I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14140

Peers' Race in Adolescence and Voting Behavior

Maria Polipciuc
Frank Cörvers
Raymond Montizaan

FEBRUARY 2021

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14140

Peers' Race in Adolescence and Voting Behavior

Maria Polipciuc
Maastricht University

Frank Cörvers
Maastricht University

Raymond Montizaan
Maastricht University and IZA

FEBRUARY 2021

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Peers' Race in Adolescence and Voting Behavior*

Using a representative longitudinal survey of U.S. teenagers, we investigate how peer racial composition in high school affects individual turnout of young adults. We exploit across-cohort, within-school differences in peer racial composition. One within-school standard deviation increase in the racial diversity index leads to a 2.2 percent increase in the probability of being registered to vote seven years later and to a 2.6 percent higher probability of voting six years later. These effects are likely due to positive interracial contact when socialization has long-lasting effects: higher racial diversity in school is linked to more interracial friendships in school and later on.

JEL Classification: D72, I24, J15

Keywords: voting behavior, school-cohort racial diversity, peers

Corresponding author:
Raymond Montizaan
Research Centre for Education and the Labour Market (ROA)
School of Business and Economics
Maastricht University
Tongersestraat 53
6200 MD Maastricht
PO box 616
The Netherlands

E-mail: r.montizaan@maastrichtuniversity.nl

* We thank seminar audiences at Maastricht University and Norwegian School of Economics, and participants at the Joint PhD Workshop in Economics (University of Bergen, Norwegian School of Economics) in October 2018, at the 2019 ERMAS conference at Babes-Bolyai University and at the 2020 EALE conference. This article benefited from helpful comments from and discussions with Joshua Angrist, Patrick Dylan Bennett, Bet Caeyers, Marion Collewet, Rita Ginja, Fanny Landaud, Linda Østergren and Veronica Petrencu. This research uses data from the National Longitudinal Study of Adolescent to Adult Health (Add Health), a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain the Add Health data files is available from the Add Health website (http://www.cpc.unc.edu/addhealth). No direct support was received from grant P01-HD31921 for this analysis.

Economists have become increasingly aware of the importance of well-functioning institutions, such as the judicial and political systems, in determining economic performance (Costa and Kahn, 2003). Several studies have established that social capital is a key determinant of good institutions and of economic growth (Knack and Keefer, 1997; Alesina and La Ferrara, 2000; Costa and Kahn, 2003; Guiso et al., 2004). A large literature finds that community heterogeneity leads to lower levels of social capital and to fewer interactions among community members, which is potentially problematic in an increasingly diverse world. Community diversity—be it ethnic, racial, or religious—negatively affects participation in social activities (Alesina and La Ferrara, 2000), trust (Costa and Kahn, 2003; Dinesen and Sønderskov, 2015), the quality and quantity of publicly provided goods (Alesina et al., 1999; Vigdor, 2004; Putnam, 2007), the willingness to redistribute income (Luttmer, 2001; Dahlberg et al., 2012), donations (Andreoni et al., 2016), and economic growth (Easterly and Levine, 1997; Montalvo and Reynal-Querol, 2005).

On a more positive note, recent empirical evidence shows that beliefs about members of other racial or ethnic groups evolve with contact and integration (Boisjoly et al., 2006; Burns et al., 2019; Finseraas et al., 2019; Lowe, 2019; Rao, 2019; Schindler and Westcott, 2020; Steinmayr, 2020) and that intergroup trust can be built even after ethnic conflict (Mousa, 2020). Exposure to diversity at an early age can create a common culture that leads to less redistributional conflict among social groups later in life and may reduce transaction costs between individuals from different social groups, by diminishing the social distance between them or by changing their beliefs about members of other groups (Gradstein and Justman, 2000, 2002). Moreover, having more racially diverse friend groups can increase support for affirmative action (Boisjoly et al., 2006).

The current literature which reports mostly negative effects of diversity generally focuses on contemporaneous measures of racial diversity and civic engagement. For this reason, it cannot establish whether there exist positive or negative long-term effects of diversity on civic engagement, as a result of evolving beliefs or intergroup trust due to earlier contact and integration. It therefore remains an empirical question whether exposure to racial diversity early on has a negative impact on the civic engagement of individuals later in life, even after they have moved to other communities with different racial compositions.[1]

Moreover, the negative effects of diversity on civic engagement are found in rather large communities, making it hard to observe evolving believes and intergroup contacts. Examples include counties, cities, or at the very

[1]Contemporaneous measurement of diversity and civic engagement also increases the likelihood that differences in racial diversity and civic engagement between communities are driven by other factors occurring at the time of measurement.

minimum, census blocks (Dinesen and Sønderskov, 2015; Algan et al., 2016; Cancela and Geys, 2016). These findings might differ in smaller units of observation, such as the school cohort level we study, because it is likely that intergroup contact works differently at this lower level of aggregation (Algan et al., 2016).

In this paper, we focus on voting, which is a type of civic engagement whose relationship to community racial composition has been investigated in a number of studies (Costa and Kahn, 2003; Oberholzer-Gee and Waldfogel, 2005; Cancela and Geys, 2016; Shertzer, 2016; Martinez i Coma and Nai, 2017; Bellettini et al., 2020).[2] We address the gaps in the literature by examining the causal long-run impact of the racial composition of one's peers in high school, which is arguably a community in which intergroup contacts are likely to evolve, on voting behavior later in life, and explore channels through which peer racial diversity could play a role. For this, we use the National Longitudinal Study of Adolescent to Adult Health (Add Health). More specifically, we focus on individual voter participation and political orientation. We use variation in student racial composition across cohorts within high schools to identify the effects of peers' race and use an exogenous effects model (Manski, 1993), under the assumption that students potentially sort into schools based on the school's average racial composition, but not based on the composition of the student's specific cohort. This method has been validated and used by several studies that have shown that variations in peers' race, gender, ability, language spoken at home, and exposure to family violence influence individual test scores and post-secondary outcomes.[3] The Add Health data set is particularly suited for measuring the long-run effects of racial diversity in high school on voting behavior since it follows multiple student cohorts from the same school into adulthood.

To our knowledge, this is the first study to use this across-cohort, within-school strategy to examine the long-term effects of the racial composition of one's peers in high school on individual voter participation and political orientation later in life. Previous studies on racial diversity focus on the short-run effects and find a negative relationship with turnout (Costa and Kahn, 2003; Martinez i Coma and Nai, 2017; Bellettini et al., 2020).[4] This paper is also closely related to an emerging literature on the effects of school desegregation policies on voting registration, turnout and political

[2]Voting is considered a measure of civic engagement because the individual bears the private costs (time spent voting, time spent informing oneself about the elections, etc.) of an action that has public benefits, at least for their group.

[3]For example, Hoxby (2000a,b); Angrist and Lang (2004); Gould et al. (2009); Hanushek et al. (2009); Carrell and Hoekstra (2010); Bifulco et al. (2011); Friesen and Krauth (2011); Lavy and Schlosser (2011); Lavy et al. (2012); Black et al. (2013).

[4]A recent review study of aggregate-level research on turnout levels by Cancela and Geys (2016), however, shows more mixed effects of ethnic diversity.

partisanship. The studies of Billings et al. (2020), Kaplan et al. (2019) and Bergman (2020) find mixed effects of the examined policies on one or more of these outcomes. Bergman (2020) analyzes the short- and long-run risks and benefits of a randomized racial desegregation program in elementary schools for minority students. He finds that being offered a transfer to a low-minority share, higher-resource school increased the likelihood to vote, although exclusively for males. Billings et al. (2020), who examine the end of race-based busing in Charlotte-Mecklenburg, North Carolina schools in 2002–2003, and Kaplan et al. (2019), who focus on a policy that bused students in Jefferson County, Kentucky, between 1975–1985, on the other hand, find no effects of these policies on voting registration and turnout. However, both studies report that these policies increased the likelihood to be registered as a Democrat.[5] Both Billings et al. (2020) and Kaplan et al. (2019) use data from Southern states, and the latter study also focuses on a policy implemented decades ago. This raises the question to what extent their results are generalizable to other states and to exposure to racial diversity in more recent years. A major benefit of our study is that, unlike data from the natural experiments from desegregation policies, which are local, it uses data which are sampled to be representative of the U.S. middle and high school population in 1994–1995. This increases the external validity of our results. Our study also differs in that we measure on racial diversity using a Herfindahl-Hirschman index of racial diversity, while the studies on desegregation policies focus on minority shares or on randomized access to schools with a lower share of one's own race. This racial diversity index is the most common way to operationalize diversity in the literature studying the effects of diversity on proxies of civic engagement. As a result, using this index allows us to compare our results with those of previous studies in this literature.

Studying the impact of peers' racial composition in high school on voting behavior later in life is of particular relevance, since the respondents are adolescents when exposed to racial diversity. Their preferences, beliefs and personality traits are still highly malleable (Borghans et al., 2008), so exposure to members of other racial or ethnic groups has the potential to change them considerably. Adolescence is the phase in which fairness and efficiency considerations, which have been proven to be good predictors of political affiliation and decisions (Fisman et al., 2017; Kerschbamer and Müller, 2020), seem to crystallize (Almås et al., 2010). Given the greater mal-

[5]Billings et al. (2020) find that a 10-percentage point increase in the share of minorities in a student's assigned school decreased their likelihood of registering as a Republican 15 years later by 8.8 percent, with the effect being driven by white students. Kaplan et al. (2019) find that white males who had been assigned to be bused in Jefferson County, Kentucky, between 1975–1985 were significantly more likely to be registered as Democrats forty years later.

leability of personality traits and preferences during adolescence, schools play a crucial role not only in transmitting knowledge, but also as a socializing force fostering civic engagement (Gradstein and Justman, 2000, 2002). Many studies find that, unlike socialization during adulthood, early socialization has a durable effect on political attitudes and voting behavior (Jennings and Markus, 1977, 1984; Malmendier and Nagel, 2011; Madestam and Yanagizawa-Drott, 2012; Giuliano and Spilimbergo, 2013; Kim and Lee, 2014; Akbulut-Yuksel et al., 2019; Algan et al., 2019). For example, using historical data from post-WWII Germany, Akbulut-Yuksel et al. (2019) show that the expulsion of Jewish professionals had long-lasting detrimental effects on the political interest and participation of Germans who were in their impressionable years (ages six to 23) during the Nazi regime, but not on adults. The authors further demonstrate that these adverse effects can be explained by the social changes brought about by the expulsions, which led to lower adult socioeconomic status (SES) and lower civic skills for individuals in their impressionable years during that time. Madestam and Yanagizawa-Drott (2012) find similar effects of attending Fourth of July celebrations on political engagement: while the impact of attendance on voter participation and political orientation was long-lasting for young individuals (ages four to 18), there were no significant long-term effects on the political behavior of adults. Hence, should there be long-run positive effects of racial diversity on voting behavior due to changes in beliefs about members of other racial or ethnic groups from contact and integration, we can expect these effects to be strongest for individuals who are relatively young when exposed to racial diversity.

Our most important finding is that exposure to a more racially diverse cohort in high school increases the probability of voting as a young adult. Greater diversity (as measured by a Herfindahl-Hirschman index of racial diversity) results in higher probabilities of being registered to vote seven years later and of having voted in the presidential elections in 2000, six years later. We measure racial diversity at the cohort level within schools; an increase in the cohort racial diversity index by one within-school standard deviation leads to an increase of 1.6 percentage points in the probability of being registered to vote seven years later (2.2 percent of the mean) and to an increase of 1.1 percentage points in the probability of having voted in the presidential elections six years later (2.6 percent of the mean). Racial diversity in high school, however, does not affect whether and with which party individuals identify as young adults.[6]

[6]Our results might seem contradictory at first with those of the studies evaluating desegregation policies. However, much of the differences in results reflects the choice of the measure of interest. While our paper uses a racial diversity index, the studies of Billings et al. (2020) and Kaplan et al. (2019) analyze the share of minorities, or randomized access to a school with a lower share of one's own race than the original

We further examine several potential channels through which racial diversity in high school could affect voting behavior positively in the long run. Our estimations show that positive intergroup experiences exist due to racial diversity: respondents in diverse cohorts have more interracial friendships both during high school and 14 years later, in Wave 4. We also find marginal evidence that a more racially diverse cohort in school has an impact on personality traits in Wave 4 (higher extraversion and conscientiousness). Both channels are indicative of an early socialization mechanism driving the change in the probability to be registered to vote and turnout. Our results thus indicate that a broader, long-run assessment of the impact of racial diversity on voting behavior is necessary. The long-term benefits due to more racially diverse friend groups and smaller social distance between these groups could outweigh their short-term costs.

This paper is organized as follows. Section 1 describes the data and presents descriptive statistics. Section 2 discusses our empirical strategy. Section 3 presents our econometric results. The last section concludes.

1 Data and descriptive statistics

1.1 Data

We use data from Add Health, a school-based longitudinal study of a nationally representative sample of adolescents in the United States. The data were collected in several ways from adolescents, their fellow students, school administrators, parents, siblings, friends, and partners. In Wave 1, an in-school survey was conducted to provide data on the school context and friendship networks. Thereafter, five in-home surveys were held in Waves 1 to 5. Register databases with information on respondents' neighborhoods and communities have been merged with the Add Health data set. In our analyses, we use information from the in-school survey, from the in-home surveys in Waves 1, 3 and 4, as well as from the merged register data sets.

The in-school survey in Wave 1 took place in 1994–1995. Around 90,000 individuals in grades 7 through 12 participated. The in-school survey gathered basic information, such as respondents' gender, race, and parents' level of education. This data enable us to construct the main explanatory vari-

school. The results of our robustness analyses using racial shares are in line with Billings et al. (2020) and Kaplan et al. (2019). That is, the share of blacks in one's cohort has no impact on registration or voting—but it does have a positive effect on the probability to identify as a Democrat. Our paper thus complements and is consistent with the existing findings in this emerging literature.

able: the racial diversity index of one's cohort.[7,8]

Add Health uses a clustered sampling design, in which the schools were sampled first and then the pupils in these schools were selected to form the core in-home sample (approximately 17 females and 17 males were selected randomly from each grade for most schools; in 16 schools, all the enrolled students were selected). This core sample was enhanced with a variety of oversamples based on race, twin status, disability status, and other categories. The enhanced sample (approximately 20,000 individuals in Wave 1) was subsequently followed longitudinally through in-depth interviews at home. The median cohort of respondents in the in-home sample consists of 213 students (minimum 12, maximum 691). Of these, the median number of respondents in the in-home survey in Wave 1 from a school cohort is 31 students (minimum seven, maximum 441).[9]

Our dependent variables reflect political attitudes and behavior, which were assessed in the in-home survey in Wave 3. Wave 3 took place in 2001–2002 and tracked approximately 15,000 of the Wave 1 in-home respondents. The great majority of the respondents in Wave 3 were aged 18 to 26. We also use a couple of variables from Wave 4 for robustness checks and for assessing potential mechanisms.

The estimation sample only includes schools with respondents in more than one cohort in both Waves 1 and 3 (a requirement of the estimation strategy; see Section 2.1) and non-missing values for a set of control variables.[10] We focus on cohorts with at least 10 respondents in the in-school survey in Wave 1. The estimation sample thus obtained consists of 12,568 individuals.

1.2 Variable definitions

1.2.1 Dependent variables

The five dependent variables in our analyses are dummy variables from the in-home survey in Wave 3. These measure voting behavior and political partisanship, as follows: 1) the first dummy measures whether the respondents were registered to vote in Wave 3, 2) the second measures

[7]Students could declare being a member of more than one race. Since for the analysis, we had to assign them to one race only, we followed Bifulco et al. (2011) and gave precedence to the answers in the following order: black, Hispanic, Asian, white, and other. For instance, if a respondent claimed to be both white and Hispanic, the respondent would be considered Hispanic in our analysis.

[8]A total of 15.75 percent of the respondents in the in-school survey selected more than one race. Figure B.1 in Appendix B shows how our results on voting are influenced by various amounts of measurement error in peers' race.

[9]See Bifulco et al. (2011) for a more detailed explanation of the sampling process.

[10]All the control variables and their definitions are listed in Table A.1 in Appendix A.

whether they voted in the 2000 presidential election (Bush versus Gore), 3) the third measures whether the respondents identified with a political party, 4) and, finally, the fourth and fifth dummies measure whether the respondents identified with the Democratic or the Republican Party, respectively. It is important to note that identification as a Democrat is not conditional on identification with a party, such that a value of one means the respondent identifies with the Democratic Party and zero means the respondent either does not identify with any party at all or identifies with another party. The same holds for the dependent variable for identification with the Republican Party.

1.2.2 Racial diversity index

Most studies examining the effects of diversity on economic outcomes operationalize diversity by an index and/or the share of the minority (or minorities) of interest.[11] In our main analysis, we use a racial diversity index.[12] To calculate the index, we construct the racial shares for each cohort in a school from the race students declare in the in-school questionnaire of Wave 1. The racial diversity index is computed as one minus the sum of the squared shares of each race possibly present in a cohort (c) within a school (s).[13] We consider five mutually exclusive categories (i) for race, ordered here by group size: white, black, Hispanic, Asian, and other:

$$RacialDiversity_{cs} = 1 - \sum_{i} s_{ics}^2$$

The racial diversity index reflects the probability that two individuals chosen at random from a cohort belong to two different racial groups. This index is the most widely used in the literature on diversity (Alesina and La Ferrara, 2005). Since we consider five possible races, this index ranges between zero (only one race is present in the school-cohort combination) and 0.8 (all five races are present, in equal proportions). The index can increase in two ways: as the racial shares become more equal, and when the number of racial groups in the cohort increases.

1.3 Descriptive statistics

Table 1 describes the characteristics of the estimation sample in column (1). Columns (2) and (3) present the means and standard deviations for

[11] For example, Alesina and La Ferrara (2000); Costa and Kahn (2003); Bifulco et al. (2011); Algan et al. (2016); Merlino et al. (2019).

[12] The results of robustness checks in Tables B.3 and B.4 in Appendix B use alternative measures of diversity.

[13] This index is known in the literature as the fractionalization index. The index is equal to one minus the Herfindahl-Hirschman index (or one minus the concentration index).

the white and minority (or nonwhite) subsamples, respectively.

It is immediately apparent that the minority respondents are in high school cohorts with higher proportions of black students and greater racial diversity. They also come from lower-income families and are more likely to live in urban areas, to be slightly older in Wave 1, and to have more pupils in their class. In Wave 3 (seven years after Wave 1), they earn less than their white peers, on average, and have a slightly lower education.

In terms of voting behavior, minorities are also slightly less likely to be to be registered to vote, as well as to have voted in the 2000 presidential elections. In the full estimation sample, approximately 43 percent declared having voted in the presidential elections in 2000 and about 72 percent were registered to vote when the data collection of Wave 3 took place (August 2001 to April 2002). We do observe a substantial difference between whites and minorities with respect to party identification. Minorities are twice as likely as whites to identify with the Democratic Party (27.3 percent versus 13.6 percent).

Table 1 also presents the within-cohort racial SES Gini index, which has similar means for white and minority respondents.[14]

Another variable showed in Table 1 is a cohort dummy indicating whether the pupils are grouped by ability in English. Roughly half of the respondents are in cohorts grouped by ability, with minorities being slightly more likely to be in such cohorts.[15]

[14]The racial SES Gini index is a Gini coefficient of the inequality in mothers' education by race in the cohort. We use the formula of Alesina et al. (2016), where, for each race, we consider two educational categories of mothers: with and without a college degree. The Gini coefficient for a cohort with n racial groups with a share y_i of college-educated mothers in group i, where i to n are indexed in nondecreasing order ($y_i \leq y_{i+1}$), is computed as

$$Gini = \frac{1}{n}[n + 1 - 2\frac{\sum_{i=1}^{n}(n+1-i)y_i}{\sum_{i=1}^{n} y_i}]$$

The coefficient is equal to zero if there is only one racial group or if the shares of college-educated mothers are equal for all the races represented in the school cohort. In a cohort with m races, the coefficient is highest when the share of college-educated mothers is zero for $m-1$ races, and one for the remaining race. In this case, the Gini coefficient is $(m-1)/m$. In this paper, we consider five races, so the Gini coefficient has an upper bound of 0.8.

[15]The information on whether a certain cohort in a school is grouped by ability is provided in Wave 1 by the school principal. For each cohort in the school, the principal answered either yes or no to the following question: "For English or language arts, does your school group classes according to ability or achievement?". In the United States, language arts refers to the area of the curriculum in which students are taught the range of skills needed to become proficient in using a language (Moreau, 2011).

2 Identification and estimation strategy

2.1 Empirical implementation

To examine the causal long-run impact of the racial composition of peers in high school on voting behavior and political partisanship, we use an analytical strategy that eliminates the bias created by families or students systematically sorting into schools, which can be observed in Table 1. We exploit idiosyncratic variation in cohort composition between adjacent grades within schools. We thus assume that, conditional on attending a certain school, the cohort composition a pupil faces is as good as random. This method was pioneered by Hoxby (2000b) and is widely used in education economics in pre-university peer studies, where random assignment is rarely feasible.[16]

To implement this strategy, we estimate a reduced-form equation using a linear probability model in which the outcome of an individual student is a linear function of the student's own observable characteristics, the mean characteristics of all the students in the same cohort and school, cohort fixed effects, school fixed effects, and school-specific linear trends. We include the racial SES Gini index and the ability grouping dummy among the controls in our analyses. We thus account for the possibility that the differences in voting behavior are not caused by racial diversity itself, but rather by exposure to the socioeconomic inequality associated with the racial diversity within school cohorts.[17] As for ability grouping, if this practice clusters students by race, the chances of interracial contact would be reduced in grouped cohorts. Since positive intergroup contact is a potential mechanism through which racial diversity positively affects voting behavior, not controlling for ability grouping could lead to an underestimation of the impact of racial diversity on voting behavior.

The reduced-form equation looks as follows:

$$\begin{aligned} PolOutcome_{ics} = \alpha + \beta_0 RacialDiversity_{cs} + \beta_1 X_{cs} + \beta_2 X_i + \\ + \delta_c + \phi_s + \lambda_s C + \epsilon_{ics} \end{aligned} \tag{1}$$

for individual i, cohort c, and school s. The variable $PolOutcome_{ics}$ is one of five dummy variables from Wave 3, indicating individuals who, respectively, are currently registered to vote, voted in the presidential elections in 2000, identify with a party, and identify with the Democratic or with the

[16] See, for instance, Carrell and Hoekstra (2010); Bifulco et al. (2011); Lavy and Schlosser (2011); Lavy et al. (2012); Black et al. (2013).

[17] We have also used alternative racial inequality Gini indices as control variables in additional robustness analyses, such as a racial SES Gini index for fathers' education, one for mothers' employment status, and one for fathers' employment status. None of these indices changes the magnitude of the coefficient of the racial diversity index for voting or being registered to vote significantly. Results are available on request.

Republican Party. Students' race is one of five mutually exclusive categories reported in the in-school questionnaire: white, black, Hispanic, Asian, and other. The variable $RacialDiversity_{cs}$ is the racial diversity index; X_{cs} includes the racial SES Gini index and the indicator for ability grouping; X_i is a set of individual characteristics; δ_c is a cohort fixed effect common to all schools; ϕ_s is the school fixed effect, which ensures that we compare cohorts within a school; C is an indicator variable for the student's cohort in Wave 1, which is allowed to vary by school; and $\lambda_s C$ allows for the possibility of school-specific linear trends in racial composition. As Bifulco et al. (2011) point out, not controlling for trends is problematic if, for instance, parents decide to enroll their children in a school based on the observed trend in the school's racial composition. The term ϵ_{ics} is a random error term. Standard errors are clustered at the school level.[18]

2.2 Identifying variation

Two conditions must be met for the resulting estimates to have a causal interpretation: (i) there should be sufficient variation in cohort composition within schools and (ii) the source of variation should be plausibly random. With regard to the latter issue, it is important to note that many of the mechanisms through which racial composition can influence outcomes are constant across cohorts in the same school. For example, a school's ability to acquire resources and parents' decisions to place their children in a school are likely influenced by the composition of the school as a whole, instead of that of a particular cohort. Our identification strategy relies on within-school variation in cohort composition so our estimates will not capture any effect that the student composition of the school as a whole has on individual outcomes. Below we present tests of conditions (i) and (ii).

Table 2 addresses condition (i): the top panel shows the standard deviation of the main explanatory variables in the estimation sample; the middle and bottom panels show the variation of the residuals obtained by regressing the respective explanatory variable on school and cohort dummies (middle panel) plus school linear trends (bottom panel). Most of the variation in racial composition is due to differences between schools. Although the within-school variation represents only a small fraction of the total variation in the racial diversity index (15 percent after removing school and cohort fixed effects and linear school trends), this is sufficient to reasonably estimate the effects of small changes in cohort composition.

[18]We also conducted robustness analyses in which we omitted the racial SES Gini index and the ability grouping indicator. Our results are robust to this exercise; the coefficients of the racial diversity index remain very similar for both being registered to vote (p-value< 0.001) and having voted (p-value< 0.1). The results are available on request.

Previous studies that use these data and method come to a similar conclusion (Bifulco et al., 2011; Merlino et al., 2019). The magnitudes reported in the middle and bottom panels are also in line with those in Bifulco et al. (2011).

To address condition (ii), we first run balancing tests, the results of which are shown in Table 3. These tests check whether deviations from school-specific fixed effects or trends in cohort composition as measured by the racial diversity index are associated with deviations in student background characteristics (except for one's own race, which is addressed by separate tests, described in the next paragraph). We regress predetermined Wave 1 background characteristics on the racial diversity index and on dummies for own race and own grade (column (1)); we then add school fixed effects (column (2)) and school linear trends (column (3)). None of the coefficients remain significant after the addition of school fixed effects and school linear trends.

Second, if variation in cohort composition is as good as random, a pupil's race should not be correlated with that of his or her peers. To test whether this is indeed the case, we need to consider all the students in cohorts from which there is a student in our estimation sample. However, one cannot simply regress an individual's own race on peers' race to test this: since an individual is always excluded from their peer group, this mechanically creates a negative correlation between the two variables, even in the presence of random variation.[19] Guryan et al. (2009) propose correcting for this bias by additionally controlling for all potential peers, that is, by controlling for the racial composition of the school as a whole. We regress each race dummy on the racial diversity index of peers (excluding oneself) in the cohort and in the school, with grade and school fixed effects and school linear trends. Appendix Table C.1 presents the results: in each column, the dependent variable is a dummy for being of a certain race. Caeyers and Fafchamps (2020) propose a different approach: regressing a transformed dummy for being of a certain race on the respective racial share of others in the cohort. The race dummy is transformed by subtracting the exclusion bias which creates the artificial negative correlation. Appendix Table C.2 shows the results for all races considered, also including grade and school fixed effects and school linear trends. From these tables, we conclude that respondents' race is not systematically correlated with that of their peers as measured by the racial diversity index at a significance level of $\alpha = 0.05$.

Third, we conduct Monte Carlo simulations to further test whether the within-school variation observed in the racial diversity index is consistent with a random process. We use two methods to flag those schools from

[19]This is true of the correlation with the share of peers of the same race. We focus on racial diversity as measured by the racial diversity index, so we adjust the tests accordingly.

the sample where this is not the case. For the first method, we compute the shares of each race present in the school. Then, for each student in the school, we randomly generate a counterfactual race using a multinomial distribution function with the probabilities of being of a certain race equal to the true racial probabilities at the school level. We repeat the process 1,000 times and compute a confidence interval for the racial diversity index, calculated using the generated counterfactual racial shares in each cohort within each school. We then flag those schools where the true average racial diversity index at the school level does not fall within the 95 percent confidence interval. This method has been used previously by Lavy et al. (2012).

The second method is very similar to the first, but it simulates assignment to a certain cohort in a school using the multinomial distribution of grades within the school. We compute the racial shares in the counterfactual grades and again flag schools where the resulting racial diversity index does not fall within the 95 percent confidence interval. This method has been used by Bifulco et al. (2011).

With our choice of the simulation seed, the first method flags 13.1 percent of the schools in the estimation sample (16 out of 122 schools, comprising 9.1 percent of the students in the main estimation sample), while the second one flags 5.7 percent (seven out of 122 schools, where 4.2 percent of the students in the main sample were enrolled).[20]

A fourth test that we perform to check whether the observed within-school variation in the racial diversity index is consistent with a random process is based on Feld and Zölitz (2017). The authors show that, if the variation one exploits is nonrandom, measurement error in the explanatory variable can create an upward bias in its coefficient. We follow Merlino et al. (2019) and introduce various amounts of measurement error in races and examine the resulting pattern of the coefficient of the racial diversity index in two regressions: that of the probability of being registered to vote and that of the probability of having voted in 2000. We introduce measurement error in race in the following way: in the in-school data set, we run a multinomial logit regression of individuals' own race (which can take on five values) on school and grade fixed effects and school linear trends, with standard errors clustered at the school level. We calculate the predicted probabilities of being of each race. We then generate a new variable that takes the value of each race with a probability equal to the predicted probability of being of that race. Then, we generate a new race variable for each level of measurement error considered, denoted *me*: 0 percent, 5

[20]As a robustness check, we rerun the main regressions on the restricted samples, which contain only respondents from the unflagged schools. Appendix Tables B.1 and B.2 present the results, which remain largely unchanged compared to results in Table 4. This result indicates that the variation we exploit is quasi-random.

percent, 10 percent, 25 percent, 50 percent, 75 percent, 90 percent and 100 percent. This variable takes the value of the true race in $100 - me$ of the cases, and the value of the new variable in me of the cases. We repeat the process 1,000 times for each error level in the measurement of race. We then construct racial diversity indices and racial SES Gini indices for all levels of measurement error in race. Appendix Figure B.1 plots the coefficients of the racial diversity index with measurement error from the regressions of being registered to vote and having voted on our most complete specification, where both indices contain measurement error. Once again, we find evidence supporting the assumption that the variation we exploit is as good as random.

A fifth and final test is a permutation test, as discussed in Guryan et al. (2009, p. 48) and Caeyers and Fafchamps (2020, p. 11–12). It works as follows: we first regress each of the five race dummies on the share of others in the cohort belonging to the same race, on cohort and school fixed effects (first specification) plus school linear trends (second specification). Because of exclusion bias, even if the allocation to cohorts in a school is as good as random after controlling for the above-mentioned fixed effects, with or without school linear trends, the coefficient for the share of others in the cohort is expected to be negative and significant. To check the assumption of random assignment we shuffle cohorts within schools, and thus create counterfactual cohort assignment. We then regress the five race dummies on the same set of independent variables, but replacing the observed share of others of the same race in the cohort with its counterfactual. We repeat this process 1,000 times and store the coefficients of the counterfactual cohort composition. For each of the five races, we then check what share of these coefficients is either above the absolute value of the true coefficient or below minus its absolute value. This share is the p-value of the test of random peer assignment for that specific race. For neither of the five races can we reject the null hypothesis of random peer assignment, either with or without including school linear trends.[21]

We conclude that the tests in this section do not reject our hypothesis that, once we control for grade and school fixed effects and school linear trends, residual deviations in the cohort racial diversity index are as good as random.

[21]Results available on request.

3 Results

3.1 Voting behavior

Table 4 presents evidence on the impact of racial diversity in one's high school cohort on voting behavior as a young adult. The dependent variable in columns (1) to (3) is being currently registered to vote. In columns (4) to (6), the dependent variable is having voted in the presidential elections in 2000. In columns (1) and (4), we control for individual characteristics measured in Wave 1. In columns (2) and (5), we additionally control for family characteristics in Wave 1. Columns (3) and (6) show the results when we also control for Wave 1 residential neighborhood characteristics. Our estimates indicate that greater racial diversity in one's cohort in high school as measured by the racial diversity index leads to a higher probability of being registered to vote and a higher probability of having voted in the 2000 presidential elections.

The estimate in the most comprehensive specification in column (3) of Table 4 shows that an increase in the racial diversity index by one within-school standard deviation leads to an increase in the probability of being registered to vote of approximately 1.6 percentage points. This represents an increase of 2.2 percent relative to the unconditional probability of being currently registered to vote (71.9 percent). If we compare the point estimates of the racial diversity index across columns (1) to (3), we further observe that the size of the coefficient is robust to the inclusion of individual, family, and neighborhood characteristics in Wave 1.

Columns (4) to (6) of Table 4 show the effects of the racial diversity index on actual voting behavior in the presidential elections in 2000. When using estimates in column (6), we find an increase in the racial diversity index by one within-school standard deviation increases the voting probability by 1.1 percentage points. This represents an increase of 2.6 percent relative to the unconditional probability of voting (42.9 percent). One's own race is also significantly related to the probabilities of being registered to vote and of having voted in the 2000 elections. Ceteris paribus, black respondents are more likely to be registered and to have voted than white respondents, while both probabilities are lower for Asians.[22]

Our results therefore suggest a positive causal long-run impact of the racial composition of one's peers in school on voting behavior later in life, even after controlling for own race. The effects we find are sizable compared to previous estimates of the (short-term) determinants of political participation. For example, de Rooij et al. (2009) find that door-to-door canvass-

[22]The coefficients in Table 4 increase slightly and remain significant at 1% level for being registered to vote, and at 5% for having voted, if we do not include either individual characteristics, family characteristics or neighborhood characteristics.

ing increases turnout by 7.1 percentage points. Furthermore, DellaVigna et al. (2016) and Rogers et al. (2016) find that the turnout rate of people who expect to be asked whether they have voted is 0.3 and 0.2 percentage points, respectively, higher than for those who do not expect to be asked.

3.2 Political partisanship

In addition to information on voting behavior, Add Health also surveyed respondents about their political attitudes in Wave 3. We have information on whether and with which party people identify. By examining the effects of early age diversity on political partisanship, we can check whether increased voting is due to political preferences becoming more alike or, on the contrary, more polarized.[23] It is, however, not clear a priori how these political views are affected by greater racial diversity in the school cohort, as measured by a racial diversity index. In both parties, numerous factions cover a wide range of the political spectrum. On some policy issues, there could therefore be overlap between certain factions from the two parties. Furthermore, there are substantial differences within the national and local divisions of these parties. Whether—and how—exposure to racial diversity measured as an index influences political partisanship is thus a purely empirical question.

Table 5 reports the results. This table is constructed similarly to Table 4, by adding the control variables mentioned in the table caption. We find that the racial diversity index in the school cohort is not significantly related to whether people indicate that they identify with a party, or to whether they identify with the Democratic or Republican Party. We do, however, find that one's own race is significantly related to political partisanship. Black individuals are more likely than white individuals (the baseline category) to identify with a party, and also more likely to identify with the Democratic Party. Asians identify significantly less with a specific party. It is also noteworthy that greater SES racial inequality in high school is linked to a higher probability to identify with the Democratic Party.

3.3 Underlying mechanisms

3.3.1 Friendships

Evidence in previous sections suggests that peer racial composition matters for voting behavior. The question is: what can explain this long-run positive impact of the racial diversity of one's peers in high school on voting behavior? The literature on racial diversity and voting behavior suggests

[23]Using data from the General Social Survey, Oberholzer-Gee and Waldfogel (2001) document that the political preferences of black and white individuals in the United States differ substantially, with the former being more liberal.

two likely mechanisms: positive intergroup contact and negative intergroup contact.

On the one hand, according to contact theory, personal contact with outgroup members can reduce prejudice and increase trust under the following conditions (Allport, 1954; Pettigrew and Tropp, 2006; Pettigrew et al., 2011): equal status, shared common goals, a cooperative setting, some form of authority, and friendship potential (summarized by Finseraas et al., 2019). If these conditions are met, exposure to people of different races and social backgrounds in childhood can lead to a more racially diverse friend group later in life. Evidence of such a relation is found by Merlino et al. (2019) for white students, also using the Add Health data set. They show that for whites, being in a cohort with more black individuals of one's own gender has a positive impact on the probability to have a black romantic partner later in life. The authors suggest this is likely due to the higher likelihood of meeting potential partners of other races via friends of one's own gender. These own gender friends are more likely to be black as the share of blacks in the cohort increases. Burns et al. (2019) find that assignment to a mixed-race room lead white students in South Africa to report a higher share of interracial friendships, have lower prejudice towards blacks, support affirmative action, and be more prosocial in an incentivized experimental game (with a partner of an unspecified race).[24] If positive intergroup experiences predominate, we expect that early interracial contact may reduce the incidence or the magnitude of negative utility from interracial contact later in life. Should this be the case, this increases one's benefits from civic participation later in life (for instance, by voting) in a society that is racially heterogeneous (Alesina and La Ferrara, 2000).

On the other hand, if interracial contact leads to negative experiences, it can stimulate more political activity due to a negative perception of other races. The possibility of ethnic diversity leading to greater prejudice and less trust toward outgroup members is in line with constrict theory (Putnam, 2007).

To determine whether racial diversity in high school leads to more positive or more negative intergroup contact on average, we examine how diversity is linked to two aspects: interracial friendship nominations in Wave 1, and interracial friendships in Wave 4. We use the same econometric specification as when examining voting behavior.

In case the greater political participation in Wave 3 is due to more collaborative and positive contact between members of different racial groups, we can expect a more racially diverse school cohort to be associated with

[24] Also for whites, Boisjoly et al. (2006) find that being assigned a black roommate in the first year of university leads to greater openness to minorities, and greater support for affirmative action. However, the study finds no effects on the share of friends of another race, or on the how frequently white students socialize with blacks.

more interracial friendships in Wave 1, while negative experiences with racial diversity in the school cohort could lead to greater racial endogamy. Table 6 looks at interracial friendships in Wave 1 in two different ways: in column (1), the dependent variable is the share of friends of other races, as computed by Add Health.[25] Column (2) contains a dummy we constructed from friendship nominations, reflecting whether the respondent has at least one minority friend if the respondent is white and at least one white friend if the respondent is a minority. This variable thus captures interracial friendships which reflect intergroup contact between whites and minorities, rather than among several minorities. In both columns, the coefficients of the racial diversity index are positive and statistically significant. The point estimates imply that an increase of one within-school standard deviation in the index is linked to an increase of 1.3 percentage points in the share of friends of other races in Wave 1 and to an increase of 3.6 percentage points in the probability of a white-minority friendship in Wave 1. This finding is a first indication that negative experiences with racial diversity in school are unlikely to be the dominant explanation for the results in Section 3.1.

Column (3) in Table 6 further shows that racial diversity has long-lasting but marginally positive effects on interracial friendships. A survey question in Wave 4 (in 2008, 13–14 years after Wave 1) asked respondents what the races of their close friends were, with the following potential answers: all the same race as myself (1), almost all the same race as myself (2), mostly the same race as myself (3), about half the same race as myself (4), mostly other races than my own (5), almost all other races than my own (6), and all other races than my own (7). Based on these answers, we constructed a dummy variable indicating whether the respondents have close friends of other races (answers (2) to (7) are coded one, and answer (1) is coded zero). Column (3) shows the results of a linear regression that uses this dummy as a dependent variable. We find a marginally positive coefficient, indicating that racial diversity positively impacts the likelihood of people having at least one friend of another race more than a decade later. The point estimate implies that an increase in the racial diversity index of one within-school standard deviation increases this probability by 0.9 percentage points.[26]

[25]This share is calculated based on all in-school friendship nominations and is not restricted to a student's own school cohort.

[26]Coefficients of interest and significance levels do not change if we use weights for longitudinal analyses with Waves 1, 3 and 4.

3.3.2 Personality

Socialization in early life influences one's personality, preferences, group identities and beliefs about politics (Madestam and Yanagizawa-Drott, 2012; Akbulut-Yuksel et al., 2019; Algan et al., 2019; Kaplan et al., 2019; Bergman, 2020; Billings et al., 2020). If racial diversity in one's school cohort shapes early life socialization, this constitutes another potential channel through which diversity can impact voting behavior and political preferences of adults.

The Add Health survey does not collect data on beliefs or preferences, but it includes a module measuring personality traits using a 20-item mini-IPIP scale for the first time in Wave 4. In Table 7, we look at whether personality traits in Wave 4 are influenced by a student's cohort composition. The dependent variables in the two panels in Table 7 aggregate the score for each personality trait in two different ways: in the top panel, the average score for all four questions for one trait is used, while in the bottom one, the score for the first component from a principal component analysis for each trait is used. Higher racial diversity leads to marginally more extraversion and conscientiousness, according to both panels. In our data, both variables are significantly correlated to the probability to be registered to vote (the coefficient is 0.034, with a standard error of 0.008 for extraversion; it is 0.022, with a standard error of 0.009 for conscientiousness). Extraversion is significantly correlated to the probability to have voted (the coefficient is 0.025, with a standard error of 0.008) (results are available on request). These results are consistent with an earlier study of Cooper et al. (2012), who find that these two traits are positively related to the probability to be registered to vote.[27]

Results in Table 7 provide suggestive evidence that racial diversity in school could have an impact on turnout and political preferences through its shaping of the socialization environment in school. These results are only tentative, and more suitable data is needed to parse out mechanisms more precisely.

3.4 Heterogeneous effects

Until now, we have seen that racial diversity in a school cohort has long-term positive effects on voting behavior and that racial diversity stimulates friendships with individuals of other races. These findings suggest that, consistent with contact theory, more collaborative and positive contact between individuals of different races could be a driving force behind the increased probability to vote through its impact on early socialization.

[27]The political psychology literature linking personality traits to voter turnout is however best described as having mixed results (Mondak et al., 2010; Mondak, 2010; Gerber et al., 2011a,b).

However, previous research has also shown that the characteristics of individuals engaging in social interactions, such as their race or SES, can influence how they experience the racial composition of their environment (Marschall and Stolle, 2004). In this section, we therefore investigate whether the results on voting behavior differ significantly by racial background or by Wave 1 family income level. We split the sample by minority status (white or minority) and annual family income in Wave 1 (above or below the median of $40,000 in our sample).

Table 8 presents the results. We find no statistically significant differences between groups, although this is potentially due to smaller sample sizes. However, the fact that there is no clear pattern in the coefficients seems to suggest a positive impact of racial diversity on the voting behavior of both whites and minorities, as well as of individuals from families with an income above or below the median.

4 Robustness checks

4.1 Robustness to attrition and weighting

A potential issue for interpreting our results is that the relationship we find between being registered to vote or having voted in the previous presidential elections and cohort racial diversity might be due to differential attrition in Wave 3. We use two different strategies to check this.[28]

First, we estimate the equation (1) using as dependent variables the probabilities to be a respondent in Waves 2, 3 and 4 (as Wave 1 was not affected by attrition). For all three dependent variables, in all three specifications (including individual characteristics, then gradually adding family characteristics, and neighborhood characteristics) the coefficients of the racial diversity index are positive and insignificant.[29]

Second, in our estimations we use Wave 3 longitudinal weights. This places more weight on those in categories from which there are more attriters. Our results may be due to some observations being weighted more heavily. To check this, we drop from the estimation sample those respondents with the highest 10% Wave 3 longitudinal weights. We re-estimate the regressions in Table 4 using this smaller sample. Results are presented in Appendix Table B.5. The coefficients of the racial diversity index decrease slightly, but they remain significant at the 5% level for being registered to vote and at 10% level for having voted in 2000. In this sample, the racial SES Gini has a negative and significant coefficient in the regression with having voted as a dependent variable.[30]

[28]The procedures in this Section have been inspired by those in Merlino et al. (2019).

[29]Results available on request.

[30]We also trimmed the sample manually using Lee bounds (Lee, 2009). This procedure

From these tests, we conclude that attrition does not affect our estimates significantly.

4.2 Robustness to different specifications of racial diversity

In Appendix Tables B.3 and B.4, we test the robustness of our results to other specifications of racial diversity. We add the racial shares, use only the racial shares, or use the racial shares plus the racial shares squared. The coefficient of the racial diversity index remains positive and significant for the probability to be registered to vote (column (1) in Table B.3), and remains positive but insignificant for the probability to have voted in 2000 (column (4) in Table B.3).

For voting behavior (Table B.3), while all shares have positive coefficients, it is only the share of Asians that is significant, regardless of whether we include the squared shares or not. For political partisanship (Table B.4), a higher share of black students in the cohort is positively related to respondents identifying with a political party, particularly with the Democratic Party. In all specifications, more interracial SES inquality (as measured by the racial SES Gini) significantly increases the probability that one identifies as a Democrat. Following Madestam and Yanagizawa-Drott (2012), we interpret this as suggesting that both a higher share of black individuals in

provides bounds for the treatment effect, under the assumption that the effect of the treatment on attrition is monotonic—in our case, that those in more racially diverse cohorts as measured by a higher racial diversity index are more likely to attrit by Wave 3 than those in less racially diverse cohorts. We define a cohort as treated if its racial diversity index is above the school's mean racial diversity index.

In practice, the method drops observations (from either the Wave 1 in-home sample or from the Wave 3 sample) in a way that equalizes the share of treated in the in-home Wave 1 sample and in the Wave 3 sample. It drops one of two types of observations: either those that contribute the most to the correlation between the treated dummy and the dependent variable (Registered to vote, Voted in 2000)—which gives a lower bound for the treatment effect, or the observations that contribute the least to the correlation between the treated dummy and the dependent variable—which gives an upper bound for the treatment effect. Ideally, the confidence interval for the treatment effect does not include zero. We manually select the observations to be dropped. For being registered to vote, the confidence interval for the treatment effect as defined by Imbens and Manski (2004) does include zero. For having voted in 2000, the confidence interval does not include zero.

This procedure rests on the assumption that the effect of the treatment on attrition is monotonic. Bounds are estimated without covariates. One can tighten the bounds by estimating them separately for each category of a predetermined covariate, such as race. With our data, this leads to an error which potentially indicates that the relationship between being treated and attrition differs by race—which contradicts the monotonicity assumption (see section 2.3 in Tauchmann, 2014, for details). For this reason, we do not stress the results of these tests, which are available on request.

one's cohort and greater racial inequality in one's cohort shift preferences to the political left, without increasing political polarization (since there is no effect on the probability of identifying with the Republican Party—results confirming this are available upon request).

These results are in line with Kaplan et al. (2019); Bergman (2020); Billings et al. (2020). These papers find that a higher share of minorities in one's school (or a proxy for it, namely assignment to busing or the option to transfer to a more racially diverse school than one's initial school) has no effect on turnout and registration. All three papers find that the specific desegregation policy they analyze makes their respondents more likely to identify as Democrats: Kaplan et al. (2019) finds this effect for white males, Bergman (2020)—for minority students, and Billings et al. (2020)—only for white students.

4.3 Relating our results to the literature

We find positive long-term effects of racial diversity in adolescence on voter registration and turnout. Is this surprising, given the negative or inconsistent short-term effects of racial diversity found in other studies?

In order to answer this question, we compare the effects of the racial diversity index on other short-term and long-term behaviors which have been investigated by previous studies. Appendix Tables B.6 and B.7 present the relationship between the racial diversity index and Wave 1 and Wave 3 behaviors which have been found to be sensitive to peer influence. The regressions use our most comprehensive specification and show that even if there is some evidence of a negative short-term correlation between the racial diversity index and behavior in Wave 1, this mostly vanishes by Wave 3. This is similar to results in Bifulco et al. (2011) and Bifulco et al. (2014), who find no long-term effects of the share of minorities (the cumulative shares of blacks and Hispanics) in one's cohort on post-secondary outcomes. We also observe that racial diversity has a marginally significant negative impact on binge drinking and a marginally positive impact on test scores and on smoking later in life.[31]

[31]We also checked whether our estimates might be driven by social desirability bias. If, for instance, individuals in more racially diverse cohorts are more agreeable or more likely to think it is important to fit in with one's group, then our estimates of voter registration and turnout might be biased upwards (since the dependent variables are self-reported). The cohort racial diversity has no impact on agreeableness in Wave 4 (see column 2 in both panels in Table 7). We also estimate equation (1) using as dependent variable a binary variable from Wave 3, reflecting the perceived importance to fit in with one's group. We coded responses of 'agree' and 'strongly agree' as '1', and the rest as '0'. The coefficient of the racial diversity index in the most complete specification (including individual characteristics, family characteristics, neighborhood characteristics, plus grade fixed effects, school fixed effects and school linear trends) is

These results indicate that our main findings are in line with previous findings. The most plausible reason for any apparent difference is therefore the different time frame for the effects (long- versus short-run), the different geographic aggregation of the data (narrowly versus broadly defined peer groups), or using a different measure of racial diversity.

4.4 How likely is it to find any long-term effect of the racial diversity index?

In subsection 4.3 we checked whether there are effects of the racial diversity index on several long-term outcomes. This raises the issue that the more tests we run, the higher the chance that some results turn out significant. We thus test the composite null of no effects of racial diversity in the long term on the following variables: our main variables of interest (is registered to vote, has voted in 2000) plus the seven variables from subsection 4.3 (is a high school dropout, has a college degree, the score in the Picture Vocabulary Test administered by Add Health, is idle (does not work and does not attend school), smokes, smokes marijuana, engages in binge drinking).

We use a resampling procedure, as discussed in Bifulco et al. (2011, section I.C). The authors combine the resampling approach by Westfall and Young (1993, p. 214–215) with a strategy by Agresti (2002, p. 97–98), to calculate the likelihood that a certain pattern of p-values might arise should the composite null hypothesis that there are no effects of racial cohort composition be true. The probability of a false positive is the sum of the probabilities of all possible outcomes that occur with a probability lower than or equal to the probability in the observed data. The share of p-values corresponding to outcomes more extreme than those in the observed data is calculated using this resampling approach.

We estimate a linear-in-means model for the nine Wave 3 outcome variables. We regress these variables on the battery of characteristics used in our main specification, excluding the racial diversity index and the racial SES Gini index, but including school fixed effects, cohort fixed effects, either with or without school linear trends. We run 10,000 simulations, following the procedure in Bifulco et al. (2011). For the racial diversity index, values more extreme than the observed one under the null hypothesis are quite unlikely (p-value = 0.002 without trends, p-value = 0.003 with trends). This is not the case for the racial SES Gini index, but since we focus on the effects of the racial diversity index, the results of the resampling method are reassuring.[32]

-0.005, and is insignificant (p-value = 0.004). We conclude that social desirability bias does not have a significant impact on our estimates.

[32]Results are available on request.

5 Discussion

This paper finds that racial diversity in high school has a positive impact on individuals' voting behavior in early adulthood. We show that this result is likely due to positive and persistent interracial contact. Respondents exposed to more quasi-randomly occurring diversity in adolescence have more friends of other races, both in high school and more than a decade later. The point estimates suggest that the effect sizes are nontrivial: an increase of one within-school standard deviation in the racial diversity index leads to an increase of 1.6 percentage points in the probability to be registered to vote seven years later and an increase of 1.1 percentage points in the probability to have voted six years later.

These results underscore that, beyond their instrumental role as transmitters of knowledge, schools are important arenas for socialization, a role that is often overlooked by research (Gradstein and Justman, 2000, 2002). This role should be considered in the design of educational policies with a focus on racial diversity. However, more research is needed to understand whether our results are generalizable to other contexts, especially to contexts in which increases in diversity are of a greater order of magnitude than those studied in this paper, or to contexts in which diversity is imposed exogenously rather than arising by chance. Another direction for future research is to better understand the channels—such as beliefs or preferences—through which early intergroup contact affects voting in the long run.

References

Agresti, A. (2002). *Categorical Data Analysis (Wiley Series in Probability and Statistics)*. Wiley-Interscience.

Akbulut-Yuksel, M., Okoye, D., and Yuksel, M. (2019). Social changes in impressionable years and adult political attitudes: Evidence from Jewish expulsions in Nazi Germany. *Economic Inquiry*.

Alesina, A., Baqir, R., and Easterly, W. (1999). Public goods and ethnic divisions. *The Quarterly Journal of Economics*, 114(4):1243–1284.

Alesina, A. and La Ferrara, E. (2000). Participation in heterogeneous communities. *The Quarterly Journal of Economics*, 115(3):847–904.

Alesina, A. and La Ferrara, E. (2005). Ethnic diversity and economic performance. *Journal of Economic Literature*, 43(3):762–800.

Alesina, A., Michalopoulos, S., and Papaioannou, E. (2016). Ethnic inequality. *Journal of Political Economy*, 124(2):428–488.

Algan, Y., Do, Q.-A., Dalvit, N., Chapelain, A. L., and Zenou, Y. (2019). Friendship networks and political opinions: A natural experiment among future French politicians. CEPR Discussion Paper No. 13771.

Algan, Y., Hémet, C., and Laitin, D. D. (2016). The social effects of ethnic diversity at the local level: A natural experiment with exogenous residential allocation. *Journal of Political Economy*, 124(3):696–733.

Allport, G. W. (1954). *The Nature of Prejudice*. Addison-Wesley Pub. Co.

Almås, I., Cappelen, A. W., Sørensen, E. Ø., and Tungodden, B. (2010). Fairness and the development of inequality acceptance. *Science*, 328(5982):1176–1178.

Andreoni, J., Payne, A. A., Smith, J., and Karp, D. (2016). Diversity and donations: The effect of religious and ethnic diversity on charitable giving. *Journal of Economic Behavior & Organization*, 128:47–58.

Angrist, J. D. and Lang, K. (2004). Does school integration generate peer effects? Evidence from Boston's Metco program. *American Economic Review*, 94(5):1613–1634.

Bellettini, G., Ceroni, C. B., and Monfardini, C. (2020). Immigration, ethnic diversity and voting: The role of individual income. *European Journal of Political Economy*, 61:101840.

Bergman, P. (2020). The risks and benefits of school integration for participating students: Evidence from a randomized desegregation program. Working paper.

Bifulco, R., Fletcher, J. M., Oh, S. J., and Ross, S. L. (2014). Do high school peers have persistent effects on college attainment and other life outcomes? *Labour Economics*, 29:83–90.

Bifulco, R., Fletcher, J. M., and Ross, S. L. (2011). The effect of classmate characteristics on post-secondary outcomes: Evidence from the Add Health. *American Economic Journal: Economic Policy*, 3(1):25–53.

Billings, S. B., Chyn, E., and Haggag, K. (2020). The long-run effects of school racial diversity on political identity. NBER Working Paper No. 27302.

Black, S. E., Devereux, P. J., and Salvanes, K. G. (2013). Under pressure? The effect of peers on outcomes of young adults. *Journal of Labor Economics*, 31(1):119–153.

Boisjoly, J., Duncan, G. J., Kremer, M., Levy, D. M., and Eccles, J. (2006). Empathy or antipathy? The impact of diversity. *American Economic Review*, 96(5):1890–1905.

Borghans, L., Duckworth, A. L., Heckman, J. J., and ter Weel, B. (2008). The economics and psychology of personality traits. *Journal of Human Resources*, 43(4):972–1059.

Burns, J., Corno, L., and La Ferrara, E. (2019). Interaction, stereotypes and performance. Evidence from South Africa. IFS Working Paper W19/03.

Caeyers, B. and Fafchamps, M. (2020). Exclusion bias in the estimation of peer effects. CEPR Discussion Paper No. 14386.

Cancela, J. and Geys, B. (2016). Explaining voter turnout: A meta-analysis of national and subnational elections. *Electoral Studies*, 42:264–275.

Carrell, S. E. and Hoekstra, M. L. (2010). Externalities in the classroom: How children exposed to domestic violence affect everyone's kids. *American Economic Journal: Applied Economics*, 2(1):211–228.

Cooper, C. A., Golden, L., and Socha, A. (2012). The big five personality factors and mass politics. *Journal of Applied Social Psychology*, 43(1):68–82.

Costa, D. L. and Kahn, M. E. (2003). Civic engagement and community heterogeneity: An economist's perspective. *Perspective on Politics*, 1(1):103–111.

Dahlberg, M., Edmark, K., and Lundqvist, H. (2012). Ethnic diversity and preferences for redistribution. *Journal of Political Economy*, 120(1):41–76.

de Rooij, E. A., Green, D. P., and Gerber, A. S. (2009). Field experiments on political behavior and collective action. *Annual Review of Political Science*, 12(1):389–395.

DellaVigna, S., List, J. A., Malmendier, U., and Rao, G. (2016). Voting to tell others. *The Review of Economic Studies*, 84(1):143–181.

Dinesen, P. T. and Sønderskov, K. M. (2015). Ethnic diversity and social trust. *American Sociological Review*, 80(3):550–573.

Easterly, W. and Levine, R. (1997). Africa's growth tragedy: Policies and ethnic divisions. *The Quarterly Journal of Economics*, 112(4):1203–1250.

Feld, J. and Zölitz, U. (2017). Understanding peer effects: On the nature, estimation, and channels of peer effects. *Journal of Labor Economics*, 35(2):387–428.

Finseraas, H., Hanson, T., Johnsen, Å. A., Kotsadam, A., and Torsvik, G. (2019). Trust, ethnic diversity, and personal contact: A field experiment. *Journal of Public Economics*, 173:72–84.

Fisman, R., Jakiela, P., and Kariv, S. (2017). Distributional preferences and political behavior. *Journal of Public Economics*, 155:1–10.

Friesen, J. and Krauth, B. (2011). Ethnic enclaves in the classroom. *Labour Economics*, 18(5):656–663.

Gerber, A. S., Huber, G. A., Doherty, D., and Dowling, C. M. (2011a). The big five personality traits in the political arena. *Annual Review of Political Science*, 14(1):265–287.

Gerber, A. S., Huber, G. A., Doherty, D., Dowling, C. M., Raso, C., and Ha, S. E. (2011b). Personality traits and participation in political processes. *The Journal of Politics*, 73(3):692–706.

Giuliano, P. and Spilimbergo, A. (2013). Growing up in a recession. *The Review of Economic Studies*, 81(2):787–817.

Gould, E. D., Lavy, V., and Paserman, M. D. (2009). Does immigration affect the long-term educational outcomes of natives? Quasi-experimental evidence. *The Economic Journal*, 119(540):1243–1269.

Gradstein, M. and Justman, M. (2000). Human capital, social capital, and public schooling. *European Economic Review*, 44(4-6):879–890.

Gradstein, M. and Justman, M. (2002). Education, social cohesion, and economic growth. *American Economic Review*, 92(4):1192–1204.

Guiso, L., Sapienza, P., and Zingales, L. (2004). The role of social capital in financial development. *American Economic Review*, 94(3):526–556.

Guryan, J., Kroft, K., and Notowidigdo, M. J. (2009). Peer effects in the workplace: Evidence from random groupings in professional golf tournaments. *American Economic Journal: Applied Economics*, 1(4):34–68.

Hanushek, E. A., Kain, J. F., and Rivkin, S. G. (2009). New evidence about Brown v. Board of Education: The complex effects of school racial composition on achievement. *Journal of Labor Economics*, 27(3):349–383.

Hoxby, C. M. (2000a). The effects of class size on student achievement: New evidence from population variation. *The Quarterly Journal of Economics*, 115(4):1239–1285.

Hoxby, C. M. (2000b). Peer effects in the classroom: Learning from gender and race variation. NBER Working Paper No. 7867.

Imbens, G. W. and Manski, C. F. (2004). Confidence intervals for partially identified parameters. *Econometrica*, 72(6):1845–1857.

Jennings, M. K. and Markus, G. B. (1977). The effect of military service on political attitudes: A panel study. *American Political Science Review*, 71(01):131–147.

Jennings, M. K. and Markus, G. B. (1984). Partisan orientations over the long haul: Results from the three-wave political socialization panel study. *American Political Science Review*, 78(04):1000–1018.

Kaplan, E., Spenkuch, J. L., and Tuttle, C. (2019). School desegregation and political preferences: Long-run evidence from Kentucky. Working paper.

Kerschbamer, R. and Müller, D. (2020). Social preferences and political attitudes: An online experiment on a large heterogeneous sample. *Journal of Public Economics*, 182:104076.

Kim, Y.-I. and Lee, J. (2014). The long-run impact of a traumatic experience on risk aversion. *Journal of Economic Behavior & Organization*, 108:174–186.

Knack, S. and Keefer, P. (1997). Does social capital have an economic payoff? A cross-country investigation. *The Quarterly Journal of Economics*, 112(4):1251–1288.

Lavy, V., Paserman, M. D., and Schlosser, A. (2012). Inside the black box of ability peer effects: Evidence from variation in the proportion of low achievers in the classroom. *The Economic Journal*, 122(559):208–237.

Lavy, V. and Schlosser, A. (2011). Mechanisms and impacts of gender peer effects at school. *American Economic Journal: Applied Economics*, 3(2):1–33.

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *Review of Economic Studies*, 76(3):1071–1102.

Lowe, M. (2019). *Types of Contact: A Field Experiment on Collaborative and Adversarial Caste Integration*. PhD thesis, Department of Economics, MIT.

Luttmer, E. F. P. (2001). Group loyalty and the taste for redistribution. *Journal of Political Economy*, 109(3):500–528.

Madestam, A. and Yanagizawa-Drott, D. (2012). Shaping of the nation: The effect of Fourth of July on political preferences and behavior in the United States. HKS Faculty Research Working Paper Series 12-034, John F. Kennedy School of Government, Harvard University.

Malmendier, U. and Nagel, S. (2011). Depression babies: Do macroeconomic experiences affect risk taking? *The Quarterly Journal of Economics*, 126(1):373–416.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, 60(3):531.

Marschall, M. J. and Stolle, D. (2004). Race and the city: Neighborhood context and the development of generalized trust. *Political Behavior*, 26(2):125–153.

Martinez i Coma, F. and Nai, A. (2017). Ethnic diversity decreases turnout. Comparative evidence from over 650 elections around the world. *Electoral Studies*, 49:75–95.

Merlino, L. P., Steinhardt, M. F., and Wren-Lewis, L. (2019). More than just friends? School peers and adult interracial relationships. *Journal of Labor Economics*, 37(3):663–713.

Mondak, J. J. (2010). *Personality and the Foundations of Political Behavior*. Cambridge University Press.

Mondak, J. J., Hibbing, M. V., Canache, D., Seligson, M. A., and Anderson, M. R. (2010). Personality and civic engagement: An integrative framework for the study of trait effects on political behavior. *American Political Science Review*, 104(1):85–110.

Montalvo, J. G. and Reynal-Querol, M. (2005). Ethnic diversity and economic development. *Journal of Development Economics*, 76(2):293–323.

Moreau, M.-C. (2011). What is "language arts"?

Mousa, S. (2020). Building social cohesion between Christians and Muslims through soccer in post-ISIS Iraq. *Science*, 369(6505):866–870.

Oberholzer-Gee, F. and Waldfogel, J. (2001). Electoral acceleration: The effect of minority population on minority voter turnout. NBER Working Paper No. 8252.

Oberholzer-Gee, F. and Waldfogel, J. (2005). Strength in numbers: Group size and political mobilization. *The Journal of Law and Economics*, 48(1):73–91.

Pettigrew, T. F. and Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, 90(5):751–783.

Pettigrew, T. F., Tropp, L. R., Wagner, U., and Christ, O. (2011). Recent advances in intergroup contact theory. *International Journal of Intercultural Relations*, 35(3):271–280.

Putnam, R. D. (2007). E pluribus unum: Diversity and community in the twenty-first century. The 2006 Johan Skytte prize lecture. *Scandinavian Political Studies*, 30(2):137–174.

Rao, G. (2019). Familiarity does not breed contempt: Generosity, discrimination, and diversity in Delhi schools. *American Economic Review*, 109(3):774–809.

Rogers, T., Ternovski, J., and Yoeli, E. (2016). Potential follow-up increases private contributions to public goods. *Proceedings of the National Academy of Sciences*, 113(19):5218–5220.

Schindler, D. and Westcott, M. (2020). Shocking racial attitudes: Black G.I.s in Europe. *The Review of Economic Studies*.

Shertzer, A. (2016). Immigrant group size and political mobilization: Evidence from European migration to the United States. *Journal of Public Economics*, 139:1–12.

Steinmayr, A. (2020). Contact versus exposure: Refugee presence and voting for the far-right. *The Review of Economics and Statistics*, pages 1–47.

Tauchmann, H. (2014). Lee (2009) treatment-effect bounds for nonrandom sample selection. *The Stata Journal*, 14(4):884–894.

Vigdor, J. L. (2004). Community composition and collective action: Analyzing initial mail response to the 2000 Census. *Review of Economics and Statistics*, 86(1):303–312.

Westfall, P. H. and Young, S. (1993). *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment*. John Wiley & Sons.

Table 1: Summary statistics

	All	White	Minority
Main variables			
Is registered to vote	0.719	0.740	0.674
	(0.450)	(0.438)	(0.469)
Voted in 2000	0.429	0.440	0.408
	(0.495)	(0.496)	(0.491)
Identifies with a party	0.333	0.333	0.332
	(0.471)	(0.471)	(0.471)
Democrat	0.181	0.136	0.273
	(0.385)	(0.343)	(0.446)
Republican	0.152	0.198	0.059
	(0.359)	(0.398)	(0.235)
Racial diversity index	0.376	0.328	0.475
	(0.199)	(0.180)	(0.199)
Racial SES Gini	0.362	0.366	0.355
	(0.181)	(0.183)	(0.174)
Shares in cohort			
Share white	0.623	0.763	0.335
	(0.302)	(0.197)	(0.275)
Share black	0.175	0.096	0.336
	(0.240)	(0.150)	(0.302)
Share Hispanic	0.123	0.074	0.224
	(0.177)	(0.091)	(0.251)
Share Asian	0.041	0.028	0.066
	(0.076)	(0.051)	(0.108)
Share other	0.039	0.039	0.039
	(0.034)	(0.028)	(0.045)
Own race			
White	0.673		
	(0.469)		
Black	0.156		
	(0.363)		
Hispanic	0.116		
	(0.320)		
Asian	0.041		
	(0.199)		
Other	0.014		
	(0.116)		
N	12,568	6,710	5,858

Notes: Summary statistics are calculated using Wave 3 longitudinal weights, which aim to produce a representative sample of individuals who were surveyed in both Waves 1 and 3.

	All	White	Minority
Wave 1: individual characteristics			
Male	0.508	0.505	0.513
	(0.500)	(0.500)	(0.500)
Age	15.916	15.855	16.044
	(1.795)	(1.782)	(1.813)
GPA	2.803	2.870	2.665
	(0.772)	(0.786)	(0.722)
Grade 7 indicator	0.178	0.185	0.164
	(0.383)	(0.388)	(0.370)
Grade 8 indicator	0.166	0.170	0.159
	(0.372)	(0.376)	(0.366)
Grade 9 indicator	0.166	0.167	0.164
	(0.372)	(0.373)	(0.370)
Grade 10 indicator	0.169	0.163	0.183
	(0.375)	(0.369)	(0.386)
Grade 11 indicator	0.157	0.153	0.166
	(0.364)	(0.360)	(0.373)
Grade 12 indicator	0.163	0.163	0.164
	(0.369)	(0.369)	(0.370)
Wave 1: family characteristics			
Mother high school dropout	0.165	0.112	0.276
	(0.372)	(0.315)	(0.447)
Mother high school graduate	0.372	0.387	0.341
	(0.483)	(0.487)	(0.474)
Mother some college	0.227	0.246	0.189
	(0.419)	(0.430)	(0.391)
Mother college graduate	0.236	0.256	0.195
	(0.425)	(0.436)	(0.396)
Family income (k)	45.051	50.061	34.750
	(38.787)	(38.925)	(36.402)
English not spoken at home	0.073	0.005	0.212
	(0.259)	(0.070)	(0.409)
Lives with both biological parents	0.573	0.617	0.482
	(0.463)	(0.461)	(0.455)
Parent civically engaged	0.503	0.548	0.409
	(0.471)	(0.474)	(0.449)
Missing parent information	0.313	0.266	0.408
	(0.464)	(0.442)	(0.491)
Wave 1: neighborhood characteristics			
Share less than high school	0.275	0.242	0.342
	(0.157)	(0.135)	(0.177)
Share with bachelor's degree	0.221	0.231	0.201
	(0.144)	(0.145)	(0.141)
Share votes for Democratic candidate 1992	0.423	0.405	0.461
	(0.095)	(0.084)	(0.106)
Share blacks	0.137	0.048	0.320
	(0.259)	(0.113)	(0.360)
Share Hispanics	0.072	0.029	0.161
	(0.162)	(0.063)	(0.246)
Share Asians & other	0.033	0.018	0.064
	(0.086)	(0.043)	(0.131)
Share below poverty level	0.144	0.110	0.213
	(0.138)	(0.104)	(0.170)
Urban area	0.511	0.427	0.682
	(0.500)	(0.495)	(0.466)
N	12,568	6,710	5,858

Notes: Summary statistics are calculated using Wave 3 longitudinal weights, which aim to produce a representative sample of individuals who were surveyed in both Waves 1 and 3.

	All	White	Minority
Wave 1: cohort controls			
Share mothers with college degree	0.296	0.296	0.295
	(0.139)	(0.138)	(0.141)
Share males	0.498	0.503	0.488
	(0.072)	(0.077)	(0.058)
Ability grouping	0.457	0.437	0.499
	(0.498)	(0.496)	(0.500)
Wave 1: other			
Average class size in school	25.805	24.965	27.531
	(4.684)	(4.379)	(4.817)
Share interracial friendships	0.233	0.196	0.324
	(0.226)	(0.209)	(0.238)
White-non white friend	0.462	0.505	0.357
	(0.499)	(0.500)	(0.479)
Region			
Northeast	0.137	0.153	0.106
	(0.344)	(0.360)	(0.308)
Midwest	0.294	0.358	0.164
	(0.456)	(0.479)	(0.370)
South	0.406	0.361	0.498
	(0.491)	(0.480)	(0.500)
West	0.163	0.129	0.232
	(0.369)	(0.335)	(0.422)
Wave 3 variables			
Married	0.167	0.178	0.142
	(0.373)	(0.383)	(0.349)
Education (years)	13.123	13.232	12.899
	(1.976)	(1.976)	(1.958)
Working	0.746	0.762	0.711
	(0.435)	(0.426)	(0.453)
Annual income (k)	13.748	14.221	12.738
	(14.888)	(14.164)	(16.282)
It is important to fit in	0.344	0.374	0.283
	(0.475)	(0.484)	(0.450)
N	12,568	6,710	5,858

Notes: Summary statistics are calculated using Wave 3 longitudinal weights, which aim to produce a representative sample of individuals who were surveyed in both Waves 1 and 3.

	All	White	Minority
Wave 1: behavior			
Smoking	0.258	0.303	0.167
	(0.438)	(0.459)	(0.373)
Marijuana use	0.136	0.134	0.140
	(0.343)	(0.341)	(0.347)
Binge drinking	0.266	0.293	0.211
	(0.442)	(0.455)	(0.408)
Wave 3: behavior			
Drop out of high school	0.172	0.154	0.207
	(0.377)	(0.361)	(0.405)
Attend college	0.565	0.594	0.504
	(0.496)	(0.491)	(0.500)
Test score	101.902	105.141	95.246
	(14.519)	(11.417)	(17.616)
Idleness	0.133	0.114	0.171
	(0.339)	(0.318)	(0.377)
Smoking	0.353	0.409	0.238
	(0.478)	(0.492)	(0.426)
Marijuana use	0.228	0.248	0.189
	(0.420)	(0.432)	(0.392)
Binge drinking	0.511	0.590	0.349
	(0.500)	(0.492)	(0.477)
Wave 4 variables			
Extraversion	3.306	3.341	3.233
	(0.772)	(0.778)	(0.752)
Agreeableness	3.804	3.835	3.741
	(0.608)	(0.609)	(0.601)
Conscientiousness	3.640	3.622	3.677
	(0.670)	(0.691)	(0.622)
Neuroticism	2.597	2.568	2.655
	(0.684)	(0.691)	(0.666)
Imagination/Intellect	3.627	3.653	3.573
	(0.618)	(0.628)	(0.593)
Has at least one friend of another race	0.549	0.516	0.617
	(0.498)	(0.500)	(0.486)
N	12,568	6,710	5,858

Notes: Summary statistics are calculated using Wave 3 longitudinal weights for variables in Wave 1, which aim to produce a representative sample of individuals who were surveyed in both Waves 1 and 3. For variables in Wave 4, we use weights for longitudinal analyses with Waves 1, 3 and 4. Personality trait scores are calculated as averages over 4 questions with answers ranging from 1 to 5, where higher numbers indicate more of that trait than lower numbers.

Table 2: Variation in cohort composition measures after removing cohort and school fixed effects and trends

	Panel A Raw cohort variables				
	N	Mean	Standard deviation	Min	Max
Share white	12,568	0.623	0.302	0.000	1.000
Share black	12,568	0.175	0.240	0.000	1.000
Racial diversity index	12,568	0.376	0.199	0.000	0.777
Racial SES Gini	12,568	0.362	0.181	0.000	0.752
	Panel B Residuals after removing school and cohort fixed effects				
	N	Mean	Standard deviation	Min	Max
Share white	12,568	0.000	0.033	−0.271	0.214
Share black	12,568	0.000	0.026	−0.295	0.126
Racial diversity index	12,568	0.000	0.039	−0.141	0.358
Racial SES Gini	12,568	0.002	0.120	−0.428	0.414
	Panel C Residuals after removing school and cohort fixed effects and trends				
	N	Mean	Standard deviation	Min	Max
Share white	12,568	0.000	0.025	−0.182	0.176
Share black	12,568	0.000	0.017	−0.170	0.192
Racial diversity index	12,568	0.000	0.030	−0.123	0.253
Racial SES Gini	12,568	0.002	0.085	−0.346	0.384

Table 3: Balancing tests: Racial diversity index

	OLS (1)	School fixed effects (2)	School fixed effects + trends (3)
Individual characteristics W1			
Male	0.042	0.140	0.188
	(0.038)	(0.121)	(0.160)
Age	−0.145	−0.149	−0.264
	(0.080)	(0.165)	(0.214)
GPA	−0.201	−0.013	−0.208
	(0.107)	(0.206)	(0.192)
Ability grouping	0.309	0.065	−0.221
	(0.197)	(0.210)	(0.386)
Family characteristics W1			
Mother's education (years)	0.260	−0.037	0.689
	(0.381)	(0.512)	(0.619)
Log of family income	0.062	0.015	0.043
	(0.033)	(0.049)	(0.065)
English not spoken at home	0.005	0.027	0.043
	(0.038)	(0.039)	(0.047)
Lives with both biological parents	−0.106	−0.045	−0.025
	(0.049)	(0.113)	(0.162)
Parent civically engaged	0.051	0.147	0.066
	(0.061)	(0.129)	(0.163)
Missing parent information	0.077	0.141	0.179
	(0.055)	(0.147)	(0.203)
Neighborhood characteristics W1			
Share less than high school	−0.116	0.004	−0.026
	(0.048)	(0.024)	(0.035)
Share with bachelor's degree	0.139	−0.027	−0.002
	(0.043)	(0.026)	(0.040)
Share votes for Democratic candidate 1992	0.037	−0.005	−0.005
	(0.040)	(0.005)	(0.007)
Share blacks	0.043	−0.036	−0.021
	(0.050)	(0.035)	(0.041)
Share Hispanics	0.058	−0.005	0.006
	(0.052)	(0.016)	(0.017)
Share Asians & other	0.117	0.003	−0.010
	(0.023)	(0.009)	(0.011)
Share below poverty level	−0.029	0.004	−0.012
	(0.039)	(0.020)	(0.025)
Urban area	0.907	0.057	−0.096
	(0.159)	(0.065)	(0.096)
N	12,568	12,568	12,568

Notes: Each coefficient is from a separate regression where each of the variables listed (measured in Wave 1) is regressed on the racial diversity index with controls including own race and own grade dummies (1), plus school fixed effects (2) and school linear trends (3). We report the coefficient of the racial diversity index. The figures in parentheses are standard errors robust to clustering at the school level. Wave 3 longitudinal weights are used.

Table 4: Voting behavior in Wave 3

Dependent variable:	Registered to vote			Voted in 2000		
	(1)	(2)	(3)	(4)	(5)	(6)
Racial diversity index	0.016	0.016	0.016	0.011	0.011	0.011
	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)
Racial SES Gini	−0.003	−0.004	−0.004	−0.006	−0.006	−0.006
	(0.006)	(0.006)	(0.006)	(0.004)	(0.004)	(0.004)
Share mothers with college degree	0.076	0.040	0.042	0.204	0.148	0.151
	(0.137)	(0.137)	(0.139)	(0.169)	(0.165)	(0.166)
Ability grouping	0.009	0.013	0.009	0.087	0.089	0.088
	(0.029)	(0.027)	(0.028)	(0.037)	(0.037)	(0.038)
Black	0.055	0.069	0.060	0.061	0.089	0.066
	(0.018)	(0.019)	(0.021)	(0.023)	(0.023)	(0.022)
Hispanic	−0.102	−0.012	−0.002	−0.121	−0.048	−0.038
	(0.027)	(0.023)	(0.023)	(0.020)	(0.020)	(0.020)
Asian	−0.205	−0.120	−0.108	−0.215	−0.160	−0.155
	(0.033)	(0.035)	(0.036)	(0.030)	(0.028)	(0.028)
Other	−0.063	−0.007	−0.005	−0.145	−0.100	−0.103
	(0.049)	(0.047)	(0.049)	(0.047)	(0.047)	(0.048)
Constant	−0.273	−0.531	−0.570	−0.909	−1.146	−1.059
	(0.201)	(0.193)	(0.217)	(0.232)	(0.223)	(0.246)
Individual characteristics	✓	✓	✓	✓	✓	✓
Family characteristics		✓	✓		✓	✓
Neighborhood characteristics			✓			✓
N	12,568	12,568	12,568	12,568	12,568	12,568

Notes: This table reports OLS estimates. The racial diversity index and the racial SES Gini index are standardized at the school cohort level, after controlling for school linear trends. Controls include school and grade fixed effects and school linear trends. The omitted category for own race is white. Wave 3 longitudinal weights are used. Standard errors (in parentheses) are clustered at the school level. All controls are listed in Table 1 under individual characteristics, family characteristics, neighborhood characteristics and cohort controls. Variable definitions are in Appendix Table A.1.

Table 5: Political partisanship in Wave 3

Dependent variable:	Identifies with a party			Identifies as a Democrat		
	(1)	(2)	(3)	(4)	(5)	(6)
Racial diversity index	−0.003	−0.003	−0.003	−0.004	−0.004	−0.004
	(0.006)	(0.006)	(0.006)	(0.004)	(0.004)	(0.004)
Racial SES Gini	0.006	0.006	0.006	0.009	0.009	0.009
	(0.005)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)
Share mothers with college degree	0.225	0.188	0.191	0.074	0.061	0.061
	(0.195)	(0.194)	(0.193)	(0.129)	(0.128)	(0.128)
Ability grouping	−0.039	−0.035	−0.036	0.003	0.006	0.006
	(0.050)	(0.050)	(0.050)	(0.037)	(0.038)	(0.038)
Black	0.096	0.115	0.114	0.223	0.226	0.213
	(0.023)	(0.024)	(0.026)	(0.024)	(0.025)	(0.025)
Hispanic	−0.083	−0.039	−0.033	0.008	0.025	0.028
	(0.026)	(0.025)	(0.025)	(0.022)	(0.021)	(0.020)
Asian	−0.159	−0.133	−0.129	−0.040	−0.031	−0.032
	(0.028)	(0.032)	(0.032)	(0.029)	(0.032)	(0.032)
Other	−0.033	−0.005	−0.003	0.053	0.064	0.062
	(0.058)	(0.057)	(0.056)	(0.041)	(0.041)	(0.040)
Constant	−0.372	−0.522	−0.495	−0.346	−0.363	−0.338
	(0.189)	(0.194)	(0.223)	(0.128)	(0.133)	(0.177)
Individual characteristics	✓	✓	✓	✓	✓	✓
Family characteristics		✓	✓		✓	✓
Neighborhood characteristics			✓			✓
N	12,568	12,568	12,568	12,568	12,568	12,568

Notes: This table reports OLS estimates. The racial diversity index and the racial SES Gini index are standardized at the school cohort level, after controlling for school linear trends. Controls include school and grade fixed effects and school linear trends. The omitted category for own race is white. Wave 3 longitudinal weights are used. Standard errors (in parentheses) are clustered at the school level. All controls are listed in Table 1 under individual characteristics, family characteristics, neighborhood characteristics and cohort controls. Variable definitions are in Appendix Table A.1.

Table 6: Friendships

Dependent variable:	Share interracial friendships W1 (1)	White-nonwhite friendship W1 (2)	Has at least one friend of another race W4 (3)
Racial diversity index	0.012	0.036	0.009
	(0.003)	(0.007)	(0.005)
Racial SES Gini	0.001	−0.005	0.001
	(0.002)	(0.007)	(0.005)
Share mothers with college degree	0.094	0.169	0.261
	(0.086)	(0.181)	(0.166)
Ability grouping	−0.011	−0.079	0.001
	(0.022)	(0.046)	(0.041)
Black	0.068	−0.175	−0.012
	(0.021)	(0.046)	(0.029)
Hispanic	0.160	0.098	0.147
	(0.018)	(0.044)	(0.028)
Asian	0.086	0.063	0.231
	(0.033)	(0.068)	(0.028)
Other	0.079	0.100	0.162
	(0.030)	(0.075)	(0.057)
Constant	−0.041	0.369	−0.302
	(0.123)	(0.311)	(0.256)
Individual characteristics	✓	✓	✓
Family characteristics	✓	✓	✓
Neighborhood characteristics	✓	✓	✓
N	8,072	8,474	10,545

Notes: This table reports OLS estimates. The racial diversity index and the racial SES Gini index are standardized at the school cohort level, after controlling for school linear trends. Controls include school and grade fixed effects and school linear trends. The omitted category for own race is white. Wave 3 longitudinal weights are used. Standard errors (in parentheses) are clustered at the school level. All controls are listed in Table 1 under individual characteristics, family characteristics, neighborhood characteristics and cohort controls. Variable definitions are in Appendix Table A.1.

Table 7: Personality in Wave 4

	Average score				
	Extraversion	Agreeableness	Conscientiousness	Neuroticism	Imagination/Intellect
Racial diversity index	0.017	−0.002	0.012	−0.003	0.002
	(0.009)	(0.009)	(0.007)	(0.006)	(0.006)
Racial SES Gini	0.002	−0.009	−0.016	0.004	−0.012
	(0.008)	(0.008)	(0.009)	(0.007)	(0.005)
Share mothers with college degree	0.831	0.570	0.162	−0.495	0.188
	(0.307)	(0.179)	(0.289)	(0.251)	(0.212)
	Principal component analysis: score for component 1				
	Extraversion	Agreeableness	Conscientiousness	Neuroticism	Imagination/Intellect
Racial diversity index	0.033	−0.005	0.026	−0.007	0.005
	(0.018)	(0.022)	(0.014)	(0.013)	(0.014)
Racial SES Gini	0.005	−0.022	−0.031	0.005	−0.029
	(0.016)	(0.018)	(0.018)	(0.014)	(0.012)
Share mothers with college degree	1.575	1.344	0.362	−1.035	0.416
	(0.597)	(0.431)	(0.605)	(0.467)	(0.491)
N	10,795	10,800	10,803	10,801	10,722

Notes: All regressions include both cohort composition variables along with controls for cohort fixed effects, school fixed effects, and school trends, as well as the individual student covariates related to the cohort variables. All dependent variables are measured using Wave 4 of the Add Health. Wave 4 longitudinal weights are used, for participants who were also interviewed at Waves 1, 3 and 4. Figures in parentheses are standard errors robust to clustering at the school level. All controls are listed in Table 1 under individual characteristics, family characteristics, neighborhood characteristics and cohort controls. Variable definitions are in Appendix Table A.1.

Table 8: Sample splits: Voting behavior

Dependent variable:	Registered to vote				Voted in 2000			
	White	Minority	Family income > 40k	Family income ≤ 40k	White	Minority	Family income > 40k	Family income ≤ 40k
Racial diversity index	0.017	0.003	0.018	0.016	0.010	0.019	0.012	0.011
	(0.005)	(0.011)	(0.005)	(0.008)	(0.006)	(0.011)	(0.006)	(0.008)
Racial SES Gini	−0.006	0.001	−0.004	0.000	−0.008	0.001	−0.010	0.000
	(0.008)	(0.010)	(0.006)	(0.008)	(0.006)	(0.008)	(0.008)	(0.005)
Share mothers with college degree	0.057	0.029	0.202	−0.128	0.197	−0.017	−0.165	0.506
	(0.165)	(0.291)	(0.188)	(0.184)	(0.209)	(0.230)	(0.224)	(0.239)
Ability grouping	0.005	0.093	−0.046	0.065	0.109	0.053	0.039	0.162
	(0.027)	(0.083)	(0.022)	(0.062)	(0.038)	(0.083)	(0.031)	(0.084)
Black			0.043	0.060			0.105	0.069
			(0.029)	(0.031)			(0.037)	(0.034)
Hispanic		−0.074	0.027	−0.002		−0.099	−0.034	−0.037
		(0.033)	(0.032)	(0.030)		(0.034)	(0.038)	(0.030)
Asian		−0.138	−0.176	0.016		−0.207	−0.164	−0.123
		(0.049)	(0.045)	(0.061)		(0.036)	(0.040)	(0.055)
Other		−0.051	−0.035	0.070		−0.148	−0.098	−0.057
		(0.055)	(0.072)	(0.073)		(0.052)	(0.067)	(0.076)
Constant	−0.259	−1.121	−0.408	−0.753	−1.352	−0.714	−0.741	−1.551
	(0.278)	(0.429)	(0.297)	(0.273)	(0.285)	(0.429)	(0.347)	(0.339)
Individual characteristics	✓	✓	✓	✓	✓	✓	✓	✓
Family characteristics	✓	✓	✓	✓	✓	✓	✓	✓
Neighborhood characteristics	✓	✓	✓	✓	✓	✓	✓	✓
p-Value, coeffs. equal	0.233		0.818		0.473		0.974	
N	6,710	5,858	6,099	6,469	6,710	5,858	6,099	6,469

Notes: This table reports OLS estimates. The racial diversity index and the racial SES Gini index are standardized at the school cohort level, after controlling for school linear trends. Controls include school and grade fixed effects and school linear trends. The omitted category for own race is white in the family income sample splits, and it is black in the minority subsample. Wave 3 longitudinal weights are used. Standard errors (in parentheses) are clustered at the school level.All controls are listed in Table 1 under individual characteristics, family characteristics, neighborhood characteristics and cohort controls. Variable definitions are in Appendix Table A.1.

Table 9: Sample splits: Political partisanship

Dependent variable:	Identify with a party				Identify as a Democrat			
	White	Minority	Family income > 40k	Family income ≤ 40k	White	Minority	Family income > 40k	Family income ≤ 40k
Racial diversity index	−0.008	0.008	−0.010	0.008	−0.007	0.006	−0.007	0.003
	(0.006)	(0.010)	(0.010)	(0.006)	(0.005)	(0.010)	(0.007)	(0.005)
Racial SES Gini	0.006	0.001	0.003	0.011	0.013	0.002	0.010	0.010
	(0.007)	(0.007)	(0.009)	(0.006)	(0.004)	(0.006)	(0.007)	(0.004)
Share mothers with college degree	0.364	−0.160	0.122	0.176	0.211	−0.428	0.067	0.039
	(0.228)	(0.366)	(0.254)	(0.258)	(0.139)	(0.263)	(0.199)	(0.156)
Ability grouping	−0.050	0.071	−0.139	0.128	0.005	0.028	−0.026	0.071
	(0.053)	(0.083)	(0.070)	(0.037)	(0.042)	(0.090)	(0.056)	(0.033)
Black			0.137	0.134			0.269	0.216
			(0.036)	(0.032)			(0.039)	(0.029)
Hispanic		−0.152	−0.073	−0.003		−0.211	−0.020	0.058
		(0.036)	(0.041)	(0.030)		(0.034)	(0.034)	(0.025)
Asian		−0.248	−0.163	−0.055		−0.286	−0.049	0.012
		(0.043)	(0.038)	(0.057)		(0.041)	(0.037)	(0.051)
Other		−0.149	−0.046	0.067		−0.188	0.025	0.106
		(0.067)	(0.067)	(0.085)		(0.061)	(0.050)	(0.059)
Constant	−0.651	−0.349	−0.675	−0.452	−0.492	0.095	−0.784	−0.087
	(0.308)	(0.365)	(0.359)	(0.257)	(0.190)	(0.376)	(0.254)	(0.205)
Individual characteristics	✓	✓	✓	✓	✓	✓	✓	✓
Family characteristics	✓	✓	✓	✓	✓	✓	✓	✓
Neighborhood characteristics	✓	✓	✓	✓	✓	✓	✓	✓
p-Value, coeffs. equal	0.156		0.149		0.212		0.206	
N	6,710	5,858	6,099	6,469	6,710	5,858	6,099	6,469

Notes: This table reports OLS estimates. The racial diversity index and the racial SES Gini index are standardized at the school cohort level, after controlling for school linear trends. Controls include school and grade fixed effects and school linear trends. The omitted category for own race is white in the family income sample splits, and it is black in the minority subsample. Wave 3 longitudinal weights are used. Standard errors (in parentheses) are clustered at the school level. All controls are listed in Table 1 under individual characteristics, family characteristics, neighborhood characteristics and cohort controls. Variable definitions are in Appendix Table A.1.

Appendix A

Table A.1: Description of variables

Variable	Wave	Description	Values
Dependent variables			
Registered to vote	3	Reports being registered to vote	No = 0, Yes = 1
Voted in 2000	3	Reports having voted	No = 0, Yes = 1
Identifies with a party	3	Reports identifying with a political party	No = 0, Yes = 1
– Democrat	3	'Yes' to previous question & reports identifying with the Democratic Party	Other/No = 0, Yes = 1
– Republican	3	'Yes' to previous question & reports identifying with the Republican Party	Other/No = 0, Yes = 1
School cohort composition variables			
Share males in cohort	1	Share of male students in one's cohort	[0,1]
Share black/Hispanic/Asian/other in cohort	1	Share of students in an individual's cohort who define themselves to be black/Hispanic/Asian/other (omitted: white)	[0,1]
Share mothers with college degree	1	Share of students in an individual's cohort whose mothers have a college degree	[0,1]
Racial diversity index	1	One minus the sum of squared racial shares in one's cohort	[0,0.8]
Racial SES Gini	1	Definition in footnote 12	[0,1]
Family characteristics			
Mother's education	1	Dummies for high school dropout, high school graduate, some college, college graduate (imputed if missing)	
Family income	1	Imputed annual family income of individual (log in regression)	in 000's. USD
English spoken at home	1	Dummy variable	No = 0, Yes = 1
Lives with both biological parents	1	Dummy variable	No = 0, Yes = 1
Parent civically engaged	1	Dummy variable if parent answering is a member of any of the following: parent/teacher organization, military veterans organization, labor union, hobby/sports group, civic or social organization (imputed if missing)	No = 0, Yes = 1
Parent dummy	1	Dummy variable if missing parent information on either mother's education, family income, parent's age, parent's civic engagement	No = 0, Yes = 1

Variable	Wave	Description	Values
Neighborhood characteristics			
Urban area	1	Respondent lives in urban area	No = 0, Yes = 1
Share of census block group[a] with less than high school education	1		[0,1]
Share of census block group with a bachelor's degree	1		[0,1]
Share votes for Democratic candidate 1992	1		[0,1]
Share of census block group black/Hispanic/Asian & other	1		[0,1]
Share of census block group below poverty level	1	Share of inhabitants in the census block group with income in 1989 below poverty level	[0,1]
Personality			
Extraversion	4	Average of 4 items (or first principal component) in 20-item mini IPIP scale	[1,5]
Agreeableness	4	Average of 4 items (or first principal component) in 20-item mini IPIP scale	[1,5]
Conscientiousness	4	Average of 4 items (or first principal component) in 20-item mini IPIP scale	[1,5]
Neuroticism	4	Average of 4 items (or first principal component) in 20-item mini IPIP scale	[1,5]
Imagination/Intellect	4	Average of 4 items (or first principal component) in 20-item mini IPIP scale	[1,5]
Other variables			
Ability grouping	1	Principal answered whether English or language arts classes in a grade are grouped by ability	No = 0, Yes = 1
Share interracial friendships	1	Share of friends of other races, as computed by Add Health	[0,1]
White-nonwhite friendship	1	At least a white/minority friend, if minority/white	No = 0, Yes = 1
Has at least a friend of another race	4	Dummy variable	No = 0, Yes = 1
Important to fit in	3	(Strongly) agreed it is important to fit in with group	No = 0, Yes = 1
Drop out of high school	3	Dummy variable	No = 0, Yes = 1
College degree	3	Dummy variable	No = 0, Yes = 1
Test score	1, 3	Standardized Add Health picture vocabulary test score	[9,123]
Idleness	3	Not in school and not working	No = 0, Yes = 1
Smoking	1, 3	Smoked in the past 30 days	No = 0, Yes = 1
Marijuana use	1, 3	Used in the past 30 days	No = 0, Yes = 1
Binge drinking	1, 3	Had drunk at least 5 drinks in a row at least once in the past 12 months	No = 0, Yes = 1

[a] A *census block group* is a cluster of census blocks within a census tract or block numbering area. It is the lowest geographical level for which the Census Bureau publishes sample data. For the 1990 census, block groups averaged 452 housing units, or 1,100 people. A typical census tract contains 4 or 5 block groups.

Appendix B Robustness checks

Table B.1: Robustness check: Restricted sample (Lavy et al., 2012)

Dependent variable:	Registered to vote			Voted in 2000		
	(1)	(2)	(3)	(4)	(5)	(6)
Racial diversity index	0.017	0.017	0.017	0.011	0.011	0.011
	(0.005)	(0.005)	(0.005)	(0.007)	(0.007)	(0.007)
Racial SES Gini	−0.005	−0.005	−0.005	−0.005	−0.004	−0.004
	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)
Share mothers with college degree	−0.083	−0.112	−0.109	0.436	0.372	0.379
	(0.200)	(0.203)	(0.204)	(0.229)	(0.223)	(0.225)
Ability grouping	0.010	0.013	0.010	0.071	0.074	0.072
	(0.033)	(0.030)	(0.031)	(0.041)	(0.040)	(0.041)
Black	0.048	0.062	0.052	0.063	0.092	0.072
	(0.019)	(0.019)	(0.022)	(0.024)	(0.024)	(0.022)
Hispanic	−0.090	−0.005	0.004	−0.113	−0.040	−0.030
	(0.028)	(0.024)	(0.025)	(0.020)	(0.019)	(0.019)
Asian	−0.205	−0.125	−0.115	−0.216	−0.160	−0.156
	(0.034)	(0.036)	(0.037)	(0.030)	(0.028)	(0.028)
Other	−0.068	−0.019	−0.020	−0.146	−0.103	−0.108
	(0.053)	(0.051)	(0.052)	(0.050)	(0.049)	(0.050)
Constant	−0.240	−0.501	−0.533	−1.049	−1.297	−1.215
	(0.230)	(0.221)	(0.248)	(0.276)	(0.269)	(0.285)
Individual characteristics	✓	✓	✓	✓	✓	✓
Family characteristics		✓	✓		✓	✓
Neighborhood characteristics			✓			✓
N	11,435	11,435	11,435	11,435	11,435	11,435

Notes: This table reports OLS estimates. The racial diversity index and the racial SES Gini index are standardized at the school cohort level, after controlling for school linear trends. Controls include school and grade fixed effects and school linear trends. The omitted category for own race is white. Wave 3 longitudinal weights are used. Standard errors (in parentheses) are clustered at the school level. All controls are listed in Table 1 under individual characteristics, family characteristics, neighborhood characteristics and cohort controls. Variable definitions are in Appendix Table A.1.

Table B.2: Robustness check: Restricted sample (Bifulco et al., 2011)

Dependent variable:	Registered to vote			Voted in 2000		
	(1)	(2)	(3)	(4)	(5)	(6)
Racial diversity index	0.016	0.016	0.016	0.015	0.015	0.014
	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)
Racial SES Gini	−0.004	−0.004	−0.004	−0.002	−0.002	−0.002
	(0.006)	(0.006)	(0.006)	(0.004)	(0.004)	(0.004)
Share mothers with college degree	0.041	0.001	0.002	0.239	0.173	0.177
	(0.144)	(0.144)	(0.147)	(0.172)	(0.168)	(0.169)
Ability grouping	0.010	0.013	0.010	0.087	0.089	0.087
	(0.029)	(0.027)	(0.028)	(0.036)	(0.036)	(0.037)
Black	0.057	0.070	0.059	0.063	0.091	0.064
	(0.018)	(0.019)	(0.021)	(0.024)	(0.024)	(0.023)
Hispanic	−0.101	−0.012	−0.002	−0.119	−0.048	−0.037
	(0.028)	(0.023)	(0.024)	(0.020)	(0.020)	(0.020)
Asian	−0.206	−0.121	−0.110	−0.214	−0.161	−0.156
	(0.033)	(0.035)	(0.036)	(0.030)	(0.028)	(0.028)
Other	−0.066	−0.010	−0.008	−0.143	−0.099	−0.102
	(0.049)	(0.048)	(0.050)	(0.048)	(0.048)	(0.049)
Constant	−0.214	−0.472	−0.521	−0.809	−1.041	−0.969
	(0.207)	(0.200)	(0.225)	(0.235)	(0.225)	(0.248)
Individual characteristics	✓	✓	✓	✓	✓	✓
Family characteristics		✓	✓		✓	✓
Neighborhood characteristics			✓			✓
N	12,036	12,036	12,036	12,036	12,036	12,036

Notes: This table reports OLS estimates. The racial diversity index and the racial SES Gini index are standardized at the school cohort level, after controlling for school linear trends. Controls include school and grade fixed effects and school linear trends. The omitted category for own race is white. Wave 3 longitudinal weights are used. Standard errors (in parentheses) are clustered at the school level. All controls are listed in Table 1 under individual characteristics, family characteristics, neighborhood characteristics and cohort controls. Variable definitions are in Appendix Table A.1.

Table B.3: Robustness check: Alternative specification, voting behavior

Dependent variable:	Registered to vote			Voted in 2000		
	(1)	(2)	(3)	(4)	(5)	(6)
Racial diversity index	0.019			0.008		
	(0.006)			(0.007)		
Racial SES Gini	−0.004	−0.003	−0.003	−0.006	−0.006	−0.006
	(0.006)	(0.006)	(0.006)	(0.004)	(0.004)	(0.004)
Share mothers with college degree	0.024	0.037	−0.004	0.130	0.136	0.129
	(0.141)	(0.143)	(0.142)	(0.165)	(0.167)	(0.164)
Ability grouping	0.006	0.002	0.015	0.083	0.081	0.086
	(0.028)	(0.027)	(0.029)	(0.039)	(0.040)	(0.040)
Share black	−0.215	0.101	−0.236	−0.002	0.138	0.001
	(0.256)	(0.262)	(0.681)	(0.235)	(0.264)	(0.713)
Share Hispanic	−0.156	0.482	1.564	0.195	0.478	0.883
	(0.388)	(0.476)	(0.506)	(0.347)	(0.396)	(0.545)
Share Asian	0.281	0.794	1.368	0.717	0.944	1.460
	(0.308)	(0.335)	(0.679)	(0.416)	(0.386)	(0.692)
Share other	−0.382	0.476	0.297	−0.169	0.211	0.257
	(0.497)	(0.370)	(0.849)	(0.532)	(0.380)	(0.710)
Share black squared			0.326			0.127
			(0.792)			(0.784)
Share Hispanic squared			−2.960			−1.152
			(0.744)			(0.741)
Share Asian squared			−1.644			−1.452
			(1.218)			(1.574)
Share other race squared			0.381			−0.931
			(6.897)			(3.310)
Black	0.060	0.061	0.061	0.066	0.066	0.066
	(0.021)	(0.021)	(0.021)	(0.022)	(0.022)	(0.022)
Hispanic	−0.002	−0.002	−0.002	−0.038	−0.038	−0.039
	(0.023)	(0.023)	(0.023)	(0.020)	(0.020)	(0.020)
Asian	−0.109	−0.109	−0.109	−0.156	−0.156	−0.156
	(0.036)	(0.036)	(0.036)	(0.028)	(0.028)	(0.028)
Other	−0.005	−0.010	−0.003	−0.106	−0.108	−0.105
	(0.050)	(0.050)	(0.050)	(0.048)	(0.049)	(0.048)
Constant	−0.483	−0.612	−0.547	−1.033	−1.090	−1.066
	(0.232)	(0.238)	(0.256)	(0.254)	(0.264)	(0.280)
Individual characteristics	✓	✓	✓	✓	✓	✓
Family characteristics	✓	✓	✓	✓	✓	✓
Neighborhood characteristics	✓	✓	✓	✓	✓	✓
N	12,568	12,568	12,568	12,568	12,568	12,568

Notes: This table reports OLS estimates. The racial diversity index and the racial SES Gini index are standardized at the school cohort level, after controlling for school linear trends. Controls include school and grade fixed effects and school linear trends. The omitted category for own race is white. Wave 3 longitudinal weights are used. Standard errors (in parentheses) are clustered at the school level. All controls are listed in Table 1 under individual characteristics, family characteristics, neighborhood characteristics and cohort controls. Variable definitions are in Appendix Table A.1.

Table B.4: Robustness check: Alternative specification, political partisanship

Dependent variable:	Identifies with a party			Identifies as a Democrat		
	(1)	(2)	(3)	(4)	(5)	(6)
Racial diversity index	−0.026			−0.008		
	(0.015)			(0.012)		
Share black	2.764	0.430	1.518	1.007	0.537	0.628
	(0.807)	(0.239)	(0.655)	(0.628)	(0.231)	(0.556)
Share Hispanic	0.631	−0.714	−0.781	−0.348	−0.707	−0.778
	(0.793)	(0.416)	(0.633)	(0.679)	(0.318)	(0.468)
Share Asian	0.976	−0.358	−0.383	0.704	0.217	0.291
	(0.909)	(0.490)	(0.697)	(0.667)	(0.300)	(0.498)
Share other	0.666	−0.185	−0.681	0.423	−0.260	0.013
	(1.013)	(0.397)	(0.685)	(0.926)	(0.268)	(0.523)
Share black squared	−2.994		−1.675	−0.516		−0.115
	(0.963)		(0.771)	(0.788)		(0.709)
Share Hispanic squared	−1.699		0.007	−0.379		0.140
	(1.080)		(0.839)	(1.001)		(0.710)
Share Asian squared	−1.721		0.014	−0.676		−0.147
	(1.617)		(1.286)	(1.140)		(0.912)
Share other race squared	2.534		4.221	−2.632		−2.118
	(4.110)		(4.624)	(2.761)		(2.514)
Racial SES Gini	0.005	0.007	0.005	0.009	0.010	0.009
	(0.005)	(0.005)	(0.005)	(0.003)	(0.003)	(0.003)
Share mothers with college degree	0.153	0.149	0.147	0.002	−0.007	0.000
	(0.176)	(0.185)	(0.180)	(0.129)	(0.129)	(0.131)
Ability grouping	−0.051	−0.039	−0.048	−0.003	−0.003	−0.002
	(0.053)	(0.051)	(0.053)	(0.039)	(0.038)	(0.039)
Black	0.112	0.113	0.112	0.211	0.211	0.211
	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)
Hispanic	−0.032	−0.032	−0.032	0.029	0.029	0.029
	(0.025)	(0.025)	(0.025)	(0.020)	(0.020)	(0.020)
Asian	−0.130	−0.128	−0.129	−0.032	−0.032	−0.032
	(0.032)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)
Other	0.005	0.002	0.005	0.065	0.065	0.065
	(0.056)	(0.056)	(0.056)	(0.040)	(0.040)	(0.040)
Constant	−0.728	−0.481	−0.526	−0.382	−0.305	−0.320
	(0.266)	(0.236)	(0.237)	(0.206)	(0.198)	(0.202)
Individual characteristics	✓	✓	✓	✓	✓	✓
Family characteristics	✓	✓	✓	✓	✓	✓
Neighborhood characteristics	✓	✓	✓	✓	✓	✓
N	12,568	12,568	12,568	12,568	12,568	12,568

Notes: This table reports OLS estimates. The racial diversity index and the racial SES Gini index are standardized at the school cohort level, after controlling for school linear trends. Controls include school and grade fixed effects and school linear trends. The omitted category for own race is white. Wave 3 longitudinal weights are used. Standard errors (in parentheses) are clustered at the school level. All controls are listed in Table 1 under individual characteristics, family characteristics, neighborhood characteristics and cohort controls. Variable definitions are in Appendix Table A.1.

Table B.5: Robustness to attrition and weighting

Dependent variable:	Registered to vote			Voted in 2000		
	(1)	(2)	(3)	(4)	(5)	(6)
Racial diversity index	0.011	0.011	0.011	0.009	0.009	0.008
	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)
Racial SES Gini	−0.002	−0.002	−0.002	−0.009	−0.009	−0.009
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Share mothers with college degree	0.265	0.234	0.233	0.064	0.035	0.035
	(0.124)	(0.127)	(0.129)	(0.151)	(0.146)	(0.147)
Ability grouping	0.012	0.017	0.015	0.087	0.089	0.089
	(0.059)	(0.056)	(0.056)	(0.035)	(0.033)	(0.034)
Black	0.049	0.054	0.043	0.051	0.070	0.051
	(0.019)	(0.019)	(0.021)	(0.021)	(0.020)	(0.022)
Hispanic	−0.085	0.009	0.020	−0.095	−0.023	−0.015
	(0.027)	(0.023)	(0.024)	(0.023)	(0.021)	(0.021)
Asian	−0.176	−0.098	−0.088	−0.173	−0.123	−0.121
	(0.031)	(0.033)	(0.032)	(0.031)	(0.031)	(0.031)
Other	−0.060	−0.014	−0.011	−0.151	−0.112	−0.112
	(0.052)	(0.044)	(0.046)	(0.050)	(0.047)	(0.047)
Constant	−0.245	−0.459	−0.479	−0.595	−0.805	−0.603
	(0.201)	(0.191)	(0.228)	(0.227)	(0.216)	(0.241)
Individual characteristics	✓	✓	✓	✓	✓	✓
Family characteristics		✓	✓		✓	✓
Neighborhood characteristics			✓			✓
N	11,307	11,307	11,307	11,307	11,307	11,307

Notes: This table reports OLS estimates. The racial diversity index and the racial SES Gini index are standardized at the school cohort level, after controlling for school linear trends. Controls include school and grade fixed effects and school linear trends. The omitted category for own race is white. Wave 3 longitudinal weights are used. Standard errors (in parentheses) are clustered at the school level. The sample drops individuals from the estimation sample who have the top 10% highest Wave 3 longitudinal weights. All controls are listed in Table 1 under individual characteristics, family characteristics, neighborhood characteristics and cohort controls. Variable definitions are in Appendix Table A.1.

Table B.6: Behavior in Wave 1

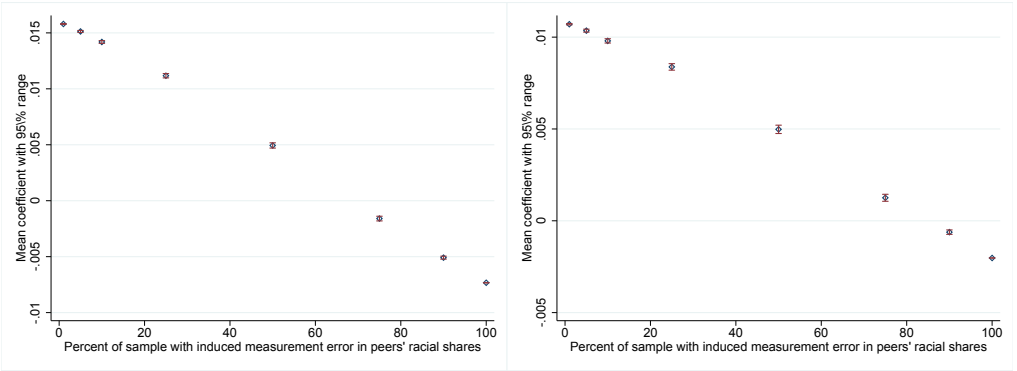
	Smoking	Marijuana use	Binge drinking
Racial diversity index	0.006	0.005	0.010
	(0.006)	(0.003)	(0.004)
Racial SES Gini	0.003	−0.003	−0.008
	(0.005)	(0.003)	(0.004)
Share mothers with college degree	−0.231	−0.240	−0.155
	(0.157)	(0.110)	(0.119)
Ability grouping	0.052	0.029	0.059
	(0.028)	(0.047)	(0.033)
Black	−0.191	−0.005	−0.123
	(0.024)	(0.022)	(0.023)
Hispanic	−0.022	0.036	0.042
	(0.025)	(0.022)	(0.022)
Asian	−0.044	−0.014	−0.077
	(0.028)	(0.028)	(0.026)
Other	−0.079	0.087	−0.091
	(0.038)	(0.045)	(0.029)
Constant	0.599	0.361	0.468
	(0.216)	(0.138)	(0.193)
Individual characteristics	✓	✓	✓
Family characteristics	✓	✓	✓
Neighborhood characteristics	✓	✓	✓
N	12,502	12,391	12,540

Notes: All regressions include both cohort composition variables along with controls for cohort fixed effects, school fixed effects, and school trends, as well as the individual student covariates related to the cohort variables. All dependent variables are measured using Wave 1 of the Add Health. Wave 1 cross-sectional weights are used. Figures in parentheses are standard errors robust to clustering at the school level. All controls are listed in Table 1 under individual characteristics, family characteristics, neighborhood characteristics and cohort controls. Variable definitions are in Appendix Table A.1.

Table B.7: Behavior in Wave 3

	Drop out of high school	College degree	Test score	Idleness	Smoking	Marijuana use	Binge drinking
Racial diversity index	0.000	−0.001	0.166	0.004	0.009	0.004	−0.011
	(0.003)	(0.004)	(0.091)	(0.003)	(0.005)	(0.004)	(0.006)
Racial SES Gini	0.002	−0.003	−0.020	0.000	0.003	0.003	0.013
	(0.003)	(0.004)	(0.097)	(0.004)	(0.005)	(0.005)	(0.006)
Share mothers with college degree	−0.188	0.277	4.085	0.110	0.171	−0.270	0.063
	(0.097)	(0.141)	(3.770)	(0.099)	(0.135)	(0.136)	(0.154)
Ability grouping	0.017	0.007	0.721	0.013	−0.004	0.017	−0.060
	(0.022)	(0.027)	(0.792)	(0.019)	(0.041)	(0.037)	(0.035)
Black	−0.050	0.050	−7.005	0.037	−0.213	−0.026	−0.254
	(0.019)	(0.018)	(0.737)	(0.019)	(0.024)	(0.022)	(0.028)
Hispanic	0.005	0.016	−1.948	0.015	−0.056	0.002	−0.025
	(0.022)	(0.023)	(0.887)	(0.021)	(0.027)	(0.027)	(0.030)
Asian	−0.022	0.086	−2.198	0.027	−0.015	−0.054	−0.154
	(0.024)	(0.029)	(0.855)	(0.028)	(0.036)	(0.033)	(0.035)
Other	0.047	0.061	−0.236	0.042	−0.051	0.038	0.010
	(0.048)	(0.036)	(1.742)	(0.049)	(0.043)	(0.055)	(0.052)
Constant	−1.073	1.211	120.945	0.683	0.560	1.149	0.804
	(0.165)	(0.195)	(5.437)	(0.182)	(0.221)	(0.208)	(0.208)
Individual characteristics	✓	✓	✓	✓	✓	✓	✓
Family characteristics	✓	✓	✓	✓	✓	✓	✓
Neighborhood characteristics	✓	✓	✓	✓	✓	✓	✓
N	12,564	11,818	12,143	12,106	12,514	12,531	12,518

Notes: All regressions include both cohort composition variables along with controls for cohort fixed effects, school fixed effects, and school trends, as well as the individual student covariates related to the cohort variables. All dependent variables are measured using Wave 3 of the Add Health. Test score is the standardized PVT score in Wave 3. Wave 3 longitudinal weights are used. Figures in parentheses are standard errors robust to clustering at the school level. All controls are listed in Table 1 under individual characteristics, family characteristics, neighborhood characteristics and cohort controls. Variable definitions are in Appendix Table A.1.



(a) Dependent variable: Registered to vote (b) Dependent variable: Voted in 2000

Figure B.1: Sensitivity of coefficients to measurement error in race variable

Notes: The y-axis variable is the average coefficient on the racial diversity index from 1,000 regressions where, before each regression, the race variable is replaced with a random value for a share of the sample. This share is indicated on the x-axis. This also affects the values of the racial diversity index and of the SES Gini index.

Appendix C Tests for non-random clustering

Table C.1: Tests for non-random clustering (Guryan et al., 2009)

	White dummy	Black dummy	Hispanic dummy	Asian dummy	Other race dummy
Racial diversity index of peers in grade	−0.095	−0.004	0.045	0.026	0.027
	(0.087)	(0.047)	(0.044)	(0.036)	(0.032)
Racial diversity index of peers in school	172.496	−26.700	−52.317	−31.927	−61.551
	(22.341)	(11.773)	(10.432)	(5.175)	(6.998)
N	79,824	79,824	79,824	79,824	79,824
Adjusted R^2	0.505	0.360	0.326	0.184	0.127

Notes: This table reports OLS estimates. Additional controls are school and grade fixed effects and school linear trends. The regressions reported in this table are run on the respondents to the Wave 1 in-school survey sample who are in cohorts containing at least one student in the estimation sample. Standard errors (in parentheses) are clustered at the school level. The data are unweighted.

Table C.2: Tests for non-random clustering (Caeyers and Fafchamps, 2020)

	Transformed white dummy	Transformed black dummy	Transformed Hispanic dummy	Transformed Asian dummy	Transformed other race dummy
Racial diversity index of peers in grade	−0.089	−0.088	0.002	0.065	0.062
	(0.131)	(0.108)	(0.081)	(0.060)	(0.038)
N	79,824	79,824	79,824	79,824	79,824
Adjusted R^2	0.601	0.574	0.496	0.262	0.043

Notes: This table reports OLS estimates. Additional controls are school and grade fixed effects and school linear trends. The regressions reported in this table are run on the respondents to the Wave 1 in-school survey sample who are in cohorts containing at least one student in the estimation sample. Standard errors (in parentheses) are clustered at the school level. The data are unweighted.