'_{ics} (e.g., students' reports of number of books at home in addition to parents' reports), using x'_{ics} and w'_{ics} as instruments for x_{ics} and w_{ics} can

correct the measurement error problem and provide consistent estimates of β and λ . However, we do not always have two measurements of the same variables, and, even if we do, the errors in the two measurements may well be correlated. For example, if parents overstate the number of books at home, so may their children. Hence, their IV approach to correct the measurement error problem may not always be feasible. Despite the classical attenuation biases, the statistical coefficient λ^* in equation (5) is still informative for two reasons. First, the attenuation bias with respect to the structural parameter λ decreases with the precision of the proxy ability measure. We expect that the lagged achievement used in our study is a more precise measure of ability compared to other indirect measures used in previous studies, such as mother's schooling and number of books at home, thus alleviating the magnitude of the attenuation bias. Second and more importantly, λ^* represents an important policy parameter – the marginal effect of average lagged peer achievement on a student's current achievement. From the policy point of view, the statistical parameter λ^* is even more informative than the structural parameter λ because of the unobservable nature of ability. For the rest of the paper, we shall refer to the empirical specification in equation (5) as the statistical model and focus attention to the consistent estimation of the statistical peer effects coefficient λ^* .

2.3 The Problem of Transitory Common Shocks in Lagged Achievement

The peer effects literature usually uses "common shocks" to refer to group-level common contextual factors that influence the achievement of all students in a group. Some of these common contextual factors, such as school resources and teacher quality, can influence student achievement through effects on student ability, while others may influence student achievement through effects on measurement errors only. In this paper, we use the term "transitory common shocks" to refer to group-level common contextual, or environmental, influences that have only transitory effects on student achievement, i.e., these influences only affect the measurement errors in student achievement but not student ability. As we have mentioned earlier, random overlapping between exam contents and classroom instructions is one source of such transitory common shocks. Another source is the across-school difference in grading standards, causing students' grades to be inflated in some schools but deflated in others. The second type of transitory common shocks can be eliminated if

standardized exams are used or grading is assigned randomly at the student level.³ However, there is no direct way to correct the transitory common shocks arising from the random overlapping of exam contents and classroom instructions.

Assumption (A4) requires the measurement errors in the lagged individual (x_{ics}) and peer achievement (w_{ics}) to be uncorrelated. But in reality, students usually take some former peers with them when moving to the next schooling phase. For example, for students in our sample, about a quarter of their middle school peer group is made up of their former peers from the same primary school. To the extent that the lagged achievement of students from the same former group is subject to transitory common shocks, the presence of a student's former peers in her current peer group leads to a positive correlation between the measurement errors in the lagged individual and peer achievement variables, i.e. $\rho = cov(v_{ics}, u_{ics}) > 0$. As w_{ics} contains u_{ics} and the error term ψ_{ics} ($= \epsilon_{ics} - \beta v_{ics} - \lambda u_{ics}$) in equation (4) contains both v_{ics} and u_{ics} , the correlation between v_{ics} and u_{ics} will carry over to w_{ics} and ψ_{ics} . Specifically, $cov(w_{ics}, \psi_{ics})$ equals $-\lambda\sigma_u^2$ when v_{ics} and u_{ics} are independent and equals $-\beta\rho - \lambda\sigma_u^2$ when v_{ics} and u_{ics} are correlated. Thus, the standard attenuation bias formula in Proposition 1 no longer applies when the measurement errors in the individual and peer achievement variables are linked through common group membership.

Proposition 2 *Under assumptions (A1) to (A3), the within-school estimate of β and λ in equation (4) converge, respectively, as follows:*

$$\begin{aligned}
p \lim \hat{\beta}_W &= \beta - \underbrace{\frac{\sigma_v^2 - \frac{\rho^2}{(\sigma_a^2 + \sigma_u^2)}}{(\sigma_a^2 + \sigma_v^2) - \frac{\rho^2}{(\sigma_a^2 + \sigma_u^2)}}}_{\text{"attenuation bias"}} \beta - \underbrace{\frac{\sigma_a^2}{(\sigma_a^2 + \sigma_u^2)} \frac{\rho}{(\sigma_a^2 + \sigma_v^2) - \frac{\rho^2}{(\sigma_a^2 + \sigma_u^2)}}}_{\text{"transitory-common-shock bias"}} \lambda \\
p \lim \hat{\lambda}_W &= \lambda - \underbrace{\frac{\sigma_u^2 - \frac{\rho^2}{\sigma_a^2 + \sigma_v^2}}{\sigma_a^2 + \sigma_u^2 - \frac{\rho^2}{\sigma_a^2 + \sigma_v^2}}}_{\text{"attenuation bias"}} \lambda - \underbrace{\frac{\sigma_a^2}{\sigma_a^2 + \sigma_v^2} \frac{\rho}{\sigma_a^2 + \sigma_u^2 - \frac{\rho^2}{\sigma_a^2 + \sigma_v^2}}}_{\text{"transitory-common-shock bias"}} \beta
\end{aligned} \tag{7}$$

³Note that random assignment of grading at the school or class level cannot eliminate the transitory common shocks arising from differences in grading standards because grading errors would still be clustered at the school or class level in this case.

where σ_a^2 , σ_v^2 , $\sigma_{\bar{a}(-i)cs}^2$, and σ_u^2 denote the variances of a_{ics} , v_{ics} , $\bar{a}_{(-i)cs}$, and u_{ics} , respectively, and ρ denotes $cov(v_{ics}, u_{ics})$.

The proof of Proposition 2 is provided in Appendix B. Let us focus our discussion on the within-school estimator of peer coefficient $\hat{\lambda}_W$. As $\sigma_u^2 \sigma_v^2 \geq [cov(u, v)]^2 = \rho^2$, $\sigma_u^2 \geq \frac{\rho^2}{\sigma_v^2} > \frac{\rho^2}{\sigma_a^2 + \sigma_v^2}$. Thus, assuming that β , λ and ρ are all positive, $\hat{\lambda}_W$ underestimates λ as both bias components are negative. The first bias component, which we refer to as the "attenuation bias," arises because the lagged peer achievement variable w_{ics} ($= a_{ics} + u_{ics}$) and the error term ψ_{ics} ($= \epsilon_{ics} - \beta v_{ics} - \lambda u_{ics}$) both contain u_{ics} . This attenuation bias component is similar to the classical attenuation bias in Proposition 1 except that it has an additional adjustment component ($-\frac{\rho^2}{\sigma_a^2 + \sigma_v^2}$) in both the numerator and denominator to correct for the correlation between v_{ics} and u_{ics} .⁴ The second bias component, which we refer to as the "transitory-common-shock bias," arises because w_{ics} is correlated with ψ_{ics} also through the correlation between u_{ics} and v_{ics} in the presence of transitory common shocks.

Substituting λ and β in equation (6) with $\frac{\sigma_a^2 + \sigma_u^2}{\sigma_a^2} \lambda^*$ and $\frac{\sigma_a^2 + \sigma_v^2}{\sigma_a^2} \beta^*$, respectively, the within-school estimator $\hat{\lambda}_W$ can be linked to the statistical coefficients β^* and λ^* in equation (5) as the following:

$$p \lim \hat{\lambda}_W = \lambda^* + \frac{\frac{\rho^2}{\sigma_a^2 + \sigma_v^2}}{\sigma_a^2 + \sigma_u^2 - \frac{\rho^2}{\sigma_a^2 + \sigma_v^2}} \lambda^* - \frac{\rho}{\sigma_a^2 + \sigma_u^2 - \frac{\rho^2}{\sigma_a^2 + \sigma_v^2}} \beta^* \quad (8)$$

Hence, $\hat{\lambda}_W$ is also a biased estimator of the statistical coefficient λ^* . It is important to note that, when β^* is sufficiently large relative to λ^* , the negative bias component ($-\frac{\rho}{\sigma_a^2 + \sigma_u^2 - \frac{\rho^2}{\sigma_a^2 + \sigma_v^2}} \beta^*$) could dominate both the positive bias component ($\frac{\frac{\rho^2}{\sigma_a^2 + \sigma_v^2}}{\sigma_a^2 + \sigma_u^2 - \frac{\rho^2}{\sigma_a^2 + \sigma_v^2}} \lambda^*$) and the statistical coefficient (λ^*), thus reversing the sign of $\hat{\lambda}_W$.

2.4 The Instrumental Variable Approach

As discussed in the previous subsection, the assumption of uncorrelated measurement errors in the lagged individual and peer achievement variables is likely to be violated in practice since a student's current peer group usually has members in common with her former peer group, whose

⁴It can be shown that this adjusted attenuation bias component is smaller in magnitude than the classical attenuation bias ($-\frac{\sigma_u^2}{\sigma_a^2 + \sigma_u^2} \lambda$).

lagged achievement is subject to the same group-specific transitory common shocks as her own lagged achievement. With correlated measurement errors in lagged achievement, the within-school estimator $\widehat{\lambda}_W$ is biased for both the structural coefficient λ and the statistical coefficient λ^* , and may even have the opposite sign. Since the problem is caused by the continuing presence of a student's former schoolmates in her current peer group, an idea for overcoming this problem is to use the average lagged achievement of a student's new peers ($w_{ics,new} = \bar{x}_{(-i)cs,new}$) to instrument for the average lagged achievement of all peers (w_{ics}). As transitory common shocks in lagged achievement are expected to be uncorrelated for students originating from different former groups, the error term $u_{ics,new}$ in the lagged achievement of new peers $w_{ics,new}$ is expected to be uncorrelated with the error term v_{ics} in the lagged individual achievement x_{ics} . In order to use $w_{ics,new}$ as instrument for w_{ics} , we replace the set of assumptions (A1) to (A4) with the following:

Assumptions (B1)-(B4):

(B1) $cov(a_{ics}, \kappa_{cs}) = 0$;

(B2) $cov(\bar{a}_{(-i)cs,new}, \kappa_{cs}) = 0$;

(B3) $cov(a_{ics}, \bar{a}_{(-i)cs,new}) = 0$;

(B4) $cov(u_{ics}, u_{(-i)cs,new}) = 0$.

Assumption (B1) is the same as (A1). Assumption (B2) indicates that the ability of one's new peers is uncorrelated with the cohort-to-cohort innovation in the common-shock component, κ_{cs} . Assumption (B3) implies that students do not choose schools based on the expected innovation new peer composition in a given year. Assumption (B4) only requires the measurement error in the lagged individual achievement to be uncorrelated with the measurement error in the lagged achievement of new peers. Compared to assumption (A4), (B4) is a much more relaxed assumption and holds if transitory common shocks in lagged achievement are uncorrelated across originating schools.

Proposition 3 *Under assumptions (B1)-(B4), the within-school, IV estimate of λ in equation (4) using $w_{ics,new}$ as an instrument for w_{ics} converges as follows:*

$$p \lim \widehat{\lambda}_{W,IV} = \frac{cov(\bar{a}_{(-i)cs}, \bar{a}_{(-i)cs,new})}{cov(\bar{a}_{(-i)cs}, \bar{a}_{(-i)cs,new}) + cov(u_{ics}, u_{ics,new})} \lambda$$

The proof of Proposition 3 is provided in Appendix C.

Lemma 1 Under assumptions (B1)-(B4) and the following assumption (B5) $\frac{\text{cov}(\bar{a}_{(-i)cs}, \bar{a}_{(-i)cs,new})}{\text{cov}(u_{ics}, u_{ics,new})} = \frac{\sigma_{\bar{a}}^2}{\sigma_u^2}$, the within-school, IV estimate of λ in equation (4) using $w_{ics,new}$ as an instrument for w_{ics} converges the statistical coefficient λ^* :

$$p \lim \hat{\lambda}_{W,IV} = \lambda^*$$

where $\lambda^* = \frac{\sigma_{\bar{a}}^2}{\sigma_{\bar{a}}^2 + \sigma_u^2} \lambda$.

Note that the additional assumption (B5) is likely to hold in practice. Two sufficient conditions for assumption (B5) are: (1) the cohort-to-cohort innovation in the ability (measurement errors) of old peers is uncorrelated with the cohort-to-cohort innovation in the ability (measurement error) of new peers, i.e., $\frac{\text{cov}(\bar{a}_{(-i)cs}, \bar{a}_{(-i)cs,new})}{\sigma_{\bar{a}_{new}}^2} = \frac{\text{cov}(u_{ics}, u_{ics,new})}{\sigma_{u_{new}}^2} = \pi_{new}$, where π_{new} denotes the average proportion of new peers in a student's current peer group; (2) the information-to-noise ratio in the lagged achievement of new peers ($\frac{\sigma_{\bar{a}_{new}}^2}{\sigma_{u_{new}}^2}$) is the same as that in the lagged achievement of all peers ($\frac{\sigma_{\bar{a}}^2}{\sigma_u^2}$). A nice feature of this IV estimation is that $\hat{\lambda}_{W,IV}$ is a consistent estimator for the peer coefficient λ^* in the statistical model. Note that $\hat{\beta}_{W,IV}$ is still a biased estimator for β^* since we do not have an instrument for the lagged individual achievement x_{ics} . But our main goal of consistently estimating the policy parameter λ^* – the marginal effect of the average lagged peer achievement on a student's current achievement – is achieved in this IV estimation.

3 Data and Descriptive Statistics

3.1 Data

We use a data set of middle school students in China to illustrate our empirical strategy. The data come from a school district in the capital city of a central China province. Based on the district's administrative records, we construct a matched panel data set that tracks three successive cohorts of middle school students who had completed middle school in this district between 2005 and 2007. The school district has 35 elementary schools and 15 middle schools. Upon graduation from elementary school, each student is assigned to a middle schools based on residency. The middle school zoning scheme, however, is not fixed over time. Every year, the school district announces a

new zoning scheme for this year’s cohort of elementary school graduates. As proximity has been taken into consideration in drawing up zoning schemes, students know in advance the set of nearby middle schools they might be assigned to, but not the exact school until the school district makes its announcement. Since elementary school enrollment is also based on residency, a student is usually assigned to a middle school together with some former schoolmates from elementary school. Middle schools in this district exhibit a high degree of heterogeneity with respect to the quality of student intakes and teachers, and some are much more preferred than others. Thus, the assignment is not completely binding: although the majority of students attend the middle school assigned by the school district, a nontrivial proportion of students opt out of their assigned school to attend a different one. Our empirical strategy, however, does not rely on the exogenous middle school assignment.

Our panel data set is constructed by matching administrative student records from two sources. Student information at the end of middle school comes from the city’s Middle School Graduation Exam (MSGGE) database, which includes each student’s middle school of graduation and test scores in four subjects examined in the citywide uniform MSGGE: math, science, Chinese, and English. Student information before the start of middle school comes from the district’s records of students’ elementary school of graduation and their math scores in a district-wide uniform exam taken at the end of 6th grade.⁵

Some limitations remain in the structure of the matched panel data set. First, the two databases do not share perfect individual identification information to guarantee unique tracking of the academic histories of all students. Specifically, we can only use the combination of name and gender to match student records in the two databases. Consequently, some students cannot be uniquely tracked due to multiple matches to common names. In addition, some students in the MSGGE database have no matched elementary school records, either because they attended an elementary school outside the district or because their names were misspelled in one or both databases.⁶ Second, we can only identify peer composition at the cohort (grade) level but not the classroom level. The classroom-level measures of peer composition may be more desirable if peer externalities take

⁵Students also take an exam in Chinese at the end of 6th grade. The Chinese exam includes a writing section. Students’ Chinese test scores are largely non-comparable across schools as grading standards differ substantially across schools.

⁶Misspelling is more likely to occur in the elementary school information records, which we obtained in handwritten paper documents and coded them into an electronic database. The MSGGE records are obtained in electronic format.

place mainly through classroom interactions. However, classroom-level measures are likely to be endogenous as school administrators and parents can have some discretion in placing students in different classes within a grade. Because of the potential sorting of students across classrooms within the same school-grade, we shall still use the cohort-level measures even if classroom-level measures were available.

Our sample consists of 7,435 students from three successive cohorts in the district whose academic histories are uniquely tracked. Students in our sample account for about 86 percent of the universe of 8,620 students who had completed middle school in the district during 2005-2007. Our data are ideal for analyzing peer effects in education for two reasons. First, it includes students' 6th-grade test scores predetermined before the peer group formation in middle school. Thus, we can proxy individual and peer abilities by lagged test scores, which may be arguably the best proxies for academic abilities that one can possibly get. Second, our data set includes students' elementary school enrollment information, which allows us to distinguish between new peers and old peers in middle school. As we have discussed in details in Section 2.3, separating new peers from old peers is crucial for the identification of peer effects when lagged test scores are used as proxies for individual and peer abilities.

3.2 Descriptive Statistics

Table 1 presents descriptive statistics for the matched sample. Panel A shows statistics for three individual-level variables: gender, 6th-grade math score, and 9th-grade math score. Panel B reports four exogenous measures of peer composition in middle school: proportion of female peers, proportion of new peers, average 6th-grade peer math score, and average 6th-grade math score of new peers. Column 1 shows the means for these variables. The sample is balanced in gender, consisting of roughly 50 percent girls. On average, three-quarters of a student's peers in middle school are new peers from other primary schools. For ease of interpretation, we normalize student test scores by cohorts to have zero means and standard deviations of one. As some students are not included in our matched sample for reasons discussed above, any observed deviation of our sample mean test score from zero reflects selection into the matched sample. For instance, the average 6th-grade test score (0.035σ) in our sample is slightly higher than the district average (which is normalized to be zero). Our explanation of this difference is that a disproportionate share of students from high

mobility families (e.g., rural migrants), who on average have lower academic achievement, opt out of the district’s middle school system and are therefore not tracked in our matched sample.

Column 2 reports the standard deviations of these individual- and peer-level variables. As any between-school variation is removed in the school fixed-effect framework, our source for identification is variation across cohorts within the same school. Hence, column 3 reports a measure of the within-school dispersion of the individual- and the peer-level variables: the standard deviation of the residual of each variable after removing school and cohort fixed effects. There remains a fair amount of cohort-to-cohort variation in peer composition within schools. For example, the within-school standard deviation of the average 6th-grade peer math score is 0.14σ and two-thirds of the schools have experienced a more than 0.3σ change in the average 6th-grade math score over the three consecutive cohorts observed in the sample.

4 Empirical Results

4.1 OLS and Within-School Estimation

Table 2 reports the OLS and the within-school estimations that regress students’ 9th-grade math scores on their 6th-grade math scores and the average 6th-grade math scores of their peers. Each column corresponds to a separate regression and the standard errors reported in parentheses are adjusted for clustering within each cohort in each school. Column 1 reports the least square estimation that does not control for school fixed effects. The OLS estimator $\hat{\lambda}_{OLS}$ (0.591 with a standard error of 0.104) shows a very strong positive relationship between one’s 9th-grade test score and the average 6th-grade test score of one’s peers in a cross-sectional setting. The large positive OLS estimator of peer coefficient, however, almost certainly confounds peer effects with “correlated effects” because of student sorting across schools based on the unobserved school characteristics. The fact that the estimated peer coefficient $\hat{\lambda}_{OLS}$ (0.591) is even larger than the estimated coefficient of own lagged test scores $\hat{\beta}_{OLS}$ (0.442) also implies that $\hat{\lambda}_{OLS}$ is likely to be biased upward due to correlated effects and that the magnitude of the bias may be quite large. As we have discussed earlier, introducing school fixed effects to the model can mitigate the bias due to school-level correlated effects. Column 2 reports the results of the within-school estimation that includes both school and cohort fixed effects. The F-test of the joint significance of the school fixed effects

has a p-value below 0.001, showing evidence of the existence of school-level correlated effects. Not surprisingly, $\widehat{\lambda}_W$ is reduced considerably in the within-school estimation. What is perhaps surprising is that $\widehat{\lambda}_W$ is now negative and significant (with a point estimate of -0.250 and a standard error of 0.112). Although the empirical literature has not reached a consensus on the existence and the magnitude of peer effects, the true peer coefficient λ is unlikely to be negative. Hence, we take the negative and significant point estimate of λ as evidence that our within-school estimator $\widehat{\lambda}_W$ is subject to a negative bias that cannot be simply explained by the attenuation bias.

Next, we revisit the assumptions ((A2) to (A4)) used to derive the within-school peer effects estimator $\widehat{\lambda}_W$ in Proposition 1 to examine the potential sources of such a negative bias. First, we assume that within-school, between-cohort variation in peer quality $\bar{a}_{(-i)cs}$ is unrelated to the de-meaned individual ability a_{ics} . This assumption implies that student sorting only occurs across schools but not across cohorts within the same school. This is plausible in our context as parents are unlikely to be well informed and sophisticated enough to condition their school choice decision on the cohort-to-cohort variation in peer quality within a school. Moreover, with classical measurement errors, within-school student sorting will introduce an upward bias in $\widehat{\lambda}_W$, opposite to what we have seen in the data. Specifically, if there exists within-school student sorting by ability, i.e., $\pi = cov(a_{ics}, \bar{a}_{(-i)cs}) > 0$, the within-school estimator $\widehat{\lambda}_W$ will converge to $\lambda - \frac{\sigma_u^2}{(\sigma_a^2 + \sigma_u^2) - \frac{\pi^2}{(\sigma_a^2 + \sigma_v^2)}} \lambda + \frac{\sigma_v^2}{(\sigma_a^2 + \sigma_v^2)} \frac{\pi}{(\sigma_a^2 + \sigma_u^2) - \frac{\pi^2}{(\sigma_a^2 + \sigma_v^2)}} \beta$. The third component of this expression can be interpreted as a correlation bias, which arises from correlation between a_{ics} and $\bar{a}_{(-i)cs}$, and has the same sign as own ability effect β . Our second assumption is that, within the same school, the cohort-to-cohort variation in peer quality $\bar{a}_{(-i)cs}$ is uncorrelated with the cohort-to-cohort variation in common-shock effects κ_{cs} . A downward bias in $\widehat{\lambda}_W$ would arise if $\bar{a}_{(-i)cs}$ is instead negatively correlated with κ_{cs} . This would be the case if, when a cohort quality is relatively poor in a school, a principal who cares about within-school equity assigns high-quality teachers to that cohort to partly compensate for the poor student quality. However, the extent of such endogenous teacher assignment, if it exists at all, is likely to be quite limited as teachers usually rotate their grade assignment on a three-year basis (grades 7 to 9). Moreover, we would not expect a principal to manipulate teacher assignment to the extreme extent to more than fully compensate the difference in peer quality such that the net effect $(\lambda \bar{a}_{(-i)cs} + \kappa_{cs})$ is negatively correlated with peer quality $\bar{a}_{(-i)cs}$.

The third assumption to derive $\hat{\lambda}_W$ in Proposition 1 is that measurement errors in the individual- and the peer-level lagged test scores are uncorrelated, i.e., $\rho = \text{cov}(v_{ics}, u_{ics}) = 0$, a condition that would hold if the error terms of lagged individual test scores are i.i.d. within each cohort in each school. However, students usually take some former peers with them when moving to the next schooling phase. In our sample, about a quarter of a student’s peers in middle school are her former peers from the same primary school. To the extent that the lagged test scores of students from the same primary school are subject to transitory common shocks, the presence of a student’s former peers in her current peer group leads to a positive correlation between v_{ics} and u_{ics} , i.e., $\rho > 0$. As shown in Proposition 2, the negative transitory-common-shock bias could reverse the sign of $\hat{\lambda}_W$ when β is large relative to λ . Thus, we take the negative and significant within-school estimator $\hat{\lambda}_W$ as evidence for the existence of transitory-common-shock bias. We shall in the following use our IV strategy to correct for the transitory-common-shock bias.

4.2 IV Estimation

Results in Table 2 show a possibly serious negative transitory-common-shock bias in the within-school estimator of peer coefficient $\hat{\lambda}_W$. Our approach to correct this transitory-common-shock problem is to use the lagged test score measures of new peers to instrument for the corresponding measures of all peers. Columns 1 and 2 in Table 3 present our reduced-form and IV estimates from the homogeneous peer effects model. The coefficient of the female dummy is statistically insignificant in both specifications, indicating no significant gender gap in 9th-grade math scores. The coefficient of lagged individual test score is highly significant and is estimated to be virtually the same (around 0.43). We now turn to the results on peer effects. We are interested in two measures of peer composition: peer gender mix and average lagged peer test score (a proxy measure for the average peer ability). Unlike some previous studies that find positive spillover effects of girls on math scores (Hoxby, 2000; Whitmore, 2003; Lavy and Schlosser, 2009), we find no evidence that peer gender composition has an impact on students’ 9th-grade math scores. The coefficient of the proportion of female peers variable is insignificant in both the reduced-form and the IV specifications. Once we replace average lagged achievement of a student’s peers with the same measure of her new peers, the negative peer coefficient for the within-school estimator (-0.250 with a standard error of 0.112) disappears, evidence supporting our transitory-common-shock bias

explanation. Unfortunately, both the reduced-form (0.121 with a standard error of 0.087) and the IV coefficients (0.218 with a standard error of 0.175) are very imprecisely estimated. Although we cannot reject the null hypothesis of no linear-in-means peer effects, we also cannot reject very large peer effects. The imprecise reduced-form and IV estimators are likely to be because of the relatively small number of clusters (cohorts \times schools) in our sample. Despite the imprecise results, the pattern of change from the within-school estimator to the reduced-form and IV estimators still shows strong evidence for the existence of a severe negative bias of using the average lagged peer test score measure when this measure is subject to transitory common shocks just like the student's own lagged test score.

Peer influences, however, may be heterogeneous and operate through the interaction between peer ability and a student's own ability. For instance, some existing computational models of peer sorting in schools assumes that peer effects exhibit single crossing, i.e., an increase in average peer ability affects high-achieving students more than low-achieving students (Nechyba, 2006). In addition, several recent empirical studies find that students seem to benefit from having peers with similar characteristics as themselves, evidence in support of tracking (e.g., Hoxby and Weingarth, 2005; Duflo, Dupas, and Kremer, 2008).

We next explore a heterogeneous peer effects model by interacting a student's lagged test score with the average lagged test score of her peers. To implement the estimation of this heterogeneous peer effects model in an IV framework, we instrument the interaction between a student's lagged test score and the average lagged test score of her peers with the interaction term between her lagged test score and the average lagged test score of her new peers. Column 3 and 4 in Table 3 present reduced-form and IV estimates of this heterogeneous peer effects model. The IV coefficient of the interaction term between the average lagged peer test score and a student's own lagged test score (0.113 with a standard error of 0.052) is positive and significant at the five percent level, indicating that an increase in the average lagged test score benefits high-achieving students more relative to low-achieving students.

5 Conclusion

We provide empirical evidence on the existence and the structure of peer effects in middle school using a unique longitudinal data set from China. The peer effects literature seems to be dominated by discussions on the relection problem and the selection issues, whereas little attention is being paid to the potential correlation in measurement errors between the individual- and the peer-level regressors, which we find important in our data. Such a correlation in measurement errors would arise if we simultaneously control for the lagged individual and peer test scores as the two measures are subject to transitory common shocks due to the continuing presence of former peers in a student's current peer group. An important contribution of this paper is to clarify the impact of lagged transitory common shocks on estimates of peer effects. We derive formally that a positive correlation in measurement errors between the individual- and the peer-level regressors will lead to a negative bias in the estimate of peer coefficient, and provide empirical evidence that the transitory-common-shock problem is more than theoretical. We propose an empirical strategy to circumvent the transitory-common-shock problem by using the lagged test score measures of new peers, whose measurement error is uncorrelated with the measurement error in lagged individual test score, to instrument for the corresponding lagged test score measures of all peers.

Our main identification strategy uses within-school variation in peer composition across adjacent cohorts to control for student sorting across schools and the unobserved school characteristics that affect student outcomes. Our within-school IV estimate of the linear-in-means peer effects coefficient is positive but insignificant. While we cannot reject the null hypothesis of no peer effects, we also cannot reject relatively large peer effects that have been found in the previous literature. Estimates of a heterogeneous peer effects model, however, show evidence in support of the single-crossing property: an increase in average peer ability benefits high-achieving students more than low-achieving students.

6 Appendices

6.1 Appendix A The Within-School Specification

We are interested in the within-school estimation of equation (1) in the text

$$Y_{ics} = \beta A_{ics} + \lambda \bar{A}_{(-i)cs} + \phi_s + \kappa_{cs} + v_{ics} \quad (1)$$

Taking average of equation (A1) for all students in school s yields

$$\bar{Y}_s = \beta \bar{A}_s + \lambda \bar{A}_s^* + \phi_s \quad (A1)$$

where $\bar{Y}_s = \frac{1}{n_s} \sum_i Y_{ics}$, $\bar{A}_s = \frac{1}{n_s} \sum_i A_{ics}$, $\bar{A}_s^* = \frac{1}{n_s} \sum_i \frac{n_{cs}-1}{n_{cs}} A_{ics}$, and n_s and n_{cs} represent, respectively, the total number of students in school s and the total number of students in cohort c in school s . Here \bar{A}_s^* differs from \bar{A}_s because we use leave-out average peer ability in equation (1). Note that κ_{cs} and v_{ics} are not included in equation (A1) as they both have zero means at the school level. We can transform Y_{isc} , A_{isc} , and $\bar{A}_{(-i)cs}$ into derivations from their school means such that $y_{ics} = Y_{isc} - \bar{Y}_s$, $a_{ics} = A_{ics} - \bar{A}_s$, and $\bar{a}_{(-i)cs} = \bar{A}_{(-i)cs} - \bar{A}_s^*$. Subtracting equation (A1) from equation (1) yields the within-school specification of the education production function

$$y_{ics} = \beta a_{ics} + \lambda \bar{a}_{(-i)cs} + \kappa_{cs} + v_{ics}$$

Under the identification assumption that κ_{cs} is uncorrelated with $\bar{a}_{(-i)cs}$, κ_{cs} and v_{ics} can be subsumed into a general error term ϵ_{ics} , such that $\epsilon_{ics} = \kappa_{cs} + v_{ics}$. Thus, the within-school specification of the education production function takes the form of equation (2) in the text:

$$y_{ics} = \beta a_{ics} + \lambda \bar{a}_{(-i)cs} + \epsilon_{ics} \quad (2)$$

Consider the following model generating the lagged test score (X_{ics}):

$$X_{ics} = A_{ics} + V_{ics}$$

where V_{ics} is a stochastic individual error term that is uncorrelated with A_{ics} and ϵ_{ics} . Let W_{ics} denote the average lagged test score of student i 's peers such that:

$$W_{ics} = \bar{X}_{(-i)cs} = \bar{A}_{(-i)cs} + U_{ics}$$

where $U_{ics} = \bar{V}_{(-i)cs}$. The within-school transformation of X_{ics} and W_{ics} can be written as follows

$$x_{ics} = X_{ics} - \bar{X}_s = (A_{ics} - \bar{A}_s) + (V_{ics} - \bar{V}_s) = a_{ics} + v_{ics} \quad (A2a)$$

$$w_{ics} = W_{ics} - \bar{W}_s = (\bar{A}_{(-i)cs} - \bar{A}_s^*) + (U_{ics} - \bar{U}_s) = \bar{a}_{(-i)cs} + u_{ics} \quad (A2b)$$

Note that the above within-school transformation allows \bar{V}_s and \bar{U}_s to be nonzero. For example, if a middle school always draw students from a primary school that manipulates the test scores of its students by lowering the grading standards, \bar{V}_s and \bar{U}_s would both be positive. Such across-school variation in measurement errors, however, is accounted for in the within-school estimation. Substituting equations (A2a) and (A2b) into equation (2) yields the empirical specification, equation (4), in the text:

$$y_{ics} = \beta x_{ics} + \lambda w_{ics} + \psi_{ics} \quad (4)$$

where $\psi_{ics} = \epsilon_{ics} - \beta v_{ics} - \lambda u_{ics}$.

6.2 Appendix B Proof of Proposition 2

Let n denotes the total number of students in the sample. Under assumptions (A1)-(A3), the plims of the variance and covariance terms are:

$$\begin{aligned} p \lim \sum \frac{(x - \bar{x})^2}{n} &= \sigma_a^2 + \sigma_v^2 \\ p \lim \sum \frac{(w - \bar{w})^2}{n} &= \sigma_a^2 + \sigma_u^2 \\ p \lim \sum \frac{(w - \bar{w})(x - \bar{x})}{n} &= \rho \\ p \lim \sum \frac{(x - \bar{x})(y - \bar{y})}{n} &= \beta \sigma_a^2 \\ p \lim \sum \frac{(w - \bar{w})(y - \bar{y})}{n} &= \lambda \sigma_a^2 \end{aligned}$$

The within-school estimator $\hat{\beta}_W$ is

$$\hat{\beta}_W = \frac{\sum(w - \bar{w})^2 \sum(x - \bar{x})(y - \bar{y}) - \sum(w - \bar{w})(x - \bar{x}) \sum(w - \bar{w})(y - \bar{y})}{\sum(x - \bar{x})^2 \sum(w - \bar{w})^2 - (\sum(w - \bar{w})(x - \bar{x}))^2} \quad (\text{A3})$$

Taking the plim of (A3) and substituting the above plims of the variance and covariance terms yield,

$$\begin{aligned} p \lim \hat{\beta}_W &= \frac{(\sigma_a^2 + \sigma_u^2) \beta \sigma_a^2 - \rho \lambda \sigma_a^2}{(\sigma_a^2 + \sigma_v^2)(\sigma_a^2 + \sigma_u^2) - \rho^2} \\ &= \beta - \frac{\sigma_v^2 - \frac{\rho^2}{(\sigma_a^2 + \sigma_u^2)}}{(\sigma_a^2 + \sigma_v^2) - \frac{\rho^2}{(\sigma_a^2 + \sigma_u^2)}} \beta - \frac{\sigma_a^2}{(\sigma_a^2 + \sigma_u^2)} \frac{\rho}{(\sigma_a^2 + \sigma_v^2) - \frac{\rho^2}{(\sigma_a^2 + \sigma_u^2)}} \lambda \end{aligned}$$

By the same argument,

$$\begin{aligned} \hat{\lambda}_W &= \frac{\sum(x - \bar{x})^2 \sum(w - \bar{w})(y - \bar{y}) - \sum(w - \bar{w})(x - \bar{x}) \sum(x - \bar{x})(y - \bar{y})}{\sum(x - \bar{x})^2 \sum(w - \bar{w})^2 - (\sum(w - \bar{w})(x - \bar{x}))^2} \\ p \lim \hat{\lambda}_W &= \frac{(\sigma_a^2 + \sigma_v^2) \lambda \sigma_a^2 - \rho \beta \sigma_a^2}{(\sigma_a^2 + \sigma_u^2)(\sigma_a^2 + \sigma_v^2) - \rho^2} \\ &= \lambda - \frac{\sigma_u^2 - \frac{\rho^2}{(\sigma_a^2 + \sigma_v^2)}}{(\sigma_a^2 + \sigma_u^2) - \frac{\rho^2}{(\sigma_a^2 + \sigma_v^2)}} \lambda - \frac{\sigma_a^2}{(\sigma_a^2 + \sigma_u^2)} \frac{\rho}{(\sigma_a^2 + \sigma_v^2) - \frac{\rho^2}{(\sigma_a^2 + \sigma_v^2)}} \beta \end{aligned}$$

For the special case where $\rho = 0$ in Proposition 1,

$$\begin{aligned} p \lim \hat{\beta}_W &= \frac{\sigma_a^2}{\sigma_a^2 + \sigma_v^2} \beta \\ p \lim \hat{\lambda}_W &= \frac{\sigma_u^2}{\sigma_a^2 + \sigma_u^2} \lambda \end{aligned}$$

6.3 Appendix C Proof of Proposition 3

We are interested in using $w_{ics, new}$ as an instrument for w_{ics} to estimate the regression:

$$y_{ics} = \beta x_{ics} + \lambda w_{ics} + \psi_{ics}$$

where $\psi_{ics} = \epsilon_{ics} - \beta v_{ics} - \lambda u_{ics}$.

$$\begin{pmatrix} \hat{\beta}_{W,IV} \\ \hat{\lambda}_{W,IV} \end{pmatrix} = (Z'X)^{-1}Z'Y = \left[\begin{pmatrix} x_{ics} \\ w_{ics,new} \end{pmatrix} (x_{ics} \ w_{ics}) \right]^{-1} \begin{pmatrix} x_{ics} \\ w_{ics,new} \end{pmatrix} (y_{ics})$$

$$\hat{\lambda}_W = \frac{\sum(x-\bar{x})^2 \sum(w_{new} - \bar{w}_{new})(y - \bar{y}) - \sum(w_{new} - \bar{w}_{new})(x - \bar{x}) \sum(x - \bar{x})(y - \bar{y})}{\sum(x - \bar{x})^2 \sum(w_{new} - \bar{w}_{new})(w - \bar{w}) - (\sum(w - \bar{w})(x - \bar{x}))(\sum(w_{new} - \bar{w}_{new})(x - \bar{x}))} \quad (A4)$$

Under assumptions (B1)-(B4), the plims of the variance and covariance terms are:

$$\begin{aligned} p \lim \sum \frac{(x - \bar{x})^2}{n} &= \sigma_a^2 + \sigma_v^2 \\ p \lim \sum \frac{(w_{new} - \bar{w}_{new})(w - \bar{w})}{n} &= \text{cov}(\bar{a}_{(-i)cs}, \bar{a}_{(-i)cs,new}) + \text{cov}(u_{ics}, u_{ics,new}) \\ p \lim \sum \frac{(w_{new} - \bar{w}_{new})(x - \bar{x})}{n} &= 0 \\ p \lim \sum \frac{(x - \bar{x})(y - \bar{y})}{n} &= \beta \sigma_a^2 \\ p \lim \sum \frac{(w_{new} - \bar{w}_{new})(y - \bar{y})}{n} &= \lambda \text{cov}(\bar{a}_{(-i)cs}, \bar{a}_{(-i)cs,new}) \end{aligned}$$

Taking the plim of (A4) and substituting the above plims of the variance and covariance terms yield,

$$p \lim \hat{\lambda}_{W,IV} = \frac{\text{cov}(\bar{a}_{(-i)cs}, \bar{a}_{(-i)cs,new})}{\text{cov}(\bar{a}_{(-i)cs}, \bar{a}_{(-i)cs,new}) + \text{cov}(u_{ics}, u_{ics,new})} \lambda$$

References

- [1] Andreas Ammermueller and Jorn-Steffen Pischke. Peer effects in European primary schools: Evidence from the Progress in International Reading Literacy Study. *Journal of Labor Economics*, 27(3):315–348, 2009.
- [2] Joshua D. Angrist and Kevin Lang. Does school integration generate peer effects? Evidence from Boston’s Metco program. *American Economic Review*, 94(5):1613–1634, 2004.
- [3] Julian R. Betts and Anne Danenberg. School accountability in california: An early evaluation. *Brookings Papers on Education Policy*, 2002:123–197, 2002.
- [4] Julian R. Betts and Andrew Zau. Peer groups and academic achievement: Panel evidence from administrative data. Unpublished manuscript, University of California, San Diego, 2004.
- [5] William A. Brock and Steven N. Durlauf. Interactions-based models. In James J. Heckman and Edward Leamer, editors, *Handbook of Econometrics*, volume 5, chapter 54, pages 3297–3367. Elsevier Science, Amsterdam, 2001.
- [6] Katherine Carman and Lei Zhang. Classroom peer effects and academic achievement: Evidence from a Chinese middle school. Unpublished manuscript, Clemson University, 2008.
- [7] Weili Ding and Steven F Lehrer. Do peers affect student achievement in China’s secondary schools? *Review of Economics and Statistics*, 89(2):300–312, 2007.
- [8] Esther Duflo, Pascaline Dupas, and Michael Kremer. Peer effects, pupil-teacher ratios, and teacher incentives: Evidence from a randomized evaluation in Kenya. unpublished manuscript, 2007.
- [9] Esther Duflo, Pascaline Dupas, and Michael Kremer. Peer effects and the impact of tracking: Evidence from a randomized evaluation in Kenya. unpublished manuscript, 2008.
- [10] Stephen Gibbons and Shqiponje Telhaj. Peer effects and pupil attainment: Evidence from secondary school transition. CEE Discussion Papers 0063, Centre for the Economics of Education, LSE, May 2006. LSE Center for the Economics of Education Discussion Papers No. 0063.

- [11] Eric D. Gould, Victor Lavy, and M. Daniele Paserman. Does immigration affect the long-term educational outcomes of natives? Quasi-experimental evidence. *The Economic Journal*, Forthcoming.
- [12] Eric A. Hanushek, John F. Kain, Jacob M. Markman, and Steven G. Rivkin. Does peer ability affect student achievement? *Journal of Applied Econometrics*, 18:527–544, 2002.
- [13] Caroline M. Hoxby. Peer effects in the classroom: Learning from gender and race variation. NBER Working Papers No. 7867, 2000.
- [14] Caroline M. Hoxby and Gretchen Weingarth. Taking race out of the equation: School re-assignment and the structure of peer effects. Unpublished manuscript, Harvard University, 2005.
- [15] Thomas J. Kane and Douglas O. Staiger. The promise and pitfalls of using imprecise school accountability measures. *Journal of Economic Perspectives*, 16(4):91–114, 2002.
- [16] Changhui Kang. Classroom peer effects and academic achievement: Quasi-randomization evidence from South Korea. *Journal of Urban Economics*, 61:458–495, 2007.
- [17] Fang Lai. How do classroom peers affect student outcomes? Evidence from a natural experiment in Beijing’s middle schools. Unpublished manuscript, New York University, 2007.
- [18] Victor Lavy and Analia Schlosser. Mechanisms and impacts of gender peer effects at school. NBER Working Papers No. 13292, 2007.
- [19] Victor Lavy, Olmo Silva, and Felix Weinhardt. The good, the bad and the average: Evidence on the scale and nature of ability peer effects in schools. Unpublished manuscript, Hebrew University of Jerusalem, 2009.
- [20] Lars Lefgren. Educational peer effects and the Chicago public schools. *Journal of Urban Economics*, 56(2):169–191, 2004.
- [21] Charles F. Manski. Identification of endogenous social effects: The reflection problem. *Journal of Economic Studies*, 14(3):115–136, 1993.

- [22] Patrick J. McEwan. Peer effects on student achievement: Evidence from Chile. *Economics of Education Review*, 22:131–141, 2003.
- [23] Robert A. Moffitt. Policy interventions, low-level equilibria, and social interactions. In Steven N. Durlauf and H. Peyton Young, editors, *Social Dynamics*, chapter 3, pages 45–82. Brookings Institution Press, Washington, D.C., 2001.
- [24] Thomas J. Nechyba. Income and peer quality sorting in public and private schools. In Eric A. Hanushek and Finis Welch, editors, *Handbook of the Economics of Education*, volume 2, chapter 22, pages 1327–1368. Elsevier Science, Amsterdam, 2006.
- [25] Bruce Sacerdote. Peer effects with random assignment: Results from dartmouth roomates. *Quarterly Journal of Economics*, 116(2):681–704, 2001.
- [26] Jacob L. Vigdor and Thomas Nechyba. Peer effects in elementary school: Learning from "apparent" random assignment. Unpublished manuscript, Duke University, 2004.
- [27] Jacob L. Vigdor and Thomas Nechyba. Peer effects in North Carolina public schools. In Ludger Woessmann and Paul E. Peterson, editors, *Schools and the Equal Opportunity Problem*. MIT Press, Cambridge, MA, 2006.
- [28] Diane Whitemore. Resource and peer impact on girls' academic achievement: Evidence from a randomized experiment. *American Economic Review Papers and Proceedings*, 95(2):199–203, 2005.
- [29] Jeffrey E. Zabel. The impact of peer effects on student outcomes in New York City public schools. *Education Finance Policy*, 3:197–249, 2008.
- [30] Hongliang Zhang. Magnet schools and student achievement: Evidence from a randomized natural experiment in China. Unpublished manuscript, Massachusetts Institute of Technology, 2009.
- [31] David Zimmerman. Peer effects in academic outcomes: Evidence from a natural experiment. *Journal of Urban Economics*, 85(1):9–23, 2003.

Table 1 Descriptive Statistics

	Mean	s.d.	within-school s.d.
Panel A: Individual characteristics:			
Female	0.497	0.500	0.500
6th-grade math score	0.035	0.927	0.865
9th-grade math score	-0.005	1.002	0.914
Panel B: Peer-group characteristics			
Proportion of female peers	0.497	0.036	0.029
Proportion of new peers	0.749	0.281	0.151
Average 6th-grade peer math score	0.035	0.363	0.140
Average 6th-grade math score of new peers	0.010	0.383	0.183
Number of observations			7,435

Table 2 OLS and Within-School Estimation

	Dependent variable: 9th-grade math score	
	OLS (1)	Within (2)
Own lagged test score	0.442*** (0.019)	0.436*** (0.019)
Average lagged peer test score	0.591*** (0.104)	-0.250*** (0.112)
Middle school fixed effects	no	yes
Number of observations		7,435

Notes: All specifications control for a female dummy, the proportion of female peers, and cohort fixed effects. Robust standard errors, adjusted for within-school-cohort clustering, are reported in parentheses. A triple asterisk (***) denotes significant at the 1 percent level.

Table 3 Within-school IV Estimates of Peer Effects

	Homogeneous Model		Heterogeneous Model	
	Reduced-form	IV	Reduced-form	IV
	(1)	(2)	(3)	(4)
Female	-0.020 (0.023)	-0.019 (0.022)	-0.021 (0.023)	-0.021 (0.022)
Proportion of female peers	-0.182 (0.614)	0.281 (0.696)	0.170 (0.622)	0.215 (0.692)
Own lagged test score	0.433*** (0.019)	0.427*** (0.020)	0.437*** (0.018)	0.433*** (0.019)
Average lagged test score of new peers	0.121 (0.087)		0.140 (0.090)	
Average lagged peer test score		0.218 (0.175)		0.246 (0.180)
Average lagged test score of new peers * lagged individual test score			0.087** (0.041)	
Average lagged peer test score * Own lagged test score				0.113** (0.052)

Notes: All specifications control for middle school fixed effects and cohort fixed effects. Robust standard errors, adjusted for within-school-cohort clustering, are reported in parentheses. A triple asterisk (***) denotes significant at the 1 percent level. A single asterisk (*) denotes significant at the 10 percent level.