University of Connecticut

Department of Economics Working Paper Series

The Effect of Classmate Characteristics on Individual Outcomes: Evidence from the Add Health

Robert Bifulco
Syracuse University

Jason Fletcher
Yale University

Stephen Ross
University of Connecticut

Working Paper 2008-21R

August 2008, revised January 2009

341 Mansfield Road, Unit 1063
Storrs, CT 06269–1063
Phone: (860) 486–3022
Fax: (860) 486–4463
http://www.econ.uconn.edu/

Abstract

We use data from the National Longitudinal Study of Adolescent Health (Add Health) to examine the effects of classmate characteristics on economic and social outcomes of students. The unique structure of the Add Health allows us to estimate these effects using comparisons across cohorts within schools, and to examine a wider range of outcomes than other studies that have used this identification strategy. This strategy yields variation in cohort composition that is uncorrelated with student observables suggesting that our estimates are not biased by the selection of students into schools or grades based on classmate characteristics. We find that increases in the percent of classmates whose mother is college educated has significant, desirable effects on educational attainment and substance use. We do not find much evidence that the percent of classmates who are black or Hispanic has significant effects on individual outcomes, on average. Additional analyses suggest, however, that an increase in the percent black or Hispanic may increase dropout rates among black students and post-high school idleness among males.

Journal of Economic Literature Classification: I21, I19, J13, J15

Keywords: Education, Peer Effects, Cohort Study, Substance Abuse

This research uses data from Add Health, a program project designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris, and funded by a grant P01-HD31921 from the National Institute of Child Health and Human Development, with cooperative funding from 17 other agencies. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Persons interested in obtaining data files from Add Health should contact Add Health, Carolina Population Center, 123 W. Franklin Street, Chapel Hill, NC 27516-2524 (addhealth@unc.edu). The authors would like to thank Joseph Altonji, Barry Hirsch, Erdal Tekin, Spencer Banzhaf, and Tom Downes who provided comments on the work presented here, as well as participants at the Syracuse University education policy seminar, the Tufts economics department seminar, the Yale labor economics lunch, and the Georgia State University labor/health economics seminar.

I. Introduction

Social scientists have long been interested in determining whether the characteristics of one's schoolmates influence important economic and social outcomes. Policy developments over the last 10 to 15 years have heightened interest in this question. Changes in the law governing racial desegregation efforts and the growth of the school choice movement have led many local school districts to replace student assignment policies focused on promoting racial integration with policies designed to expand parents' discretion over what school their child attends. Several studies suggest that such policy changes may increase the isolation of minority students and the stratification of schools by measures such as parental education and academic achievement.[1] Whether or not such changes can be expected to exacerbate social and economic inequalities depends on how the student composition of a school influences individual outcomes.

The fundamental problem facing studies of schoolmate or peer effects is that individual children or their parents choose the students' peers. In higher education, several studies including Carrell, Fullerton and West (2008), Foster (2006), Kremer and Levy (2008), Li and Li (In press), Lyle (2007), Sacerdote (2001), Siegfried and Gleason (2006), Stinebrickner and Stinebrickner (2006), and Zimmerman (2003), use the random assignment of students to residential facilities to test for peer effects. For primary education in the United States, the opportunities to exploit random assignment to investigate peer effects has been much more limited. In the only examples of which we are aware, Whitmore (2005) and Cascio and

[1] Since 1990, school segregation has declined more slowly than neighborhood segregation, and the isolation of black students in many areas has increased (Clotfelter, 2004; Vigdor & Ludwig, 2007). Clotfelter, Ladd, and Vigdor (2006) provide evidence that federal court rulings during the period have contributed to these trends. Several studies, including but not limited to Bifulco, Ladd, and Ross (In press), Cullen, Jacob, and Levitt (2005), and Hastings, Kane, and Staiger (2006), indicate that students with college educated parents and high achievers are more likely than others to use expanded choice of schools to avoid concentrations of educationally disadvantaged students and to enroll in schools with other educationally advantaged students. Brunner, Imazeki, and Ross (In Press) find that voting patterns for a school choice program in California were consistent with increases in school segregation. As a result many different types of school choice programs can be expected to increase stratification of schools.

Schanzenbach (2007) both use random assignment in Tennessee's project STAR to examine variation in the gender and age composition, respectively.[2]

In a recent innovation introduced by Hoxby (2000b), variation in student composition across cohorts within schools has been used to identify the effect of peers under the assumption that parents and their children are not able to sort across schools based on differences between the demographic composition of the child's cohort and the average composition of the school. Recent studies applying this approach include Angrist and Lang (2004), Friesen and Krauth (2008), Gould, Lavy, and Passerman (2004), Hanushek, Kain, and Rivkin (2002), Hoxby (2000a, 2000b), Lavy and Schlosser (2007), Lavy, Passerman, and Schlosser (2008). However, because this approach requires data on multiple cohorts from the same schools, studies that use it have had to rely on state and local administrative data sets which provide information on only a small set of outcomes, usually limited to student test scores. As a result our knowledge of the effects of student composition on individual outcomes is still quite limited.

In this study, we use data from the National Longitudinal Study of Adolescent Health (Add Health) to extend this line of research on the effects of school composition. The Add Health is a longitudinal survey program that collects information on a wide range of individual outcomes beginning during the teenage years. The study sample includes students from multiple cohorts in a nationally representative set of schools. These aspects of the Add-Health allow us to use comparisons across cohorts within schools by controlling for school fixed effects and trends to estimate the effect of classmate characteristics on a much wider range of outcomes than have previous studies.

[2] Also see Boozer and Cacciola (2001) who examine the effect of randomly assigned new entrants to STAR classes on student performance exploiting the empirical observation that these students appear to be lower in quality in order to detect the influence of peers. In a developing country context, see Duflo, Dupas, and Kremer (2008).

Our analysis focuses on the effects of the percent minority and the percent with a college educated mother among the students in one's school cohort. Distinguishing the effects of school racial composition from social class composition is potentially important. Analysis by Reardon, Yun, and Kurlaender (2006) demonstrate that policies to promote integration by social class might not significantly reduce racial integration, and vice versa. Estimates of the distinct effects of racial composition and social class composition can inform choices about policy priorities.

Several diagnostic analyses support our use of across cohort variation in student composition to identify the effect of peers on student outcomes. First, we run simulations to examine how much within school variation in cohort composition would be expected if students were assigned randomly to school-specific cohorts, and compare this amount of variation to the amount we actually find in our sample.[3] The results of this comparison indicate that the amount of variation across cohorts within schools that we observe in our sample is quite consistent with random assignment. Second, as suggested by Schlosser and Lavy (2007), we conduct balancing tests which examine whether across cohort variation in peer composition can explain predetermined student attributes; if not, these tests imply that students have not sorted on their observables across cohorts within schools. Third, following Hoxby (2000b), we run placebo tests in which we replace the actual cohort composition measures for each student with measures of the composition of a randomly selected cohort from the same school. In both the balancing and placebo tests, we find no evidence of systematic relationships above what would be expected due to type I error, suggesting that our estimates are not biased by either differences in students across cohorts within schools or by school characteristics which might be correlated with cohort composition measures. Finally, we find that after removing school fixed effects and trends, the residuals of share minority and share of students whose mother has a college degree are

[3] We thank Joe Altonji for this suggestion.

uncorrelated. This result provides further confirmation that unobserved factors that influence within school variation in both cohort composition and student outcomes are not confounding our effect estimates.

Our results indicate that a having a higher percentage of classmates[4] with a college educated mother decreases the likelihood of dropping out of high school, increases the likelihood of attending college, and reduces the likelihood of smoking cigarettes in high school and using marijuana after high school. In addition, a higher share of minority classmates is associated with a higher likelihood of using marijuana in high school and a lower likelihood of binge drinking. The number of significant findings far exceed what might be expected based on type I error rates. We also decompose the sample by race/ethnicity, parent education level, and gender. While these later results are somewhat weak statistically, the findings are suggestive that having a higher percentage of classmates with college educated mothers decreases the likelihood of dropping out primarily among white and Hispanics. Analyses also suggest that the effects of classmates with college educated mothers on educational attainment are large and significant for males, but negligible and insignificant for females. A higher share of minority classmates also appears to increase dropout rates among black students and increase rates of post-high school idleness among males.

The paper is organized as follows. Section II briefly reviews the prior research on the effects of student composition. Section III describes the data we use, section IV explains our identification strategy, and Section V provides evidence on the validity of this strategy. Section VI presents our primary results, and Section VII presents extensions that investigate differences

[4] Throughout this paper we will use the term “classmates” to refer to the students in an individual’s school specific cohort.

in the effects of cohort composition across subgroups. Section VIII concludes by considering the policy implications of our findings.

II. Prior Research

Early empirical work on the effects of classmate characteristics focused on the effects of racially desegregated schools. These studies primarily examined outcomes related to academic achievement and racial attitudes and focused largely on the short-run effects of deliberately moving students to less racially segregated schools. Comprehensive reviews of this early research suggest that the results of desegregation were quite mixed, with some evidence of small, positive effects on the academic achievement of black students and little evidence of consistently positive effects on racial attitudes (Cook, 1984; Schofield, 1995). Much of this literature is based on comparisons of students who attended desegregated schools with students who remained in segregated schools, and has been criticized for failing to adequately control for unobserved differences between these two groups of students. Also, Hanushek, Kain, and Rivkin (2002) point out that desegregation efforts were often accompanied by conflict and resistance, and thus, estimates of the short run effects of desegregation might be contaminated by factors related to the desegregation process.

More recent research has focused on the relationship between student composition and outcomes rather than on the effects of specific desegregation efforts.[5] This more recent research has used two different approaches. The first approach uses arguably exogenous variation across

[5] Some recent studies have examined the effects of desegregation. Using variation in the timing of court ordered desegregation, Guryan (2004) finds that desegregation plans in the 1970s decreased black dropout rates and Ludwig, Lutz & Weiner (2007) find that desegregation decreased homicide victimization rates for both blacks and whites.

schools in student composition to identify effects and the other uses variation across cohorts within schools.[6]

Any study that draws on variation in student composition across schools must address the fact that the composition of students in a school influences parents' decisions about whether or not to enroll their child. As a result, students in integrated schools are likely to differ from students in less integrated schools in difficult to observe ways, and these differences are likely to confound estimates of the effect of student composition. Vigdor and Nechyba (in press) illustrate the potential bias using data from North Carolina. They find evidence of strong peer effects using methods that compare students with high and low achieving classmates, but no evidence of peer effects using comparisons that exploit arguably exogenous changes in school composition associated with administrative redistricting.

To address the nonrandom choice of schools, several studies have tried to use measures of racial composition or segregation from higher levels of aggregation to instrument for school racial composition. Rivkin (2000) uses district level variation in exposure to whites, and finds that racial composition has no effects on test scores, educational attainment or earnings.[7] Boozer, Krueger, and Wolken (1992) use variation across time and states in school racial composition, and find that high white enrollment shares are associated with higher educational and occupational attainment. Evans, Oates, and Schwab (1992) use metropolitan level measures of socioeconomic well-being as instruments, and find no relationship between the percent of economically disadvantaged schoolmates and either teenage pregnancy or drop-out rates. Cutler and Glaeser (1997) and Card and Rothstein (2007) also draw on metropolitan level variation and

[6] Most studies avoid examining variation in composition across classrooms due to concerns of non-random assignment of students into classrooms. See Vigdor and Nechyba (2004) and Zabel (2008) for examples of attempts to examine peer effects within the classroom.

[7] Rivkin's effect estimates control for the academic achievement gains made by students in the school, which of course is one of the mechanisms through which school peers can influence student outcomes.

find that residential segregation by race is associated with lower high school graduation rates, lower wages, and higher rates of single parenthood for blacks and a larger black-white test score gap, respectively.

Although these analyses do not require exogenous selection into schools, they are subject to potential biases related to unobserved differences in students across districts, metropolitan areas, or states. Another limitation of studies that use metropolitan level variation is that they often cannot distinguish between peer effects in schools and the effects of processes that occur in the broader urban environment. Card and Rothstein (2007), for instance, find that more segregated metropolitan areas have larger black-white test score gaps. However, after controlling for residential segregation, school segregation is unrelated to their measure of the test score gap, and they cannot conclusively distinguish the effects of school segregation from the effects of residential segregation.[8]

A second approach to estimating the effect of classmate characteristics exploits variation across cohorts within schools. These studies use data drawn from state or local administrative sources to estimate models that control for school-by-grade fixed effects. Such models arguably isolate idiosyncratic variation in student composition across cohorts within a school. Focusing on within school variation reduces concerns about nonrandom selection across schools and also helps to isolate the effects of student composition from any aspects of school quality that are constant across cohorts.

[8] A recent study by Friesen and Krauth (2007) makes efforts to address both of these limitations. Using data from Alberta, Canada, they examine the relationship between segregation across schools within a community and the variance in high school test scores. To control for the possibility that unobserved heterogeneity among students within a community causes higher levels segregation, they focus on the relationship between changes in the level of segregation and changes in test score variance between sixth and ninth grade, and use plausibly exogenous sources of variation in changes in the level of segregation. They find that increases in sorting by parent education level increases variance in test scores, but increases in ethnic and income segregation does not influence test score variance.

Hoxby (2000b) pioneered this approach using data from Texas, and finds that elementary students have lower test scores when they are in cohorts with a larger share of black students, and that the negative effects are larger for black students than other students.[9] Hanushek, Kain, and Rivkin (2002), using similar data and methods, find that the negative effects of percent black on test scores are significantly higher for high ability black students than either low ability black students or students from other ethnic groups. Angrist and Lang (2002) use data from the Boston area's Metco program, which allows minority students from Boston to attend schools in suburban districts. In contrast to the Texas studies, they find only small effects of an increased share of Metco students on the test scores of non-Metco students--effects that are limited to minority girls.[10] Other studies have used the cross cohort approach to examine the effect of other classmate characteristics. Using data from Israel, Lavy and Schlosser (2007) find significant effects of variation in gender composition on student test scores and Lavy, Schlosser, and Passerman (2007) find significant effects of peer ability. Friesen and Krauth (2008) find that the home language spoken by peers influences academic performance in British Columbia.

The composition of students in one's school or class might influence individual choices and outcomes through a variety of mechanisms. First, schools with concentrations of disadvantaged students might have difficulty garnering educational resources.[11] A second set of mechanisms work through group dynamics, including teacher expectations and motivation, student achievement norms and motivation, pace of instruction, and levels of classroom

---

[9] Hoxby (2000a) uses this approach to examine the impact of class size on student performance.
[10] The Metco students who transferred into suburban schools were a select sample of Boston students, and their effects on group dynamics, may not be typical, which might account for differences between the Metco and Texas findings.
[11] Several studies indicate that schools with concentrations of minority and poor students are less able to attract and retain highly qualified teachers (Betts, Reuben, & Dannenberg, 2000; Clotfelter, Ladd, & Vigdor, 2005; Lankford, Loeb, & Wyckoff, 2002).

disruption. In addition, students from different family backgrounds might bring to school different attitudes towards behaviors such as the use of alcohol or illicit drugs influencing school norms. Finally, school composition might influence the opportunity to make the personal contacts or to develop the modes of behavior that that are crucial for gaining access to jobs and educational opportunities (Clotfelter, 2004; Wells, 1995). Unlike studies that use across school variation, cohort studies only capture mechanisms that can operate through one time (cohort specific) changes in student composition. Naturally, group dynamics and student attitudes can change from cohort to cohort and be captured in cohort studies, while other mechanism such as school resources and culture, selection of teachers into schools, and community engagement are less likely to be represented.

Our study employs this cross cohort approach, and thus plausibly addresses biases associated with self-selection into schools and provides estimates of the effects of classmate characteristics that are clearly distinguished from the effects of residential segregation and other metropolitan level processes. Unlike the studies that have used metropolitan level variation, which have focused on a broad range of outcomes, previous studies that have used the cross-cohort approach have been limited to estimating effects on test scores. We, however, are able to estimate the effects of classmate characteristics on a broad range of outcomes including post high school outcomes like college attendance or employment. Our study is also the first to conduct this type of analysis on a nationally representative sample of students.

III. Data

The data for this study come from the restricted version of the National Longitudinal Study of Adolescent Health (Add Health). The Add Health is a school-based, longitudinal study of the health-related behaviors of adolescents and their outcomes in young adulthood. Beginning with an in-school questionnaire administered to a nationally representative sample of students in grades 7 through 12 in 1994-95 (Wave 1), the study follows up with a series of in-home interviews of students approximately one year (Wave 2) and then six years later (Wave 3). Other sources of data include questionnaires for parents, siblings, fellow students, and school administrators. By design, the Add Health survey included a sample stratified by region, urbanicity, school type, ethnic mix, and size.[12]

Over 20,000 individuals completed the full, Wave 1 survey. However, because we are interested primarily in post-high school outcomes, we drop the 6,000 students who were not in grades 9-12 (grades 10-12 for three year high schools) during Wave 1 and approximately 80 additional students who report still being in high school during Wave 3. The remaining students range from 20 to 24 years old at the time of the Wave 3 survey. In addition, we drop approximately 4,500 individuals who were not followed through Wave 3, and, because our identification strategy depends upon having multiple cohorts within schools, we drop 500 students who are in schools that do not have a 10th, 11th, and 12th grade. Finally, we drop approximately 150 students who did not identify themselves as either white, black, Hispanic, or Asian and 60 students in grades with fewer than 10 sample students. The samples restrictions leave an analysis sample of approximately 9,500 students in 75 high schools, although the sample varies slightly by the outcome of interest.

Among these various sample edits, the approximately 4,500 individuals who were dropped because they were not followed through Wave 3 are of particular concern. If within-school variation in cohort composition is systematically related to the probability of non-response in Wave 3, then our estimates of the effects of cohort composition could be biased. To test this possibility, we regressed an indicator of whether or not a student was followed through

[12] See Udry 2003 for full description of the Add Health data set.

Wave 3 on our cohort composition measures, the set of control variables described below, school fixed effects, and school specific trends. The results indicate that the cohort composition measures are not related to probability of non-response in Wave 3, and thus, dropping non-responders should not introduce any bias into our effect estimates.

We create our cohort-level variables by using the items from the in-school sample of Add Health at Wave 1. The in-school survey was administered to over 90,000 students and asked a limited amount of information, including race/ethnicity and maternal education, for (in principle) a census of students in each sampled school. This feature of the Add Health allows us to reduce the error in our aggregate measures of classmate characteristics.

Table 1 provides descriptive statistics for all of the variables that we use in the analyses that follow. The variables include those we use as outcome measures, our key cohort composition variables, a set of baseline controls that include grade fixed effects and student attributes directly related to the cohort variables, an extended set of controls that are unlikely to be influenced by school experiences plus a test score as a measure of ability and an additional set of family variables. Table 2 provides means and standard deviations for the outcome variables and our cohort composition variables for different racial groups and for groups defined by the level of mother's education.

Black students, Hispanic students, and students with lower levels of parental education do worse on several outcome measures. Each has relatively high dropout rates, low rates of college attendance, low test score levels and high rates of idleness. Black and Hispanic students also attend schools with relatively high percentages of minority students, and Hispanic students and students with lower levels of parental education attend schools with low percentages of college educated mothers. The key task of the analyses that follow is to determine whether any of these relatively poor outcomes of students from disadvantaged groups might be attributed to the relatively high share of their classmates who are black or Hispanic or whose mother's lack a college education. White students, for their part, are more likely to report that they smoke, use marijuana and binge drink than are other groups, and we are also interested in whether any of those differences might be attributed to the group dynamics fostered by a higher percentage of white classmates.

IV. Identification Strategy

To avoid issues of selection across schools and to isolate the effect of classmate characteristics from other aspects of school quality, our identification strategy relies on variation across cohorts within schools. To implement this strategy we estimate regressions of the following form:

$$y_{isc} = \alpha_c + \beta_s + \delta_s c + x_{isc}\phi + \varphi P^E_{sc} + \gamma P^M_{sc} + \varepsilon_{isc}$$

y_{isc} is an outcome measure for individual i from school s and cohort c; α_c is cohort or grade specific effect; β_s is a school fixed effect; $\delta_s c$ is a school-specific time or cohort trend where c takes the value of 0 for the oldest cohort and increases by 1 for each successive cohort;[13] x_{isc} is a vector of student level covariates; P^E_{sc} is the percent of students in school s and cohort c with a college educated mother; P^M_{sc} is the percent of students in school s and cohort c who are either black or Hispanic; and ε_{isc} is a random error term which might be correlated across observations from the same school.[14]

[13] All students are observed at the same points in time, so referring to these are school-specific cohort or grade trends is more accurate than referring to them as time trends. However, in studies that use administrative data variation across cohorts is often referred as variation within schools over time, and so we use the two terms interchangeably.
[14] Thus, for all our regressions we compute standard errors that are robust to any type of clustering within schools.

We examine several different outcome measures including whether or not the individual has dropped out of high school, has attended college, is idle (i.e. neither working nor attending school), uses cigarettes, uses marijuana or engages in binge drinking. Each of these variables is measured using Wave 3 of the Add Health and thus represent post-high school outcomes. We also examine the individual's post-high school PVT test score as a measure of cognitive ability[15] and indicators of cigarette use, marijuana uses and binge drinking from Wave 1 when the individuals are still in high school.

Students from different cohorts are in different grades during the initial wave of the Add Health, and thus we include a cohort specific effect, α_c, to control for these differences in age during the initial Wave 1 interview. Including school fixed effects, β_s, ensures the estimation of classmate effects is based on comparisons across cohorts within a school, and controls for unobserved differences in average student characteristics across schools as well as for aspects of school quality that are constant across cohorts within a school.

Although school fixed effects provide powerful controls for selection across schools, differences in peer characteristics across cohorts within a school might be systematically correlated with unobserved variables that affect achievement. Schools that show systematic trends in peer characteristics are of particular concern. For instance, parents might be able to discern when the share minority in a school is increasing over time, and as a result, students from older cohorts who select into the school might differ in systematic, but unobserved ways from students in younger cohorts. Similarly, the quality of teachers who can be attracted and retained to teach younger cohorts might differ from those who can be attracted and retained to teach the

[15] The Add Health Picture Vocabulary Test (AHPVT) is a computerized, abridged version of the Peabody Picture Vocabulary Test-Revised (PPVT-R). The AHPVT is a test of hearing vocabulary, designed for persons aged 2 1/2 to 40 years old who can see and hear reasonably well and who understand standard English to some degree. The test scores are standardized by age. Some psychologists interpret PVT scores as a measure of verbal IQ. Information on the test is provided online at http://www.cpc.unc.edu/projects/addhealth/files/w3cdbk/w3doc.zip.

older cohorts. In either case, unobserved differences in student and teacher quality across cohorts within the same school could be correlated with differences in the share minority, and would confound estimates of the effect of share minority on outcomes. To address this concern we control for school specific linear trends, $\delta_s c$. As a result our effect estimates are based on the correlation between deviations from the school specific trend in a cohort's student composition and deviations from school specific trends in student outcomes.

The cohort fixed effects together with the school specific linear trends also help address another problem created by the structure of our data. Unlike school administrative data, we do not observe multiple cohorts passing through the same grade, but rather observe all cohorts at the same time in different grades. Therefore, we cannot explicitly control for school-grade fixed effects, and systematic changes in cohort variables across grades might be correlated with differences in outcomes across grades. For example, because minorities and those whose parents have less education are more likely to drop out sometime between grades 9 and 12, the cohorts that are in later grades during Wave 1 will have lower percentages of minority and higher percentages of students with college educated mothers than cohorts in earlier grades during Wave 1. Also, because the least motivated students are more likely to dropout as they age, students in the later grades during Wave 1 might be systematically different than students in the earlier grades on unobserved characteristics that influence outcomes. The average effect of any systematic, unobserved differences between older cohorts and younger cohorts that arise because of this type of attrition will be controlled for by the cohort fixed effects. Because the effects of dropouts on cohort composition and on unobserved student characteristics are likely to be larger in some schools than others, however, cohort fixed effects may not be sufficient to eliminate potential biases. If we assume, however, that the effects of dropouts or other grade specific effects on cohort composition and on unobserved student characteristics are approximately linear in grades in most schools, then school specific trends will break any correlation between the two variables, and thus minimize any potential biases. While this limitation of our data requires more assumptions than traditional applications of the cohort approach, it does not impact the validity of our diagnostics for instrument exogeneity, and therefore simply requires us to lean more heavily on those diagnostics.

Deviations from school trends in student composition, which are difficult for parents and students to predict, are unlikely to influence their decision to attend a school, and thus, such deviations from trend are arguably idiosyncratic. Nonetheless, race and parents' education are likely to be correlated with several other factors that influence outcomes. Thus, even if deviations from school trends in cohort composition are truly idiosyncratic, the students in cohorts with higher than predicted percentages of minority students or college educated parents will differ from students in other cohorts in systematic and potentially important ways. Including individual controls for race and parent education will prevent these systematic differences from confounding our effect estimates. Also, even if deviations from school trends in student composition do not influence a student's initial decision to attend a school, students might systematically opt out of a school that they find unsatisfactory after their initial experience, potentially introducing a source of omitted variable bias into our school trend model. Thus, we include a full set of controls for individual student characteristics measured during Wave 1, x_{isc}. To ensure that our results do not hinge on judgments about which individual controls to include, we estimate models with three different specifications of x_{isc}.

Our baseline model only includes controls for the individual students variables directly related to the school cohort variables: race/ethnicity[16] and years of education for the parent who responded to the parent survey.[17] A second set of models include additional controls for clearly exogenous student characteristics including gender, age, whether or not the responding parent reports being born in the U.S., number of years the family has lived in the U.S.,[18] a dummy variable indicating whether the parent information was reported or imputed, plus the student's PVT test score during Wave 1 which, while potentially influenced by cohort composition, is our best available proxy for a student's underlying cognitive ability.[19] A third set of models adds an extended set of family background variables including log of family income, a single parent family indicator, an indicator of whether or not a student lives with both biological parents, the number of older siblings, and indicators of whether the student reports having discussed school or grades with a parent in the last month, whether one of the student's parents report being a member of a parent/teacher organization, and whether the responding parent reports that he/she or the student's other biological parent has alcoholism. All these variables are measured during Wave 1. These variables provide powerful protection against any potential omitted variables

[16] We include mutually exclusive and exhaustive categories of race and ethnicity, including non-Hispanic white, non-Hispanic black, Hispanic, and Asian. Students who reported being multirace were designated as black if the races were white and black, and designated as Asian if the races were Asian and white.

[17] Years of education of survey respondent is used since it is our most error free measure of parental education levels. In principle, we might have included dummy variables for mother's educational attainment paralleling the construction of the mother's education cohort variable. Models controlling for those variables instead of parental education yield results that are very similar to the estimates presented in the paper.

[18] As reported by the student, the variable is set equal to the age of the parent if the student was born in the U.S.

[19] With the exception of the model for Wave 3 test scores (where estimates are insignificant anyway), the estimates on cohort variables are nearly identical whether or not the set of controls for student attributes includes Wave 1 test scores. While including test score has no substantive effects on our estimates, in principle, including this variable changes the interpretation of our estimates slightly. The baseline models can be interpreted as estimating the total effect of changes in classmate characteristics that operate through dynamics that vary across cohorts within the school. The models that include the Wave 1 test score give us estimates of the effects of classmate characteristics that operate through dynamics that vary across cohorts within a school and independently of any effects on cognitive development through Wave 1.

bias. Many of them, however, might be influenced by a student's experiences in school and by the student's behavior, and thus we do not include them in our baseline models.

Our variables of interest are measures of student composition for each cohort within each school. We focus on the percent disadvantaged minority, which is the percent black plus the percent Hispanic, in the school specific cohort, P_{sc}^{M}, and the percent of students in the cohort who have a college educated mother, P_{sc}^{E}. The racial composition of schools has been a leading policy concern dating back to the Supreme Court's landmark ruling in Brown v. Board of Education (1954), and much of the literature on the effects of student composition has focused on racial composition.[20] It is also important, however, to focus on segregation by other family background characteristics, and particularly parental education. Evidence from a wide range of school choice programs indicates that students whose parents have higher levels of education are more likely than other parents to use expanded schooling options to avoid schools with concentrations of disadvantaged students and to attend schools with higher levels of achievement. Thus, the growth of student assignment policies that emphasize parental choice is likely to increase stratification of schools by levels of parental education. Also, it is important to distinguish the effects of racial composition from class composition because policies to decrease segregation by class will not necessarily decrease segregation by race, and vice-versa.

As mentioned above, some of the mechanisms through which student composition might influence individual outcomes are constant across cohorts within schools. For instance, a school's ability to garner resources is likely to be determined largely by the composition of the school as a whole and may not vary across cohorts within the school. Similarly, teacher

[20] We also ran models that use percent black rather than percent black or Hispanic. In most cases, the estimated effects of percent black were similar to the estimated effects of percent black or Hispanic. An important exception for models predicting marijuana use in high school is noted below.

expectations and motivation might be influenced as much by the composition of preceding cohorts as by the composition of the current cohort. By relying on within school variation in cohort composition, however, our estimates will miss any effect that the student composition of the school as a whole has on student outcomes. Thus, we will interpret our estimates as the effects of cohort composition that operate through the mechanisms of cohort specific group dynamics, holding other aspects of school quality constant. It is important to realize that this effect may be only part of the total effect that school composition has on student outcomes.

In order to interpret our estimates as the effects that operate through cohort specific group dynamics, the variation in cohort composition that we use to derive our estimates must be uncorrelated with other aspects of school quality related to teachers, resources or other factors within schools. School fixed effects and trends should ensure that this condition is met. To verify this assumption, we run a placebo test for each of our regressions. Following Hoxby (2000b), our placebo test replaces the measures of student composition for student i's actual cohort with the same measures for a randomly selected younger or older cohort from the same school. The composition of a cohort other than a student's own should not have any causal effect on the student. Thus, if controlling for school fixed effects and trends successfully breaks the correlation between the composition of a student's cohort and other aspects of school quality, then the composition of other cohorts in the same school should not show any effect on student outcomes in these placebo regressions.[21]

V. Evidence on the Identification Strategy

As Lavy and Schlosser (2007) point out in a similar analysis of gender composition effects, the success of our identification strategy rests on two things. First, in order to obtain

[21] This placebo test also helps to diagnose small sample bias. Our data consists of 75 schools with four cohorts each, and we use two degrees of freedom per school to estimate school specific trends. If we have any results driven by small sample bias, these should show up in the placebo regressions as well.

precise estimates, we need sufficient variation in our cohort composition measures after controlling for school fixed effects and trends. Second, in order to make causal interpretations of our effect estimates plausible, deviations from school specific trends in student composition must be uncorrelated with differences in student characteristics across cohorts. In this section we investigate whether or not these conditions are met.

Table 3 examines the extent of variation in cohort composition that is left after removing school fixed effects and trends. As we would expect, most of the variation in our student composition measures is across schools rather than within schools. Removing school fixed effects and trends reduces the standard deviations in the percent of students with college educated mothers by nearly 80 percent and the standard deviation in percent black or Hispanic by more than 90 percent. Thus, our effect estimates are based on small, marginal changes in student composition, and cannot tell us about the effects of moving an individual student across schools with very different student compositions.

Table 3 does, however, suggest that we have sufficient variation to estimate the effects of small changes in cohort composition with reasonable precision. The precision of our estimates depends on our sample size and on the absolute magnitude of the variation we use. The variation in our data in the percent of mothers with college and the percent black or Hispanic after removing school fixed effects and trends is 20 to 80 percent greater than the within school variation in gender composition reported by Lavy and Schlosser (2007), which was enough variation for those authors to obtain statistically significant estimates of gender composition effects. It is fortunate that we have greater within school variation in our student composition measures than Lavy and Schlosser, because our data has roughly one-third as many schools and thus fewer school-specific-cohorts than they do, which reduces the precision of our estimates.

Our identification strategy assumes that variation in student composition across cohorts within a school is generated randomly. To test whether the amount of variation observed in Table 3 is consistent with random assignment of students across school specific cohorts we ran a series of simulations. In each simulation, we randomly match students in our sample to the school and grade specific slots in our sample, and use the resulting distribution of students across schools and cohorts to compute standard deviations for the cohort composition variables. Across 50 simulations of this kind, the average standard deviation for percent of students with college educated mothers in the same school and cohort was 0.029 and for percent of black or Hispanic the average standard deviation was 0.025, which are quite similar to the standard deviations of 0.031 and 0.025 reported in Table 3.[22] These results indicate that the amount of variation across cohorts within schools that we observe in our sample is quite consistent with random assignment of students.

Further, any systematic selection process that might generate variation in cohort composition measures would be expected to have similar effects on the percent minority and the percent of students with college educated mothers. Thus, variation across schools in race and parent education are often highly correlated. Our identification strategy, however, assumes exogenous variation in student composition across cohorts within a school, after controlling for school specific trends. Thus, we expect the variation in percent of students with college educated parents and the variation in the percent black or Hispanic isolated by our regressions to be uncorrelated. Therefore, the correlation between the residuals of our two cohort composition measures, after removing school fixed effects and trends, provides another, informal test of our

[22] The standard deviation around the mean standard deviation for percent with college educated mothers was 0.002 and for percent minority was 0.001 placing the actual standard deviations well within the 95% confidence intervals.

key identifying assumption. The correlation between these residuals is -0.01, which confirms the expectations associated with exogenous cohort variation.

Another informal test of our key identifying assumption can be conducted by checking whether deviations from school specific trends in our cohort composition measures are correlated with deviations from school specific trends for a variety of student background characteristics (balancing tests). If these deviations are uncorrelated, the analysis supports the premise that school trends capture any systematic selection (due to either sorting or attrition) on student observables. Further, if one uses the degree of selection on observables as a guide to the degree of selection on unobservables as suggested by Altonji, Elder and Tabor (2005), null results on the balancing tests would support the assumption that our model specification identifies variation in cohort composition unrelated to unobservables that determine student outcomes.[23]

We performed this check by regressing different student background characteristics on our measures of cohort composition controlling for cohort fixed effects, school fixed effects, school trends, the student's race and the education level of the student's mother. If deviations from school trends in parent education levels and student composition are truly idiosyncratic, then once we control for the student's own race and parent's education level, any correlation between deviations from school trends in the cohort composition variables and deviations from school trends in other student background characteristics should be removed.

Table 4 presents the results of these balancing tests. The results of 11 separate regressions and a total of 22 coefficient estimates are presented. In the absence of any systematic selection of students, we would expect two to three of these coefficients to be significantly different than zero at the 0.10 confidence level and one to be statistically significant

[23] Similar logic has been used in recent studies of neighborhood effects by Grinblatt, Keloharju, and Ikaheimo (2008) and Bayer, Ross, and Topa (In press) that document no sorting on unobservables over space conditional on their models.

at the 0.05 level due merely to chance. On the whole the results from our t-tests are equivocal. Four coefficients are significantly different than zero at the 0.10 level, which is more than we would expect, but none are significantly different from zero at the 0.05 level, which is less than we would expect. Tests using t-statistics from the same regression, however, are not independent. More telling are F-tests for the joint significance of the two cohort variables in each regression. Only one of the eleven F-statistics reported is significantly different than zero at the 10 percent level and that one has a p-value very close to 0.10. One rejection of the null at the 0.10 level in eleven tests is just what we would expect if the variation of cohort composition measures were in fact idiosyncratic. The balancing tests, then, provide general evidence that school specific trends are sufficient to isolate variation in cohort composition that is unrelated to student observables, and thus, there is little reason to suspect that differences in unobserved student characteristics across cohorts within a school are biasing our effect estimates.

Despite our appeal to type I error, one might be concerned about the specific rejections of the null hypothesis in our balancing tests. To address this concern we estimate models with and without the variables examined in the balancing test. The first specification includes only the covariates also used as controls in our balancing tests, the second specification adds a substantial number of controls including three of the controls that fail the balancing test at the 0.10 level, and the third specification includes all remaining covariates including one additional variable that failed the balancing tests. If our identification strategy is working, adding these control variables should have no influence on the estimated coefficients for our cohort composition variables. Again following the intuition behind Altonji, Elder and Tabor (2005), the impact of including observable student attributes on peer effect estimates likely provide a good indication of the potential bias from unobservables, and so if adding observable controls has little impact on estimates it is reasonable to presume that there is little bias from student unobservables.

VI. Primary Results

Table 5 presents estimates of the effects of our cohort composition variables on ten different outcomes. All of the estimates presented in Table 5 are from regressions that include controls for cohort fixed effects, school fixed effects, and school trends. For each outcome, we present estimates from regressions that include the baseline set of student covariates listed above, the baseline set of covariates plus extended covariates that include the Wave 1 PVT test score, and the extended set of covariates plus a set of additional family background controls. The estimates on our cohort composition variables are quite stable across each specification of student covariates. The robustness of our estimates with respect to choice of student covariates provides additional support for the results of the balancing tests presented in Table 4.

As seen in Table 5, the percent of students in the cohort with a college educated mother shows significant effects on the decisions to drop out of high school, to attend college, to smoke during high school, and to use marijuana post-high school. Most would consider the direction of these significant effects desirable. The point estimates imply that a 1 percentage point increase in the percent of students whose parents are college educated is associated with a decrease in the likelihood of dropping out of about 0.3 percentage points, an increase in the likelihood of attending college of between 0.4 to 0.5 percentage points, and decreases in the likelihoods of smoking during high school and of using marijuana after high school of about 0.4 percentage points each.[24]

[24] We also estimated alternative versions of these regressions using the average years of mother's education in the cohort instead of percent of students with a college educated mother. For all of our findings, the point estimates on this variable were in the same direction and implied effects of a similar magnitude as the coefficient on percent with college educated mothers. However, the level of statistical significance fell for drop-out and college attendance. In

To help think about the magnitude of these effect estimates, we can estimate how much reducing disparities in exposure to classmates with college educated mothers would reduce disparities in dropout and college attendance rates. The figures in Table 2 indicate that the percent of college educated mothers among the classmates of students whose own parents are college graduates is 11.1 percentage points higher than that of students whose own mothers have no college experience. Also, among individuals in our sample, those whose own parents are college educated are 12.3 percentage points less likely to drop out of high school than students whose mothers do not have any college. The effect estimates in Table 5 imply that reducing the gap in exposure to classmates with college educated mothers by half (5.5 percentage points), would decrease the gap in dropout rates nearly 14 percent (from 12.3 to 10.6 percentage points). Similar calculations indicate that reducing the gap in exposure to classmates with college educated mothers by half, would decrease the gap in college attendance between individuals whose own parents are college graduates and individuals whose own parents have no college by nearly 7 percent (from 34.7 to 32.2 percentage points).

Higher shares of students who are black or Hispanic in a cohort are associated with a greater likelihood of using marijuana in high school and a smaller likelihood of binge drinking after high school. A one percentage point increase in the percent minority is associated with an increase in the likelihood of using marijuana during high school of about 0.4 percentage points and a decrease in binge drinking after high school of more than 0.5 percentage points. That increases in the percent minority increase marijuana use is somewhat surprising given that black students are less likely than whites to report using marijuana in high school (see Table 2). Hispanic students, however, report a higher rate of marijuana use than other groups during the

addition, higher average years of mother's education is associated with significantly lower likelihood of using marijuana in high school while the estimates based on percent of college educated mothers are small and statistically insignificant on this outcome.

high school years, and when we estimate the effect of percent black and percent Hispanic in separate regressions, we find that the estimated effect is more than three times larger for percent Hispanic than for percent black, and is only significant for percent Hispanic (results not shown). Just as noteworthy as these significant effects on substance use, the estimated effects of minority share on educational attainment, post-high school test scores, and idleness are small and statistically insignificant.

In general, then, the results in Table 5 suggest that the percent of classmates whose mother's are college educated can have significant effects on educational attainment, and both percent of college educated mothers and the share of classmates who are black or Hispanic can have significant effects on substance use. Higher levels of parent education among one's classmates generally have desirable effects, while higher levels of minorities have mixed effects. Interestingly, the level of parental education in one's cohort has no lasting influence on cognitive development as indicated by the PVT test score, and thus, the potentially important effects of school cohort composition on individuals would be missed by studies focused solely on test score measures.

Table 6 presents the results of our placebo tests, which are coefficient estimates from regressions that replace the actual cohort composition measures for a student with the composition measures for another randomly selected cohort in the same school. Across the 10 outcomes, 30 regressions and 60 coefficient estimates presented in Table 5, only two coefficients are significantly different than zero at the 0.10 level or lower, only one of these is significant at the 0.05 level, well less than the number of significant coefficients that we would expect to emerge by chance. Both significant coefficients are on the percent black plus percent Hispanic variable in models that have post-high school binge drinking as the outcome. Therefore, with two cohort variables and ten outcomes, significance only occurs for one variable on one outcome with the finding ranging between insignificant and significant at the 0.05 level across the three specifications. These results stand in stark contrast with the causal estimates in Table 5 where out of 10 outcome variables the estimated effects for two different outcomes are significant near or above the 0.01 level across the three specifications, estimated effects on three additional outcomes are significant at the 0.05 level, and estimated effects on one more outcome is significant at the 0.10 level.

The negative and statistically significant coefficients on percent black plus percent Hispanic in the post-high school binge drinking placebo regressions is not unexpected. Recall that the variables in the regressions in Tables 5 and 6 are deviations from school specific trends, and thus, since we cannot randomly assign a cohort to itself, we expect a negative correlation between the actual cohort composition measures used in Table 5 and the cohort compositions of randomly selected different cohorts from the same school used in Table 6. This correlation combined with the strong, positive relationship between the percent black or Hispanic in the student's actual cohort and post-high school binge drinking (see last row of Table 5) is the most likely explanation for the smaller, negative coefficient with marginal statistical significance in our placebo regression.[25] Overall, then, the results in Table 6 provide strong evidence that the estimated effects of the actual cohort composition measures presented in Table 5 are not biased by selection across schools or omitted school characteristics.

By definition, exceptionally large deviations from school trends are unlikely to arise by chance, and one might suspect that non-random factors that cause large changes in cohort

[25] Bayer, Ross, and Topa (2008) find a similar small sample bias whey then test whether block average attributes correlate with the attributes of individual residents because each individual is excluded from calculation of the block means for their own block. Bayer, Ross, and Topa obtain consistent estimates of the correlation by sampling only one person from each block, but they have thousands of blocks in their sample. In our application, this strategy would provide a placebo test with very low power due to the smaller number of within school cohorts.

compositions within a school can simultaneously cause changes in the unobserved characteristics of students in the school. For instance, if the district a school is located in adopts a school choice program sometime between when the twelfth graders and the ninth graders in our sample entered high school, that could simultaneously cause differences in student composition and unobserved student "quality" across cohorts within a school. To test the sensitivity of our results to the inclusion of cases with large deviations of individual cohort compositions from school trends, we identified cases of large deviations, dropped them from our sample, and reestimated our regressions. The results from these alternative regressions were very similar to the results in Table 5.[26]

VII. Results by Subgroups

In this section we examine whether the effects of cohort composition vary across racial/ethnic groups, groups based on mother's education, and student gender. Admittedly, the estimates based on these subsamples are fairly noisy. Taking the results presented in Tables 7 through 9 as a whole, we find four rejections at the 5 percent level or better and a total of eight rejections at the 10 percent level or better out of 60 tests. We would expect three and six rejections at the 5 percent and 10 percent levels, respectively, based just on type I error. Therefore, these findings should be interpreted as suggestive.

First, we estimated our regression models separately for black students, white students, Hispanic students, and Asian students.[27] The results of these regressions are presented in Table

[26] Specifically, we regressed the student's cohort composition measures on a set of school fixed effects and trends, and if the residual from this regression for a particular observation was more than three times the standard deviation of all such residuals, we dropped that observation. Generally, the significant effect estimates became slightly larger and slightly less precise when cases of large deviations were dropped. In no cases, did the results of inference tests change.

[27] Differences in the effects of marginal changes in student composition across different types of schools is also potentially interesting. However, with only 75 schools in our sample and relatively small amounts of variation across cohorts within schools, the power we have to distinguish differences in effects across school types is limited. We did run estimates in which we split the sample in half and into thirds based on the percent of college educated

7. These results are from models with the cohort composition variables entered singly, and thus each coefficient reported in Table 7 is from a separate regression. Each regression includes controls for cohort fixed effects, school fixed effects, and school trends as well as the full set of individual and family covariates.[28] The last column presents the results of a Wald test for the joint null hypotheses that the coefficients are equal.[29]

As indicated in the top panel of Table 7, the effects of cohort composition on the decision to drop out vary considerably across ethnic and racial groups. On average, across all groups, the level of parent education among one's classmates has a negative, statistically significant effect on the likelihood of dropping out (see Table 5). In Table 7, we see that this negative effect is driven by moderate effects on whites and large effects on Hispanic students, and the Wald test rejects equality at nearly the 1 percent level. The point estimates in the top panel of Table 7 indicate that a 1.0 point increase in the percent of students with a college educated mother reduces dropout rates among Hispanics by nearly 1.6 percentage points, much more than for white and Asian students. This result is important given the high dropout rates among Hispanics.

While just missing statistical significance at the 10 percent level, differences in the effects of cohort racial composition on the dropout rates are sizable and potentially important. While an increase in the percent black or Hispanic does not have a significant effect on the likelihood of dropping out on average, a 1.0 percent increase in the percent minority increases the likelihood of dropping out among black students by 0.8 percentage points, a result which is substantively large and statistically significant. This result is consistent with findings from

mothers or percent disadvantaged minority in the school. Given the imprecision of our estimates, however, we typically could not reject the null hypotheses of no differences in effects across schools.

[28] Because our identification strategy successfully eliminates any correlation between percent with college educated mothers and percent black or Hispanic, entering the two variables jointly or singly provides very similar results. Using different specifications of the student covariates also did not substantially change the estimated effects.

[29] We ran placebo tests for each separate sample using the same procedure that we used for the pooled sample. The placebo regressions confirm that that the estimated effects of the actual cohort composition measures presented in Tables 8 and 9 are not biased by omitted school characteristics.

Guryan (2004) that desegregation efforts in the 1970's decreased drop rates among African Americans.[30] The estimated impact of exposure to minorities on Hispanic drop out rates is statistically insignificant, but the magnitude of the effect is similar to the substantial negative effect on black students while literally no effect is observed for whites or Asians.

The estimated effects of classmate characteristics on smoking and marijuana use also vary significantly across racial/ethnic groups. In general these differences highlight the fact that even when classmate composition measures do not show any effects on average, they can have significant effects on subgroups. For instance, although an increase in the percent black or Hispanic does not have significant effects on smoking on average (see Table 5), it does significantly decrease the likelihood of post-high school smoking among Asians. In general, however, while the effects of peers on smoking and marijuana use appear to be concentrated among Asians and Hispanics, the differences across groups in the effects of classmate composition on these outcomes are difficult to explain.

We also estimated the same set of regressions presented in Table 7 for four different groups defined by the mother's level of education. The results are presented in Table 8. The only statistically significant differences in effects between groups is that an increase in the percent of students with college educated mothers increases rates of idleness of students whose own mothers do not have a college education and decreases rates of idleness among those whose mothers have at least some college. Perhaps, when more of their classmates come from educated families, an accelerated pace of instruction or difficulties competing academically discourage

[30] Guryan reports that implementation of desegregation plans in the 1970s increased black exposure to whites an average of 15 points and estimates that these efforts decreased black dropout rates 2 to 3 percentage points. The estimate in Table 7 implies that a 15 percent increase in black exposure to whites would decrease black dropout rates several times more than Guryan estimates. The estimate in Table 7, however, is fairly imprecise. For instance, the hypothesis that a 15 percent increase in exposure to whites would reduce black dropout rates by 2 percent is well within the 95 percent confidence interval of the estimate in Table 7. Also, given that our estimates are based on relatively small changes in percent minority, predicting the impacts of a 15 percent increase in exposure requires extrapolation well beyond the variation available in our data.

children with less educated parents, undermining their cognitive development or their aspirations.

Finally, Table 9 presents the results from estimating our models separately for males and females. The upper portion of Table 9 indicates that effects of classmates are stronger for males than for females. The estimated effect of a change in the percent of classmates with college educated mothers on dropouts is three times as large for males as for females and the estimated effect on college attendance is 16 times as large for males as for females. Only the latter difference is statistically significant. Perhaps the most interesting result in Table 9 concerns the effect of percent black or Hispanic on post-high school idleness. The estimates from Table 5 indicate that on average changes in the percent of classmates who are black or Hispanic has no influence on rates of idleness. Here we see, however, that the effect of black or Hispanic classmates on the idleness rates for males is significantly different both from the effect on females and from zero. The exceptionally high and growing rates of idleness among black males is a growing policy concern (Edelman, Holzer, and Offner, 2006; Mincy, 2006). Thus, the finding that increasing the percentage of minority classmates can significantly affect the rates of idleness among males is potentially important.

Table 9 also indicates that increases in the percent of classmates with college educated mothers significantly increases post-high school smoking among females. This result suggests that while classmates may be more influential in determining educational attainment and employment outcomes for males than for females, females might be more susceptible to the influence of classmates than are males when it comes to other kinds of choices. Also, this finding suggests that the influence of classmates with college educated parents is not universally positive.

VIII. Conclusions

Our analyses use data from the Add-Health to estimate the effects of classmate characteristics on a range of student choices and outcomes. The unique structure of the Add-Health allows us to estimate these effects using comparisons across cohorts within schools, and to examine a wider range of outcomes than other studies that have used this identification strategy.

We find evidence that classmate characteristics do matter for potentially important individual outcomes. Most importantly, we find that increases in the percent of students with a college educated mother in one's cohort have several desirable affects on individual outcomes. Specifically, higher levels of parent education among one's classmates is associated with higher rates of college attendance and lower rates of dropping out of high school, smoking in high school and using marijuana after high school. Also, additional analyses suggest that increases in the average education level of classmates' parents results in substantial reductions in dropout rates among Hispanic students and more moderate reductions among whites. These are effects that would have been missed by studies focusing solely on test scores as outcome measures.

We do not find much evidence that the share of students from disadvantaged minority groups negatively affects student outcomes, on average. An increase in the percent of black or Hispanic classmates does increase the likelihood of using marijuana in high school, but not post-high school, and actually decreases rates of binge drinking post-high school. The percent of minority classmates might, however, have important effects for subgroups of students. Although the statistical evidence is weak, increases in the percent black or Hispanic appear associated with an increase dropout rates among black students. Also, an increase in the percent black or Hispanic classmates significantly increases rates of post-high school idleness among males.

The primary policy implication of these findings is that programs that decrease stratification across schools by parent education are likely to decrease disparities in educational attainment. Likewise, programs that increase stratification of this kind, including many types of school choice programs, are likely to exacerbate disparities in educational attainment. Our findings also suggest that programs that decrease the isolation of black students may help to decrease dropout rates among black students and decrease rates of idleness among black males.

A crucial caveat to all our findings is that our estimates only capture the effects of classmate characteristics that operate through the mechanisms we have referred to collectively as group dynamics. Any effects that operate through the school's ability to attract resources are missed. Also, any effects of student composition that operate schoolwide, such as, perhaps, effects on teacher expectations, are not captured by our estimates. Thus, our results might underestimate the total effect of classmate characteristics on individual outcomes. Nonetheless, our results provide evidence that classmate characteristics may well play a role in explaining disparities in outcomes across groups, and argue that policies that influence the grouping of students into schools may, therefore, have important costs and benefits.

REFERENCES

Altonji, Joseph G., Todd E. Elder, Christopher R. Tabor. 2005. "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools." *Journal of Political Economy*, 113: 151-184.

Angrist, Joshua and Kevin Lang. 2004. "Does School Integration Generate Peer Effects? Evidence from Boston's Metco Program." *American Economic Review*, 94(5): 1613-1634.

Bayer, Patrick, Stephen L. Ross, and Giorgio Topa. In press. "Place of Work, Place of Residence: Informal Hiring Networks and Labor Market Outcomes. *Journal of Political Economy.*

Betts, Julian, Kim Reuben, and K. Dannenberg. 2000. "Equal Resources, Equal Outcomes? The Distribution of School Resources and Student Achievement in California." Public Policy Institute of California.

Bifulco, Robert, Helen F. Ladd, and Stephen L. Ross. In press. “Public School Choice and Integration: Evidence from Durham, North Carolina.” *Social Science Research*.

Boozer, Michael and Stephen Cacciola. 2005. “Inside the ‘Black Box’ of Project Star: Estimation of Peer Effects using Experimental Data,” Economics Growth Center, Discussion Paper No. 832, Yale University.

Boozer, Michael A., Alan B. Krueger, and Shari Wolken. 1992. “Race and School Quality Since Brown v. Board of Education.” *Brookings Papers on Economic Activity. Microeconomics*, pp. 269-338.

Brunner, Eric, Jennifer Imazeki, and Stephen L Ross. In Press. “Universal Vouchers and Racial and Ethnic Segregation.” *Review of Economics and Statistics*.

Card, David and Jesse Rothstein. 2007. “Racial Segregation and the Black-White Test Score Gap.” *Journal of Public Economics* 91: 2158-2184.

Carrell, Scott E., Richard L. Fullerton and James E. West. 2008. “Does Your Cohort Matter? Measuring Peer Effects in College Achievement. NBER Working Paper #14032.

Cascio, Elizabeth and Diane Whitmore Schanzenbach. 2007. “First in the Class? Age and the Education Production Function.” *NBER Working Paper No. 13663*.

Clotfelter, Charles T. 2004. *After Brown: The Rise and Retreat of School Desegregation.* Princeton, NJ: Princeton University Press.

Clotfelter, Charles T., Helen F. Ladd, & Jacob L. Vigdor. 2005. “Who Teaches Whom? Race and the Distribution of Novice Teachers.” *Economics of Education Review* 24: 377-92.

Clotfelter, Charles T., Helen F. Ladd, & Jacob L. Vigdor. 2006. “Federal Oversight, Local Control, and the Specter of ‘Resegregation’ in Southern Schools.” *American Law and Economics Review* 8 (Summer): 1-43.

Cook, Thomas. 1984. “What Have Black Children Gained Academically from School Integration? Examination of the Meta-Analytic Evidence.” In *School Desegregation and Black Achievement*, ed. Thomas Cook, David Armor, N. Miller, W. Stephan. Herbert Walberg, & P. Wortman, pp. 6-42. Washington, DC: National Institute of Education.

Cullen, Julie Berry, Brian A. Jacob, and Steven Levitt. 2005. “The Impact of School Choice on Student Outcomes: An Analysis of the Chicago Public Schools.” *Journal of Public Economics*, 89 (June): 729-760.

Cutler, David M. and Edward L. Glaeser. 1997. “Are Ghettos Good or Bad?” *Quarterly Journal of Economics* 112: 827-872.

Duflo, Ester, Pascaline Dupas, and Michael Kremer. 2008. “Peer Effects and the Impact of Tracking: Evidence from a Randomized Experiment in Kenya.” Paper presented at the NBER Spring 2008 Education Program Meeting.

Edelman, Peter, Harry J. Holzer, and Paul Offner. 2006. *Reconnecting Disadvantaged Young Men.* Washington, D.C.: Urban Institute Press.

Evans, William N., Wallace E. Oates, & Robert M. Schwab. 1992. Measuring Peer Group Effects: A Study of Teenage Behavior. *Journal of Political Economy* 100 (October): 966-91.

Foster, Gigi. 2006. It's Not Your Peers, and It's Not Your Friends: Some Progress toward Understanding the Educational Peer Effect Mechanism. *Journal of Public Economics* 90: 1455-1975.

Friesen, Jane and Brian Krauth. 2008. “Enclaves, peer effects, and student learning outcomes in British Columbia.” Unpublished paper.

Friesen, Jane and Brian Krauth. 2007. “Sorting and Inequality in Canadian Schools.” *Journal of Public Economics* 91: 2185-2212.

Gould, Eric, Victor Lavy, and Daniele Passerman. 2004. “Does Immigration Affect the Longterm Educational Outcomes of Natives? Quasi-Experimental Evidence,” *NBER Working Paper No. 10844*.

Grinblatt, Mark, Matti Keloharju, and Seppo Ikaheimo 2008. “Interpersonal Effects in Consumption: Evidence from the Automobile Purchases on Neighbors.” *Review of Economics and Statistics* 90: 735-753.

Guryan, Jonathan. 2004. “Desegregation and Black Dropout Rates.” *American Economic Review* 94(4): 919-943.

Hanushek, Eric A., John F. Kain, and Steven G. Rivkin. 2002. “New Evidence about *Brown v. Board of Education*: The Complex Effects of Racial Composition on Achievement.” *National Bureau of Economic Research Working Paper No. 8741.*

Hastings, Justine S., Thomas J. Kane, and Douglas O. Staiger. 2006. “Parental preferences and school competition: Evidence from a public school choice program.” *National Bureau of Economic Research Working Paper No. 11805.*

Hoxby, Caroline M. 2000a. “The Effects of Class Size on Student Achievement: New Evidence from Population Variation.” *Quarterly Journal of Economics* 115(4): 1239-1285.

Hoxby, Caroline M. 2000b. “Peer Effects in the Classroom: Learning from Gender and Race Variation.” *National Bureau of Economic Research Working Paper No. 7867.*

Kremer, Michael and Dan M. Levy. 2008. Peer Effects and Alchohol Use Among College Students, *Journal of Economic Perspectives* 22(3): 189-206.

Lankford, Hamilton, Susanna Loeb, & Jim Wyckoff. 2002. “Teacher Sorting and the Plight of Urban Schools: A Descriptive Analysis.” *Educational Evaluation and Policy Analysis*, 24(Winter): 38-62.

Lavy, Victor and Analia Schlosser. 2007. “Mechanisms and Impacts of Gender Peer Effects at School.” *National Bureau of Economic Research Working Paper No. 13292.*

Lavy, Victor, M. Daniele Passerman, and Analia Schlosser. 2008. “Inside the Black-Box of Peer Ability Effects: Evidence from Variation in High and Low Achievers in the Classroom.” *NBER Working Paper No. 14415*.

Li, Han and Tao Li. In press. “The Gender Difference of Peer Influence in Higher Education.” *Economics of Education Review*.

Ludwig, Jens, Byron F. Lutz and David A. Weiner. 2007. “The Effects of School Desegregation on Crime.” Unpublished paper.

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

Mincy, Ronald B. 2006. *Black Males Left Behind*. Washington, D.C.: Urban Institute Press.

Reardon, Sean F., John T. Yun, and Michal Kurlaender. 2006. “Implications of Income-Based School Assignment Policies for Racial Segregation.” *Educational Evaluation and Policy Analysis* 28(Spring): 49-76.

Rivkin, Steven G. 2000. “School desegregation, academic attainment, and earnings.” *Journal of Human Resources* 35(2): 333-346.

Sacerdote, Bruce L. 2001. “Peer Effects with Random Assignment: Results from Dartmouth Roommates.” *Quarterly Journal of Economics* 116(2): 671-704.

Schofield, Janet Ward. 1995. “Review of Research of School Desegregation’s Impact on Elementary and Secondary School Students.” In *Handbook of Research on Multicultural Education*, ed. James A. Banks and Cherry A. McGee Banks, pp. 597-616. New York: Macmillan Publishing.

Siegfried, John J. and Michael A. Gleason. 2006. “Academic Roommate Peer Effects.” Working Paper.

Stinebrickner, Ralph and Todd R. Stinebrickner. 2006. “What Can Be Learned About Peer Effects Using College Roommates: Evidence from New Survey Data and Students from Disadvantaged Backgrounds.” *Journal of Public Economics* 90: 1435-1454.

Udry, 2003 J.R. Udry, The National Longitudinal Study of Adolescent Health (Add Health), Waves I and II, 1994–1996; Wave III, 2001–2002, Carolina Population Center, University of North Carolina, Chapel Hill, NC (2003) machine-readable data file and documentation.

Vigdor, Jacob and Jens Ludwig. 2007. Segregation and the Black-White Test Score Gap. *National Bureau of Economic Research Working Paper No. 12988.*

Vigdor, Jacob and Thomas Nechyba. In press. "Peer Effects in North Carolina Public Schools." In *Schools and the Equal Opportunity Problem*, eds. P.E. Peterson and L. Woessmann. Cambridge, MA: MIT Press.

Vigdor, Jacob and Thomas Nechyba. 2004. "Peer Effects in Elementary School: Learning from Apparent Random Assignment." Unpublished paper.

Wells, Amy Stuart. 1995. "Reexamining Social Science Research on School Desegregation: Long- versus Short-Term Effects." *Teachers College Record* 96 (Summer): 691-706.

Whitmore, Diane. 2005. "Resource and Peer Impact on Girls Academic Achievement: Evidence from a Natural Experiment." *American Economic Review Papers and Proceeding* 95 (2): 199-203.

Zabel, Jeffrey E. 2008. "The Impact of Peer Effects on Student Outcomes in New York City Public Schools." *Education Finance and Policy* 3(2): 197-249.

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

Table 1: Sample Descriptives

	N	Mean	Standard Deviation
Outcome Variables			
Drop Out of High School	9398	0.136	0.343
Attend College	9043	0.586	0.493
Post High School Test Score	9051	0.144	0.873
Idleness Post High School	9052	0.130	0.336
High School Smoking	9350	0.312	0.463
Post High School Smoking	9361	0.338	0.473
High School Marijuana Use	9244	0.170	0.376
Post High School Marijuana Use	9371	0.211	0.408
High School Binge Drinking	9372	0.345	0.475
Post High School Binge Drinking	9356	0.517	0.500
Cohort Variables			
Percent black or Hispanic in cohort	9398	30.4	29.4
Percent with college educated mother in cohort	9398	28.8	14.0
Baseline Controls			
Black	9398	0.163	0.370
Hispanic	9398	0.119	0.323
Asian	9398	0.048	0.207
Parent Education	9398	13.62	2.27
Grade 10 Indicator	9398	0.255	0.436
Grade 11 Indicator	9398	0.239	0.426
Grade 12 Indicator	9398	0.256	0.436
Extended Controls			
Male	9398	0.505	0.500
Age	9398	16.95	1.25
Parent Age	9398	42.59	5.82
Parent Native Born	9398	0.872	0.302
Parent Years in US	9398	35.70	13.09
Parent Information Missing	9398	0.335	0.472
PVT Score	8953	0.180	0.926
Additional Family Controls			
Log Family Income	9398	0.358	0.209
Single Parent	9398	0.264	0.405
Live with Both Parents	9398	0.573	0.456
Number Older Siblings	9385	0.834	1.179
Talk about School with Parents	9398	0.638	0.464
Parent Involvement	9398	0.310	0.426
Parent Alcoholic	9398	0.149	0.328

Table 2: Student Outcomes and Cohort Composition, by Race and Mother's Educational Attainment

	White Students	Black Students	Hispanic Students	Asian Students
Drop Out of High School	0.121 (0.326)	0.159 (0.366)	0.206 (0.404)	0.097 (0.296)
Attend College	0.607 (0.488)	0.519 (0.500)	0.489 (0.500)	0.764 (0.425)
Post High School Test Score	0.344 (0.682)	-0.399 (1.024)	-0.184 (1.103)	-0.028 (1.003)
Idleness Post High School	0.116 (0.320)	0.194 (0.395)	0.140 (0.347)	0.100 (0.300)
High School Smoking	0.370 (0.483)	0.155 (0.362)	0.237 (0.425)	0.202 (0.402)
Post High School Smoking	0.390 (0.488)	0.204 (0.403)	0.260 (0.439)	0.254 (0.435)
High School Marijuana Use	0.173 (0.379)	0.160 (0.367)	0.180 (0.384)	0.132 (0.338)
Post High School Marijuana Use	0.229 (0.420)	0.181 (0.385)	0.184 (0.387)	0.131 (0.338)
High School Binge Drinking	0.389 (0.487)	0.200 (0.400)	0.350 (0.477)	0.211 (0.408)
Post High School Binge Drinking	0.599 (0.490)	0.253 (0.435)	0.460 (0.498)	0.390 (0.488)
Percent black or Hispanic in cohort	17.0 (17.9)	62.6 (29.1)	57.9 (29.7)	41.7 (23.0)
Percent with college educated mother in cohort	28.5 (13.8)	29.6 (14.0)	26.4 (13.1)	37.2 (14.9)
Sample Size	4920	1921	1701	856
	High School Drop-Out	High School Graduate	Some College	College Graduate
Drop Out of High School	0.250 (0.433)	0.143 (0.350)	0.127 (0.264)	0.057 (0.232)
Attend College	0.356 (0.479)	0.518 (0.500)	0.620 (0.485)	0.808 (0.394)
Post High School Test Score	-0.297 (1.026)	0.111 (0.793)	0.240 (0.835)	0.398 (0.792)
Idleness Post High School	0.185 (0.388)	0.139 (0.346)	0.126 (0.332)	0.085 (0.278)
High School Smoking	0.290 (0.454)	0.341 (0.474)	0.317 (0.465)	0.278 (0.448)
Post High School Smoking	0.295 (0.456)	0.376 (0.484)	0.347 (0.476)	0.303 (0.460)
High School Marijuana Use	0.160 (0.367)	0.172 (0.378)	0.179 (0.384)	0.165 (0.371)
Post High School Marijuana Use	0.147 (0.354)	0.214 (0.410)	0.229 (0.420)	0.235 (0.424)
High School Binge Drinking	0.333 (0.471)	0.343 (0.475)	0.373 (0.484)	0.330 (0.470)
Post High School Binge Drinking	0.377 (0.485)	0.528 (0.499)	0.528 (0.499)	0.583 (0.493)
Percent black or Hispanic in cohort	44.0 (32.9)	27.2 (28.0)	28.1 (27.7)	28.2 (28.1)
Percent with college educated mother in cohort	23.4 (10.5)	26.1 (11.8)	29.1 (12.3)	36.3 (17.2)
Sample Size	1657	3083	2189	2469

Means and standard deviations in parentheses.

Table 3: Variation in cohort composition measures after removing school fixed effects and trends.

	Raw cohort variables				
Full Sample	N	Mean	Std Dev	Min	Max
Percent mothers with college	9384	0.302	0.139	0.000	0.877
Percent black or Hispanic	9398	0.377	0.312	0.000	1.000
	Residuals after removing school fixed effects and trends				
Full Sample		Mean	Std Dev	Min	Max
Percent mothers with college	9384	0.000	0.026	-0.159	0.143
Percent black or Hispanic	9398	0.000	0.030	-0.203	0.122

Table 4: Balancing tests for cohort composition measures

Dependent Variable	% black or Hispanic	% with college educated mother	F-statistic
Male	-0.431* (0.231)	-0.151 (0.200)	1.980
Age (in years)	-0.120 (0.312)	-0.076 (0.273)	0.113
Parent's age (in years)	-3.918 (4.012)	3.908* (2.044)	2.025
Parent born in the U.S.	-0.174* (0.101)	-0.074 (0.085)	1.856
Missing parent information	0.260 (0.257)	-0.105 (0.262)	0.675
PVT test score	9.923 (6.652)	0.668 (5.497)	1.123
Log of family income	0.095 (0.094)	0.139 (0.096)	1.655
Single parent family	0.396 (0.270)	0.007 (0.219)	1.117
Live w/both biological parents	-0.218 (0.247)	0.442* (0.259)	2.396*
Number of older siblings	-0.251 (0.638)	-0.167 (0.398)	0.193
Parent alcoholism reported	0.039 (0.120)	-0.113 (0.178)	0.322

The figures in each row are coefficients from regressions that include in addition to the cohort composition measures controls for cohort fixed effects, school fixed effects, school trends, the student's race, and the student's mother's years of education. All variables are measured using Wave 1 of the Add Health. Figures in parentheses are standard errors robust to clustering at school level. The F-statistics is for the joint effect of percent black or Hispanic and percent with college educated mothers. * designates significantly different from zero at 0.10 or an F-Statistics greater than 2.303.

Table 5: Estimated impacts of cohort composition on student outcomes

	Baseline Controls	Baseline + Extended Controls	Baseline + Extended Controls+ Additional Family	Baseline Controls	Baseline + Extended Controls	Baseline + Extended Controls+ Additional Family
	Drop Out of High School			*Attend College*		
% College Educated Mother	-0.327**	-0.312***	-0.299***	0.515**	0.504**	0.439**
	(0.131)	(0.106)	(0.112)	(0.223)	(0.210)	(0.189)
% black + % Hispanic	0.080	0.104	0.064	0.034	0.027	0.060
	(0.188)	(0.169)	(0.166)	(0.296)	(0.273)	(0.267)
	Post High School Test Score			*Idleness Post High School*		
% College Educated Mother	0.239	0.210	0.232	0.039	0.020	0.042
	(0.322)	(0.249)	(0.257)	(0.130)	(0.136)	(0.137)
% black + % Hispanic	0.591*	0.342	0.327	0.085	0.123	0.118
	(0.334)	(0.229)	(0.222)	(0.160)	(0.147)	(0.150)
	High School Smoking			*Post High School Smoking*		
% College Educated Mother	-0.452*	-0.410	-0.399*	0.126	0.215	0.248
	(0.259)	(0.256)	(0.235)	(0.190)	(0.196)	(0.197)
% black + % Hispanic	0.128	0.159	0.136	0.311	0.295	0.269
	(0.289)	(0.279)	(0.272)	(0.250)	(0.237)	(0.237)
	High School Marijuana Use			*Post High School Marijuana Use*		
% College Educated Mother	-0.272	-0.240	-0.239	-0.474***	-0.435**	-0.422**
	(0.169)	(0.179)	(0.170)	(0.166)	(0.172)	(0.169)
% black + % Hispanic	0.395**	0.428**	0.412**	0.267	0.254	0.218
	(0.166)	(0.173)	(0.177)	(0.195)	(0.205)	(0.204)
	High School Binge Drinking			*Post High School Binge Drinking*		
% College Educated Mother	-0.118	-0.184	-0.181	-0.205	-0.096	-0.116
	(0.258)	(0.261)	(0.248)	(0.213)	(0.208)	(0.209)
% black + % Hispanic	0.060	-0.069	-0.085	-0.452	-0.578**	-0.599**
	(0.265)	(0.258)	(0.251)	(0.276)	(0.291)	(0.294)

All regressions include controls for cohort fixed effects, school fixed effects, and school trends as well as the individual student covariates related to the cohort variables. Figures in parentheses are standard errors robust to clustering at the school level. * designates significantly different from zero at 0.10, ** significantly different than zero at 0.05 level, and *** significantly different from zero at 0.01 level.

Table 6: Results of Placebo Regressions

	Baseline Controls	Baseline + Extended Controls	Baseline + Extended Controls+ Additional Family	Baseline Controls	Baseline + Extended Controls	Baseline + Extended Controls+ Additional Family
	Drop Out of High School			*Attend College*		
% College Educated Mother	-0.043	-0.009	0.003	-0.153	-0.123	-0.144
	(0.103)	(0.103)	(0.098)	(0.164)	(0.168)	(0.146)
% black + % Hispanic	-0.068	-0.099	-0.091	-0.191	-0.110	-0.102
	(0.085)	(0.088)	(0.087)	(0.141)	(0.128)	(0.129)
	Post High School Test Score			*Idleness Post High School*		
% College Educated Mother	0.247	0.163	0.156	0.037	0.098	0.098
	(0.336)	(0.213)	(0.213)	(0.070)	(0.072)	(0.073)
% black + % Hispanic	-0.030	-0.126	-0.085	0.032	0.011	-0.002
	(0.221)	(0.192)	(0.134)	(0.080)	(0.078)	(0.078)
	High School Smoking			*Post High School Smoking*		
% College Educated Mother	-0.120	-0.133	-0.107	0.050	0.018	0.036
	(0.172)	(0.186)	(0.172)	(0.153)	(0.151)	(0.160)
% black + % Hispanic	0.022	0.050	0.047	-0.040	-0.004	-0.009
	(0.163)	(0.170)	(0.167)	(0.130)	(0.143)	(0.139)
	High School Marijuana Use			*Post High School Marijuana Use*		
% College Educated Mother	-0.036	-0.053	-0.043	-0.152	-0.135	-0.105
	(0.111)	(0.120)	(0.114)	(0.159)	(0.169)	(0.166)
% black + % Hispanic	-0.062	-0.107	-0.105	-0.139	-0.121	-0.117
	(0.103)	(0.101)	(0.103)	(0.099)	(0.105)	(0.107)
	High School Binge Drinking			*Post High School Binge Drinking*		
% College Educated Mother	-0.103	-0.012	0.023	0.213	0.138	0.142
	(0.199)	(0.213)	(0.192)	(0.143)	(0.163)	(0.163)
% black + % Hispanic	-0.088	-0.034	-0.025	0.158	0.267*	0.285**
	(0.158)	(0.155)	(0.158)	(0.153)	(0.137)	(0.138)

All regressions include controls for cohort fixed effects, school fixed effects, and school trends as well as the individual student covariates related to the cohort variables. Figures in parentheses are standard errors robust to clustering at the school level. * designates significantly different from zero at 0.10, ** significantly different than zero at 0.05 level, and *** significantly different from zero at 0.01 level.

Table 7: Estimated impacts of cohort composition on student outcomes, by race

	Black Students	White Students	Hispanic Students	Asian Students	F-Statistic
			Drop Out of High School		
% College Educated Mother	0.283	-0.297**	-1.648***	-0.222	3.642**
	(0.365)	(0.128)	(0.466)	(0.465)	
% black + % Hispanic	0.823**	-0.166	0.932	0.100	2.064
	(0.393)	(0.186)	(0.905)	(0.292)	
			Attend College		
% College Educated Mother	0.311	0.536**	1.113	-0.197	0.495
	(0.577)	(0.215)	(0.823)	(0.782)	
% black + % Hispanic	0.076	0.185	-1.029	0.132	0.465
	(0.497)	(0.266)	(0.994)	(0.743)	
			Post High School Test Score		
% College Educated Mother	-0.134	0.406	-2.603*	-0.675	1.843
	(0.570)	(0.265)	(1.318)	(1.845)	
% black + % Hispanic	-0.220	0.181	0.956	-0.399	0.295
	(0.583)	(0.389)	(1.222)	(1.645)	
			Idleness Post High School		
% College Educated Mother	0.308	0.023	0.072	-0.524	0.341
	(0.466)	(0.161)	(0.456)	(0.683)	
% black + % Hispanic	0.921	0.135	0.479	0.981**	1.405
	(0.580)	(0.163)	(0.608)	(0.484)	
			High School Smoking		
% College Educated Mother	-0.025	-0.573**	-0.200	0.241	0.656
	(0.355)	(0.285)	(0.813)	(0.817)	
% black + % Hispanic	0.070	0.310	0.314	-1.313	1.148
	(0.636)	(0.304)	(0.512)	(0.838)	
			Post High School Smoking		
% College Educated Mother	0.344	0.192	0.809	1.756***	1.999
	(0.552)	(0.261)	(0.746)	(0.599)	
% black + % Hispanic	1.014	0.246	-0.097	-1.437**	2.674**
	(0.781)	(0.321)	(0.858)	(0.600)	
			High School Marijuana Use		
% College Educated Mother	-0.043	-0.163	-1.576**	-1.537***	3.499**
	(0.372)	(0.207)	(0.704)	(0.485)	
% black + % Hispanic	0.469	0.271	1.332*	0.507	0.686
	(0.477)	(0.169)	(0.752)	(0.582)	

			Post High School Marijuana Use		
% College Educated Mother	-1.198***	-0.364*	-0.655	-0.211	1.362
	(0.406)	(0.211)	(0.651)	(0.352)	
% black + % Hispanic	-0.510	0.242	1.369***	0.951*	2.599*
	(0.571)	(0.260)	(0.491)	(0.547)	
			High School Binge Drinking		
% College Educated Mother	-0.125	-0.190	-0.862	1.207**	2.021
	(0.444)	(0.304)	(0.772)	(0.578)	
% black + % Hispanic	-0.650	0.019	0.075	-1.531**	1.411
	(0.663)	(0.331)	(0.669)	(0.743)	
			Post High School Binge Drinking		
% College Educated Mother	-0.703	-0.067	-0.505	0.746	0.864
	(0.519)	(0.258)	(0.699)	(0.854)	
% black + % Hispanic	-0.862	-0.577	-0.547	-1.269	0.210
	(0.728)	(0.390)	(0.791)	(0.855)	

Each figure reported is a coefficient from a separate regression. All regressions include controls for cohort fixed effects, school fixed effects, school trends, and the extended set of individual student and additional family covariates. Figures in parentheses are standard errors robust to clustering at the school level. F-statistic is based on a Wald test with three degrees of freedom for the joint hypotheses that coefficients obtained using each sample are equal. * designates significantly different from zero at 0.10 or an F statistic greater than 2.084, ** significantly different than zero at 0.05 level or an F of 2.605, and *** significantly different from zero at 0.01 level or an F of 3.782.

Table 8: Estimated impacts of cohort composition on student outcomes, by mother's education level

	Mother's Education				
	High School Drop-Out	High School Graduate	Some College	College Graduate	F-Statistic
			Drop Out of High School		
% College Educated Mother	-0.381	-0.534**	-0.144	-0.278*	0.359
	(0.698)	(0.244)	(0.363)	(0.156)	
% black + % Hispanic	0.877	-0.141	0.004	0.085	0.925
	(0.559)	(0.261)	(0.299)	(0.217)	
			Attend College		
% College Educated Mother	0.920*	0.323	0.413	0.395	0.353
	(0.514)	(0.287)	(0.431)	(0.346)	
% black + % Hispanic	-0.183	0.173	-0.683	0.494	1.215
	(0.554)	(0.387)	(0.430)	(0.520)	
			Post High School Test Score		
% College Educated Mother	-1.487	0.399	0.183	0.360	1.199
	(0.945)	(0.425)	(0.518)	(0.486)	
% black + % Hispanic	0.412	0.231	1.083*	-0.382	1.471
	(0.889)	(0.586)	(0.609)	(0.380)	
			Idleness Post High School		
% College Educated Mother	0.904	0.443**	-0.732**	-0.244	3.441**
	(0.703)	(0.209)	(0.361)	(0.342)	
% black + % Hispanic	0.326	0.362	0.058	-0.144	0.312
	(0.733)	(0.318)	(0.241)	(0.548)	
			High School Smoking		
% College Educated Mother	-0.377	-0.538*	-0.344	-0.355	0.061
	(0.709)	(0.308)	(0.538)	(0.411)	
% black + % Hispanic	0.430	0.231	0.147	0.136	0.060
	(0.643)	(0.507)	(0.348)	(0.464)	
			Post High School Smoking		
% College Educated Mother	-0.672	0.128	0.126	0.525	0.685
	(0.689)	(0.509)	(0.410)	(0.472)	
% black + % Hispanic	0.421	0.602	0.472	-0.212	0.487
	(0.616)	(0.450)	(0.336)	(0.555)	

	High School Marijuana Use				
% College Educated Mother	-0.330	-0.180	-0.565	-0.052	0.409
	(0.459)	(0.372)	(0.367)	(0.304)	
% black + % Hispanic	1.018**	0.292	0.214	0.317	0.698
	(0.474)	(0.340)	(0.382)	(0.348)	
	Post High School Marijuana Use				
% College Educated Mother	-0.443	-0.837***	-0.371	-0.049	0.831
	(0.508)	(0.297)	(0.506)	(0.422)	
% black + % Hispanic	0.542	-0.106	0.129	0.562	0.338
	(0.624)	(0.417)	(0.400)	(0.840)	
	High School Binge Drinking				
% College Educated Mother	-0.466	-0.619*	0.437	0.237	1.130
	(0.567)	(0.338)	(0.660)	(0.493)	
% black + % Hispanic	0.416	0.019	-0.032	-0.897	0.774
	(0.469)	(0.609)	(0.299)	(0.727)	
	Post High School Binge Drinking				
% College Educated Mother	-0.157	-0.329	0.164	0.589	0.898
	(0.634)	(0.330)	(0.432)	(0.482)	
% black + % Hispanic	-1.368**	-0.580	-0.008	-1.105*	1.220
	(0.669)	(0.403)	(0.448)	(0.655)	

Each figure reported is a coefficient from a separate regression. All regressions include controls for cohort fixed effects, school fixed effects, school trends, and the extended set of individual student and additional family covariates. Figures in parentheses are standard errors robust to clustering at the school level. F-statistic is based on a Wald test with three degrees of freedom for the joint hypotheses that coefficients obtained using each sample are equal. * designates significantly different from zero at 0.10 or an F statistic greater than 2.084, ** significantly different than zero at 0.05 level or an F of 2.605, and *** significantly different from zero at 0.01 level or an F of 3.782.

Table 9: Estimated impacts of cohort composition on student outcomes, by gender

	Males	Females	t-statistic	Males	Females	t-statistic
	Drop Out of High School			*Attend College*		
% College Educated Mother	-0.487***	-0.166	1.42	0.811***	0.051	1.67*
	(0.174)	(0.144)		(0.265)	(0.370)	
% black + % Hispanic	0.118	-0.056	0.56	-0.282	0.396	1.38
	(0.252)	(0.185)		(0.404)	(0.282)	
	Post High School Test Score			*Idleness Post High School*		
% College Educated Mother	0.118	0.200	0.14	-0.202	0.301	1.93*
	(0.455)	(0.337)		(0.173)	(0.195)	
% black + % Hispanic	0.276	0.550	0.45	0.505**	-0.165	2.09**
	(0.355)	(0.491)		(0.224)	(0.230)	
	High School Smoking			*Post High School Smoking*		
% College Educated Mother	-0.392	-0.480	0.20	-0.245	0.553**	1.79*
	(0.295)	(0.324)		(0.343)	(0.252)	
% black + % Hispanic	0.007	0.107	0.18	0.182	0.194	0.03
	(0.383)	(0.389)		(0.374)	(0.284)	
	High School Marijuana Use			*Post High School Marijuana Use*		
% College Educated Mother	-0.249	-0.303	0.17	-0.629**	-0.227	1.25
	(0.250)	(0.202)		(0.248)	(0.205)	
% black + % Hispanic	0.505*	0.301	0.55	0.264	0.089	0.44
	(0.289)	(0.228)		(0.264)	(0.292)	
	High School Binge Drinking			*Post High School Binge Drinking*		
% College Educated Mother	0.064	-0.387	1.27	-0.261	-0.000	0.66
	(0.259)	(0.243)		(0.292)	(0.268)	
% black + % Hispanic	0.266	0.052	0.48	-0.690**	-0.788*	0.18
	(0.332)	(0.220)		(0.291)	(0.462)	

All regressions include controls for cohort fixed effects, school fixed effects, and school trends as well as the individual student covariates related to the cohort variables. Figures in parentheses are standard errors robust to clustering at the school level. t-statistics are for the difference between parameter estimates for males and females. * designates significantly different from zero at 0.10, ** significantly different than zero at 0.05 level, and *** significantly different from zero at 0.01 level.