NBER WORKING PAPER SERIES

GIFTED & TALENTED PROGRAMS AND RACIAL SEGREGATION

Owen Thompson

Working Paper 29546
http://www.nber.org/papers/w29546

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2021

Gifted & Talented Programs and Racial Segregation
Owen Thompson
NBER Working Paper No. 29546
December 2021
JEL No. I24,J15

ABSTRACT

Racial segregation can occur across educational programs or classrooms within a given school, and there has been particular concern that gifted & talented programs may reduce integration within schools. This paper evaluates the contribution of gifted & talented education to racial segregation using data on the presence and racial composition of gifted & talented programs at virtually all US elementary schools over a span of nine school years. I first show that, consistent with widespread perceptions, gifted & talented programs do disproportionately enroll white and Asian students while Black, Hispanic and Native American students are underrepresented. However, I also show that accounting for the within-school racial sorting caused by these programs has little or no effect on standard measures of overall racial segregation. This is primarily because gifted & talented programs are a small share of total enrollments and do enroll non-negligible numbers of under-represented minority students. I also estimate changes in race-specific enrollments after schools initiate or discontinue gifted & talented programs, and find no significant enrollment changes after programs are eliminated or initiated. I conclude that gifted & talented education is a quantitatively small contributor to racial segregation in US elementary schools.

Owen Thompson
Department of Economics
Williams College
24 Hopkins Hall Drive
Williamstown, MA 01267
and NBER
ot3@williams.edu

A data appendix is available at http://www.nber.org/data-appendix/w29546

"For years, New York City has essentially maintained two parallel public school systems. A group of selective schools and programs geared towards students labeled gifted and talented is mostly filled with white and Asian children. The rest of the system is open to all students and is predominantly Black and Hispanic." (New York Times, October 8, 2021)

"With Black and White students largely segregated within the schools they attend, racialized tracking has made it possible to have desegregation without integration" (Tyson 2011)

Introduction

Reducing the racial sorting of students across schools has been among the most contentious and impactful areas of educational policy in the US for decades, if not for centuries. The dismantling of racially identifiable schools was the focus of the landmark 1954 *Brown* ruling, was among the central legislative aims of the 1964 Civil Rights Act, and was the stated goal of district-level busing policies in large urban districts nationwide during the 1970s and 1980s, while more recently integration levels have regressed as judicial enforcement and legislative efforts waned. The levels, causes, and consequences of racial segregation across schools is the subject of vast academic literatures within and beyond economics, with interdisciplinary reviews provided by Rivkin & Welch (2006) and Reardon & Owens (2014).

But as the opening epigraphs emphasize, beyond the racial sorting of students across schools there is also widespread concern that educational segregation occurs across classrooms or instructional groups even within racially diverse school buildings. Much of the recent discussion of within-school racial segregation, particularly at the primary school level, has pertained to the impacts of "gifted & talented" (G&T) programs. Public attention to these programs was elevated by the 2021 announcement that New York City Public Schools planned to phase out its G&T programs, with the effects of these programs on racial segregation explicitly cited as a primary motivation for their elimination.

In this paper I use data on the existence and racial composition of G&T programs at virtually every elementary school in the US over a nine year period to provide transparent, systematic evidence on how G&T programs contribute to racial segregation. I specifically investigate the relative racial compositions of G&T versus general education programs; calculate how standard indices of racial segregation change when G&T programs are treated as effectively separate schools; and estimate the dynamic effects of G&T programs on race-specific enrollment patterns using event studies. Perhaps surprisingly, the balance of the evidence I present suggests that G&T programs are a minimal contributor to racial segregation in US elementary schools.

1 Background

There is little question that within-school tracking programs like G&T education have historically been used as an intentional strategy by segregationists to subvert legally required school integration. Many southern districts initiated testing-based classroom assignments in the wake of strong school desegregation enforcement in the 1970s, and many of these programs were successfully challenged in the courts, although the general practice of ability grouping was never ruled unconstitutional per-se (Wright 2013). Contemporary implementations of G&T programming are rarely seen as explicit attempts to resurrect de-jure racial segregation. But racial gaps in tests scores as well as other common features of G&T screening, such as the use of teacher referrals and parental interventions in the admissions process, have the strong potential to result in de-facto racial imbalances in G&T programs and contribute to overall racial segregation.

The actual structure and content of G&T programs are locally determined and highly heterogeneous. While no official data is collected on the basic operations of G&T programs, a 2019 survey of more than 1,200 G&T teachers and coordinators conducted by *Education Week* (Kurtz et al. 2019) provides some systemic information on the common practices of G&T programs.

This survey found that the most common method of delivering G&T instruction was "pull-outs" where G&T students are removed from the mainstream classroom for a portion of instructional time, with 86% of G&T educators reporting any use of pull-outs, compared to 32% reporting any use of self-contained classrooms. Most educators also reported that their districts used multiple overlapping methods of screening students for G&T services, with 79% of districts using non-IQ standardized tests, 71% using teacher referrals, and 66% using IQ-type standardized tests. Previous work suggests that many of these screening processes are subject to some degree of racial bias (Useem 1992; Grissom & Redding 2016). Finally, the most common G&T service was "content enrichment" where deeper coverage of grade-level topics was provided (90% of programs), although "content acceleration" in which students move more quickly to new topics was common as well (78%).

Given its prominence in recent policy discussions, there is surprisingly little academic research directly addressing the effects of G&T programs on racial segregation.

A large body of research, primarily in educational sociology, documents that G&T programs disproportionately enroll white and Asian students and investigates systematic barriers that disadvantaged students face in accessing G&T education (Lewis & Diamond 2015; Roda 2015; Smith-Peterson 2021). This literature primarily consists of informative qualitative case studies, but does not attempt to quantify the overall impacts of G&T education on racial segregation. Another large interdisciplinary literature studies educational tracking, including but not limited to G&T programs (Oakes 2005; Betts 2011; Bui et al. 2014; Card & Giuliano 2016). This literature is primarily focused on tracking's effects on student outcomes, rather than its effects on racial segregation.

The previous work most closely related to the current study includes Conger (2005), Clotfelter (2021), and Francis & Darity (2021), who all use classroom-level data to evaluate the relative importance of within-school versus between-school segregation. I build on these previous studies by studying nearly all US elementary schools, whereas these previous papers all focused on a single district (New York City) or a single state (North Carolina); by studying the longitudinal effects of G&T programs on race-specific enrollment patterns, a potentially important aspect of these program's impacts; and by focusing explicitly on G&T programs, rather than overall classroom sorting, which is a highly relevant policy margin given that school administrators directly choose whether to offer G&T programs.

2 Data

Data is drawn from Civil Rights Data Collection surveys (CRDC), conducted by the Department of Education's Office of Civil Rights. The CRDC is typically conducted biennially, and I use the five CRDC waves covering the 2009-10, 2011-12, 2013-14, 2015-16 and 2017-18 school years, which is the range of years for which the survey's timing and content relating to G&T programs was implemented in a consistent fashion. Below I refer to school years using the calendar year occurring in the spring, with for example "2010" referring to 2009-10 school year.

The CRDC collects information on a wide variety of school characteristics including enrollments, discipline, teacher characteristics, expenditures, and curricular offerings. Relative to other federal school data collection efforts like the Common Core or EdFacts, the CRDC is unique in that it disaggregates most data elements by race and ethnicity (as well as by sex, English proficiency, and disability status). For the present study the key data elements are simply a binary indicator of whether each school operated a G&T program in each year, as well as the race-specific enrollments of both G&T programs (when present) and of the full school.

The CRDC is a *census* of public schools, rather than a sample: Virtually all public schools are required by law to participate and certified data is collected for more than 99% of the nation's schools. In practice the data provided for CRDC by school districts is most commonly drawn directly from district IT systems and batch uploaded, rather than being entered manually by school employees (National Forum on Education Statistics 2018). Since the data is essentially administrative, there is limited scope for intentional manipulation, and any measurement error would primarily be due to inaccuracies within district's internal data collection.

I include public schools offering any grade from K through 6 in all 50 states plus Washington DC. I include charter, magnet and alternative schools but exclude schools that offered only a special education curriculum. Finally, I exclude approximately 6% of schools that were not observed in all five survey years, typically because they were founded, disbanded, or merged over the study period.[1] The resulting data set is a balanced panel of 46,704 public elementary schools observed five times over a span of nine school years.

[1] None of the results below are sensitive to any of these basic sample restrictions.

I calculate race-specific enrollments using the five racial and ethnic categorizations that were collected in a consistent fashion across all survey years: Hispanics of any race, and non-Hispanic Whites, Blacks, Asian/Native Hawaiian/Pacific Islanders, and American Indian/Alaskan Natives.[2] For brevity and because many common segregation measures are defined in terms of two groups, most of the analyses below further aggregate race and ethnicity into (1) white and Asian students and (2) Black, Hispanic and American Indian students, and I refer to the latter group as under-represented minorities (URMs). In addition to expositional and computational convenience, these more aggregated categorizations are often used in policy discussions around segregation and G&T education. In the appendix I reproduce every result in the paper using disaggregated racial and ethnic categories, and none of the key findings are changed.

3 Results

3.1 The Racial Composition of Gifted & Talented Programs

G&T programs are widely perceived to disproportionately enroll white and Asian students relative to Black, Hispanic and American Indian students. I begin by providing systematic national evidence on the whether these often anecdotally motivated perceptions are accurate.

Columns 1-5 of Table 1 report the average share of students in G&T and non-G&T programs who are from each of the five racial and ethnic groups.[3] Columns 1 and 2 show that the average G&T program enrolls 60.1% white students and 8.2% Asian students, whereas in non-G&T programs the analogous shares are substantially lower at 50.9% and 5.1%. Columns 3-5 respectively show that Black students account for 17% of enrollments in non-G&T programs but only 11% of enrollments in G&T programs; that Hispanic students account for 25.5% of non-G&T enrollments and 19.8% of G&T enrollments; and that American Indian students account for 1.4% of non-G&T enrollments and .9% of G&T enrollments.

The enrollment shares reported in Columns 1-5 pooled all schools and school years, such that the racial composition differences by G&T status may partially reflect sorting of students across schools that do and do not offer G&T programs or changes in G&T prevalence and racial compositions over time. The final column of Table 1 accounts for these possibilities by reporting the results of regressing the share of students who are white or Asian onto a G&T dummy and school-by-year fixed effects. The highly statistically significant coefficient of .134 indicates that, on average, the share of white and Asian students in G&T programs is 13.4 percentage points higher than the share of white and Asian students in the non-G&T program offered by

[2] Beginning in 2012, seven race and ethnicity categories were used rather than five, with the additional categories being Native Hawaiian or Other Pacific Islander and Two or More Races. To maintain consistent definitions, I collapse later surveys into the five pre-2012 categories.

[3] Table 1 reports unweighted means, such that each school is given equal weight, but the results are very similar when weighting by total enrollment.

the same school in the same years.[4]

On balance, the results in Table 1 suggest that the widespread perception that G&T programs disproportionately enroll white and Asian students is basically accurate, although URM students are enrolled in G&T programs at non-negligible levels as well.

3.2 Gifted & Talented Programs and Racial Segregation

The generally unrepresentative racial compositions of G&T programs has sometimes been used to argue that standard between-school segregation measures substantially understate the true level of racial segregation, and in turn that eliminating G&T programs could be an effective method of promoting integration. I evaluate this argument by exploiting the availability of race-specific G&T enrollments in the CRDC to calculate how standard segregation indices are changed when G&T programs are treated as fully separate schools.

I specifically calculate two versions of both the Dissimilarity Index and the Exposure Index, the two segregation measures most widely used in the literature (Massey & Denton 1988). I first calculate each index using only between-school variation, and then recalculate each index using variation both between-schools and between G&T versus non-G&T programs within schools, and observe the magnitudes of the changes.

The Dissimilarity Index is defined as:

$$Dissimilarity_d = \frac{1}{2} \sum_s \left| \frac{A_s}{A_d} - \frac{B_s}{B_d} \right|$$

where A_d and B_d denote the number of students in groups A and B within the overall unit (typically a school district), while A_s and B_s denote the number of students in groups A and B within some sub-unit (typically a school). The Dissimilarity Index is a measure of the evenness with which two groups are distributed across sub-units, and can be interpreted as the share of students from Group A (or Group B) who would need to move across sub-units in order to make the racial composition of each sub-unit (most commonly schools) match that of the larger unit (most commonly districts). The Dissimilarity Index therefore ranges from 0 to 1, with larger values indicating greater segregation.

Using the same notation, the Exposure Index (of Group A to Group B) is given by:

$$Exposure_d = \sum_s \frac{A_s}{A_d} \times \frac{B_s}{A_s + B_s}.$$

The Exposure Index is a measure of how intensively one group is exposed to another group, and can be interpreted as the probability among members of Group A that a randomly selected peer will be from group B, such that larger values indicate less segregation (note that this is the opposite of the Dissimilarity Index).

[4] Given the constructions of the aggregated racial and ethnic groups, the enrollment share of URMs is conversely 13.4 percentage points lower in G&T programs than in the non-G&T programs.

The Exposure Index is asymmetric - the exposure of Group A to Group B is not the same as the exposure of Group B to Group A. In the current application I focus on the exposure of URM students to white and Asian students, not the converse. The minimum possible value of the Exposure Index is zero, which would occur in cases where students from Group A had no exposure to students from Group B. Its maximum value is the overall share of the district's students who are from Group B, reflecting the fact that exposure to other groups will be inherently limited in racially homogeneous (e.g. all URM) school districts.

To isolate the influence of G&T programs on racial segregation, I first calculate these two indices using schools as the sub-unit and school districts as the overall unit, which reflects only segregation occurring *between* the schools in a district. I then re-calculate each index using the combination of school and G&T program status as the sub-units. In terms of the notation above, this is simply defining the s subscript as *school* $\times$ *G&T program* rather than only as *school*. When calculated over *school* $\times$ *G&T program* sub-units, the indices reflect both between school segregation and any *within* school segregation that results from G&T programs. The difference between the two measures can be interpreted as the change in segregation that would occur if the G&T programs were discontinued and the students enrolled in them fully returned to the non-G&T programs of their schools, which is a conceptually reasonable measure of how G&T program contribute to overall racial segregation.[5]

Note that this approach effectively assumes that G&T programs are *separate schools*, and observes how this assumption affects the segregation indices. Because many G&T programs operate as "pull outs" and students in the two programs likely have at least some interactions with each other, this approach likely overstates the contribution of G&T programs to segregation, and the resulting estimates can in this sense be viewed as upper bounds.

Table 2 reports average values of the Dissimilarity Index and the Exposure Index for several different sets of districts. In Panel A each district is given equal weight in calculating the average index values, while in Panel B each district is weighted by its total enrollment, giving more influence to larger districts.

Beginning with the unweighted averages in Panel A, Column 1 shows that when calculated using only between-school variation, the Dissimilarity Index in the full sample has a value of .151 (row 1). When the Dissimilarity Index is re-calculated using variation between both schools *and* G&T programs, its value increases to .172 (row 2). The change of .021 (row 3) is the within-school component attributable to racial sorting in G&T programs. In the enrollment-weighted calculations in Panel B, the strictly between-school Dissimilarity Index is .309, while the index value between both schools and G&T programs is .325, such that the contribution of G&T programs within-schools is .016.

Column 2 of Table 2 has the same structure as Column 1 but uses the Exposure Index. In the unweighted calculations in Panel A the Exposure Index falls by .004 after accounting for G&T programs (recall that

[5] The presence of G&T programs may also affect race-specific enrollments over time, a possibility I study directly in the next sub-section.

higher values of the Exposure Index indicate *less* segregation), while in the weighted results in Panel B the within-school component is -.005.

Columns 3 and 4 of Table 2 repeat the same calculations for district-years where at least one school had a G&T program, excluding the approximately 30% of district-years that had no G&T offerings and where G&T policy is arguably less relevant. Columns 5 and 6 report calculations for large, racially diverse district-years, defined as district-years with at least 35,000 total students and URM shares between 10% and 90%, which is intended to focus on contexts with greater scope for racial sorting and to mitigate potential measurement issues when calculating the Dissimilarity Index as discussed by Allen et al. (2015).

In the unweighted results from Panel A, changes in the Dissimilarity Index are somewhat larger among district-years with any G&T offerings than in all district-years (.030 rather than .021) while they are somewhat smaller in large diverse district-years than in all district-years (.013 rather than .021). Changes in the Exposure Index remain very small in both subsets of district-years. Applying enrollment weights in Panel B further reduces the magnitudes of the changes in the Dissimilarity Index, which are .018 in district-years with any G&T programming and .011 in large diverse district-years, while changes in the Exposure Index continue to be uniformly close to zero in the weighted calculations.[6]

On balance, the calculations for the Dissimilarity Index in Table 2 suggest that students of different racial and ethnic groups would be more evenly distributed across education programs in the absence of G&T education, but the magnitudes of these changes could reasonably be described as modest or small, while the calculations for the Exposure Index suggest that eliminating G&T programs would have essentially no impact on the exposure of URM students to white and Asian students.

The over-representation of white and Asian students in G&T programs, as shown in Table 1, may seem incongruent with the lack of substantive changes in segregation measures after accounting for G&T programs, as shown in Table 2. The primary explanation for this contrast is that G&T programs account for a small overall share of enrollments, and that notwithstanding the significant racial imbalances shown in Table 1, G&T programs do enroll substantial numbers of URM students. For instance during the 2017-18 school year, the schools included in the current analysis enrolled a total of 23.6 million students of whom 1.6 million were in G&T programs, a non-negligible but quantitatively modest 6.9% share of overall enrollment. Furthermore, 27.3% of total G&T enrollments consisted of URM students, which is less than their overall enrollment share of 47.4% but is still substantial.

[6] In New York City public schools, the largest school district in the US and a focus of G&T policy debate, similar calculations show an increase in the Dissimilarity Index of .002 and decrease in the Exposure Index of .002.

3.3 The Dynamic Effects of Gifted & Talented Programs on Race-Specific Enrollments

While the calculations in Table 2 suggest that G&T programs contribute very modestly to racial segregation at a point in time, G&T programs may also have dynamic effects on school's racial compositions. For instance the discontinuation of a G&T program may cause white and Asian students to unenroll from a particular school, or conversely the establishment of a G&T program may disproportionate attract students of particular racial and ethic groups.

The longitudinal structure of the CRDC data makes it well-suited for studying enrollment patterns surrounding the introduction or discontinuation of G&T programs, and within-school policy changes were reasonably common over the study period: Of the 46,704 schools in the balanced panel 12,037 either initiated or eliminated a G&T program during the study period.

I use this variation to estimate the effects of G&T programs on race-specific enrollments in an event-study framework, and estimate separate specifications for schools eliminating versus initiating G&T programs.[7] In the models estimating the dynamic effects of *eliminating* G&T programs, I use the sample of all schools that had G&T programs during the first school year available in the data, 2010. There are 34,513 such schools, of which 7,686 eliminated their G&T program at a subsequent point during the study period and 26,827 maintained a G&T program throughout the study period. The schools that eliminated their programs comprise the "treated units" while I also retain the schools that maintained their programs throughout the study period and use them as "never-treated' controls. Conversely, in the models that estimate the impacts of *initiating* G&T programs, I retain the 12,191 schools that did not have a G&T program in the base year, of which 4,351 added a program later in the study period and the remaining 7,480 serve as never-treated controls.[8]

Given that data is available for a relatively narrow window of five observations spanning nine school years, and that events are spread fairly evenly across the sample years, estimating a standard event-study specification using a panel that is balanced in event-time is quite restrictive. For instance to estimate event-study models that include two pre-treatment observations and two post-treatment observations on a balanced panel would require restricting the analysis to events occurring in 2014, eliminating a large majority of the events. Conversely, to include events occurring in the middle three years of the sample (2012, 2014 and 2016) would only allow one pre-treatment period and one post-treatment period to be evaluated.[9]

As an alternative, I divide the treated schools in my sample into four distinct "treatment cohorts" based on the year in which they eliminated (initiated) a G&T program. I then estimate and report four separate

[7] The effects of eliminating versus initiating a G&T program are clearly related, but are also distinct, and there is no ex-ante reason to assume that the effects of these two policy changes are equal and opposite.

[8] In the rare instances where a school initiates and then discontinues a G&T program all within the study period (or vice-versa), I study only the first policy change and treat it as an "absorbing" state.

[9] This is because events occurring in 2012 have only one observable pre-period (2010) while events occurring in 2016 only have one observable post-period (2018).

event-study specifications, one for each treatment cohort, where in each of the four specifications I use never-treated units as controls and include the full range of event-times that are estimable for that treatment cohort.

For instance, the earliest observed events occur in 2012, and since each school is observed every two years from 2010-2018, the 2012 treatment cohort has estimable coefficients for event times -2, 0, 2, 4 and 6. I therefore use a sample that contains schools with 2012 events as well as never-treated schools to estimate the following specification:

$$Enrollment_{sy} = \sum_{k=-2}^{k=6} \gamma_k I(year - event\ year_s = k) + \delta_s + \eta_{dy} + \varepsilon_{sy}$$

where the dependent variable is a race-specific enrollment measure for school s in year y; δ_s and η_{dy} are respectively school and district-by-year fixed effects; and γ_k are the coefficients for indicators of whether a G&T program was eliminated or (initiated) k years ago, or event time. The observation immediately prior the event (event time -2) is omitted and serves as the reference category, and event time is set to -2 for all untreated units. The school fixed effects in this specification will account for all time-invariant school characteristics affecting enrollment patterns, while the district-by-year fixed effects will account for all time-varying enrollment determinants that are general to the district containing the individual school.

I estimate similar specifications for the treatment cohorts of 2014, 2016 and 2018, with the only difference being that these other treatment cohorts have different estimable ranges of event time. Specifically for the 2014 treatment cohort event times -4 through 4 are estimable, for the 2016 treatment cohort event times -6 through 2 are estimable, and for the 2018 treatment cohort event times -8 through 0 are estimable. I report event study diagrams for each of the four treatment cohorts, and taken together these diagrams provide a "moving window" that characterizes all estimable enrollment trends in the years surrounding the elimination or initiation of G&T programs over a nine year period.

In addition to characterizing enrollment trends for as many schools and event times as possible, the described approach addresses concerns related to two-way fixed effects specifications in settings where treatment timing varies across units, which is the topic of an influential recent methodological literature (Callaway & Sant'Anna 2021; Goodman-Bacon 2021; Sun & Abraham 2021). A key issue raised by this literature is that two-way fixed effects estimates with variable treatment timing are partially based on comparisons of later-treated units to earlier-treated units and vice-versa. A particular concern is that if early-treated units follow differential trends that continue into later periods due to the treatment itself, those trends will inaccurately characterize the counterfactual outcomes of later-treated units.

The specifications I estimate here are effectively an implementation of the procedure suggested by Sun & Abraham (2021) to address these issues in event-study settings. In particular Sun & Abraham (2021) suggest

that researchers estimate separate event-study coefficients for each treatment cohort using never-treated units as the controls when available, as is done here. The authors refer to the event time coefficients for each treatment cohort as "cohort-specific average treatment effects on the treated" or CATTs, and their full procedure then takes a weighted average of the cohort-specific event time coefficients. Here I simply report all estimable CATTs, rather than their weighted average, which allows me to estimate a broader range of event time coefficients. Intuitively, my adapted Sun & Abraham (2021) approach circumvents potential issues from using schools treated in one period as controls for schools treated in another period because the event-study coefficients for each treatment cohort are based solely on comparisons of the schools treated in that year to untreated schools, and not on any comparisons of early-treated schools and later-treated schools.

Figure 1 reports results for the elimination of G&T programs, while Figure 2 reports results for the initiation of G&T programs. The dependent variable in both figures is the share of the student population that is white or Asian, since the possibility that G&T programs disproportionately attract white and Asian families has been a focus of policy attention, but note that given the utilized racial and ethnic groupings the share of URM student shares are identical to those shown in Figures 1 and 2 but with the signs reversed.

Overall, the patterns in Figures 1 and 2 do not provide any consistent evidence for a causal effect of G&T programs on school's race-specific enrollments. None of the eight cohort specific event studies shown in the figures display a visually apparent trend break in white and Asian enrollment after the elimination or initiation of a G&T program, only one of the 20 coefficients for post-treatment event times is statistically significant at the 5% level, and relative to a baseline white and Asian enrollment share in pre-treatment periods of approximately .50, the coefficient magnitudes of less than .005 (one half of one percentage point) are qualitatively small. There are also no indications from the pre-trends in Figures 1 and 2 that G&T programs are initiated or discontinued in *response* to changing racial compositions, for instance that schools with increasing shares of URM students are more likely to create G&T programs to retain or attract white and Asian families.[10]

4 Conclusion

On balance, the findings reported above indicate that G&T programs are a small or negligible contributor to racial segregation in US elementary schools. Eliminating all G&T programs nationally would have a minimal estimated impact on standard measures of racial segregation, and the presence of a G&T program does not appear to causally impact the racial composition of enrollments over time.

One caveat to these findings is that G&T education is primarily a feature of elementary schools in the US, and previous research has found that within-school segregation is less extensive in primary schools than in secondary schools (Betts 2011). An analysis of tracking and racial segregation at the high school level, for

[10] In the appendix I also show estimates for *total* enrollments, which indicate no dynamic effects from starting a G&T program, and some suggestive evidence of modest total enrollment declines from ending a G&T program.

instance in Advanced Placement or International Baccalaureate classes, might yield different conclusions. Another caveat is that there may be subtle interactions between racial segregation between and within schools. For instance Clotfelter et al. (2021) report descriptive evidence that North Carolina school districts with less segregation between schools also have more segregation within schools, and further research on this and related patterns would be valuable.

These caveats notwithstanding, the current study's findings suggest that any impacts of G&T programming on racial segregation at the elementary school level are likely to be minimal. Given this, the question of whether a given G&T policy is desirable depends primarily on whether G&T programs provide benefits to the students who enroll in them or have deleterious impacts on the students who remain in the general education population. The effects of G&T programs on academic outcomes is the topic of a large literature, but currently lacks consensus findings (see for example Rees et al. 2000; Betts & Shkolnik 2000; Bui et al. 2014 and Card & Giuliano 2016, as well as the review by Betts 2011). While the analyses reported here do not directly prescribe whether G&T programs are a desirable overall education policy, they do indicate that the effects of existing G&T programs on racial segregation should not be a first-order policy consideration.

References

Allen, R., et al. (2015). More reliable inference for the dissimilarity index of segregation. *The Econometrics Journal*, 18(1), 40-66.

Betts, J. (2011). The economics of tracking in education. In *Handbook of the Economics of Education* (Vol. 3, pp. 341-381): Elsevier.

Betts, J. & Shkolnik, J. (2000). Key difficulties in identifying the effects of ability grouping on student achievement. *Economics of Education Review*, 19(1), 21-26.

Bui, S. Craig, S. & Imberman, S. (2014). Is gifted education a bright idea? Assessing the impact of gifted and talented programs on students. *American Economic Journal: Economic Policy*, 6(3), 30-62.

Callaway, B., & Sant'Anna, P. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2), 200-230.

Card, D., & Giuliano, L. (2016). Can tracking raise the test scores of high-ability minority students? *American Economic Review*, 106(10), 2783-2816.

Clotfelter, C. et al. (2021). School segregation at the classroom level in a Southern 'New Destination' state. *Race and Social Problems*, 13(2), 131-160.

Conger, D. (2005). Within-school segregation in an urban school district. *Educational Evaluation and Policy Analysis*, 27(3), 225-244.

Francis, D. & Darity, W. (2021). Separate and unequal under one roof: How the legacy of racialized tracking perpetuates within-school segregation. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, 7(1), 187-202.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*.

Grissom, J. & Redding, C. (2015). Discretion and disproportionality: Explaining the underrepresentation of high-achieving students of color in gifted programs. *AERA Open*, 2(1), 2332858415622175.

Kurtz, H. et al. (2019). Gifted Education: Results of a National Survey. Editorial Projects in Education.

Lewis, A. & Diamond, J. (2015). *Despite the best intentions: How racial inequality thrives in good schools*. Oxford University Press.

Massey, D., & Denton, N. (1988). The dimensions of residential segregation. *Social Forces*, 67(2), 281-315.

National Forum on Education Statistics (2018). Forum Guide to Reporting Civil Rights Data, (NFES 2017-168). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Oakes, J. (2005). *Keeping track: How schools structure inequality*. Yale University Press.

Reardon, S., & Owens, A. (2014). 60 years after Brown: Trends and consequences of school segregation. *Annual Review of Sociology*, 40, 199-218.

Rees, D., Brewer, D., & Argys, L. (2000). How should we measure the effect of ability grouping on student performance? *Economics of Education Review*, 19(1), 17-20.

Rivkin, S., & Welch, F. (2006). Has school desegregation improved academic and economic outcomes for blacks? *Handbook of the Economics of Education*, 2, 1019-1049.

Roda, A. (2015). *Inequality in gifted and talented programs: Parental choices about status, school opportunity, and second-generation segregation*. Springer.

Smith-Peterson, M. (2021). Excavating Histories of the Present in" GT". Doctoral Dissertation, University of Minnesota.

Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2), 175-199.

Tyson, K. (2011). *Integration interrupted: Tracking, Black students, and acting White after Brown*. Oxford University Press.

Useem, E. L. (1992). Middle schools and math groups: Parents' involvement in children's placement. *Sociology of Education*, 263-279.

Wright, G. (2013). *Sharing the Prize*. Harvard University Press.

Table 1: Enrollment Shares by Gifted & Talented Classification

	(1) White	(2) Asian	(3) Black	(4) Hispanic	(5) American Indian	(6) White or Asian
Gifted & Talented	60.1%	8.2%	11.0%	19.8%	0.9%	
Non Gifted & Talented	50.9%	5.1%	17.0%	25.5%	1.4%	
Gifted & Talented Indicator						0.134***
						(0.001)
Constant						0.556***
						(0.000)

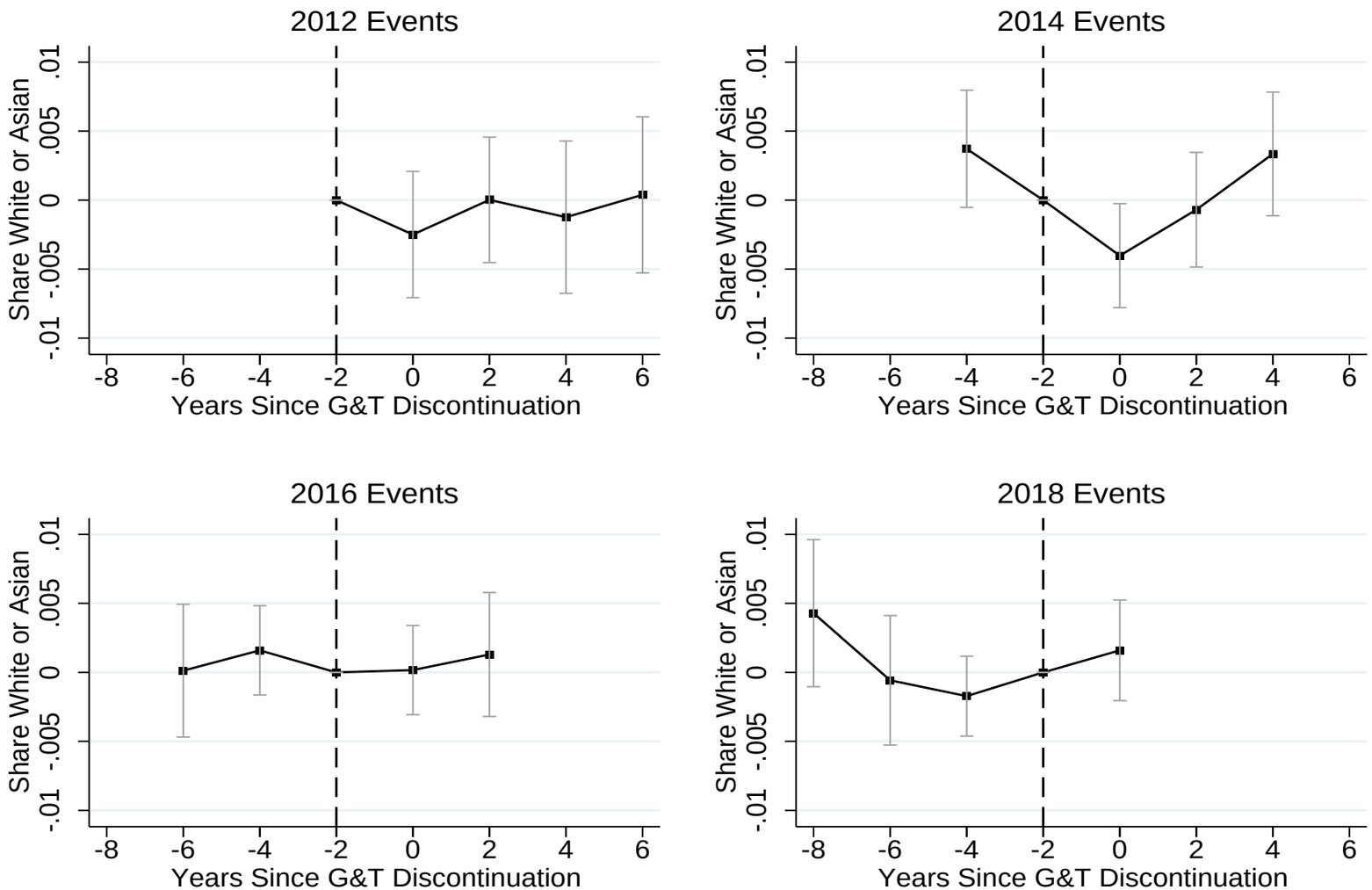
Notes: Columns 1-5 report the average share of students in G&T and non-G&T programs who are from the indicated racial and ethnic group, pooling all schools and school years. Column 6 reports the results of regressing the share of students who are white or Asian onto a G&T indicator and school-by-year fixed effects, such that the reported coefficient estimates the average share of white and Asian students in G&T programs relative to the non-G&T programs in the same school and same year. All schools are given equal weight. For the regression in Column 6 standard errors are clustered at the school level and *, ** and *** indicate statistical significant at the 10%, 5% and 1% levels, respectively.

Table 2: Gifted & Talented Programs and Segregation Indices

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Unweighted	All Districts		Districts with G&T Programs		Large Diverse Districts	
	Dissimilarity Index	Exposure Index	Dissimilarity Index	Exposure Index	Dissimilarity Index	Exposure Index
Segregation Between Schools	0.151	0.671	0.171	0.649	0.439	0.335
Segregation Between Schools *and* G&T Programs	0.172	0.667	0.201	0.643	0.452	0.329
Difference (within-school segregation from G&T)	0.021	-0.004	0.030	-0.006	0.013	-0.006
Observations (district-years)	30,229	30,229	21,353	21,353	287	287
Panel B: Enrollment Weighted	All Districts		Districts with G&T Programs		Large Diverse Districts	
	Dissimilarity Index	Exposure Index	Dissimilarity Index	Exposure Index	Dissimilarity Index	Exposure Index
Segregation Between Schools	0.309	0.488	0.323	0.467	0.493	0.279
Segregation Between Schools *and* G&T Programs	0.325	0.483	0.342	0.461	0.503	0.273
Change (within-school component)	0.016	-0.005	0.018	-0.006	0.011	-0.005
Observations (district-years)	30,229	30,229	21,353	21,353	287	287

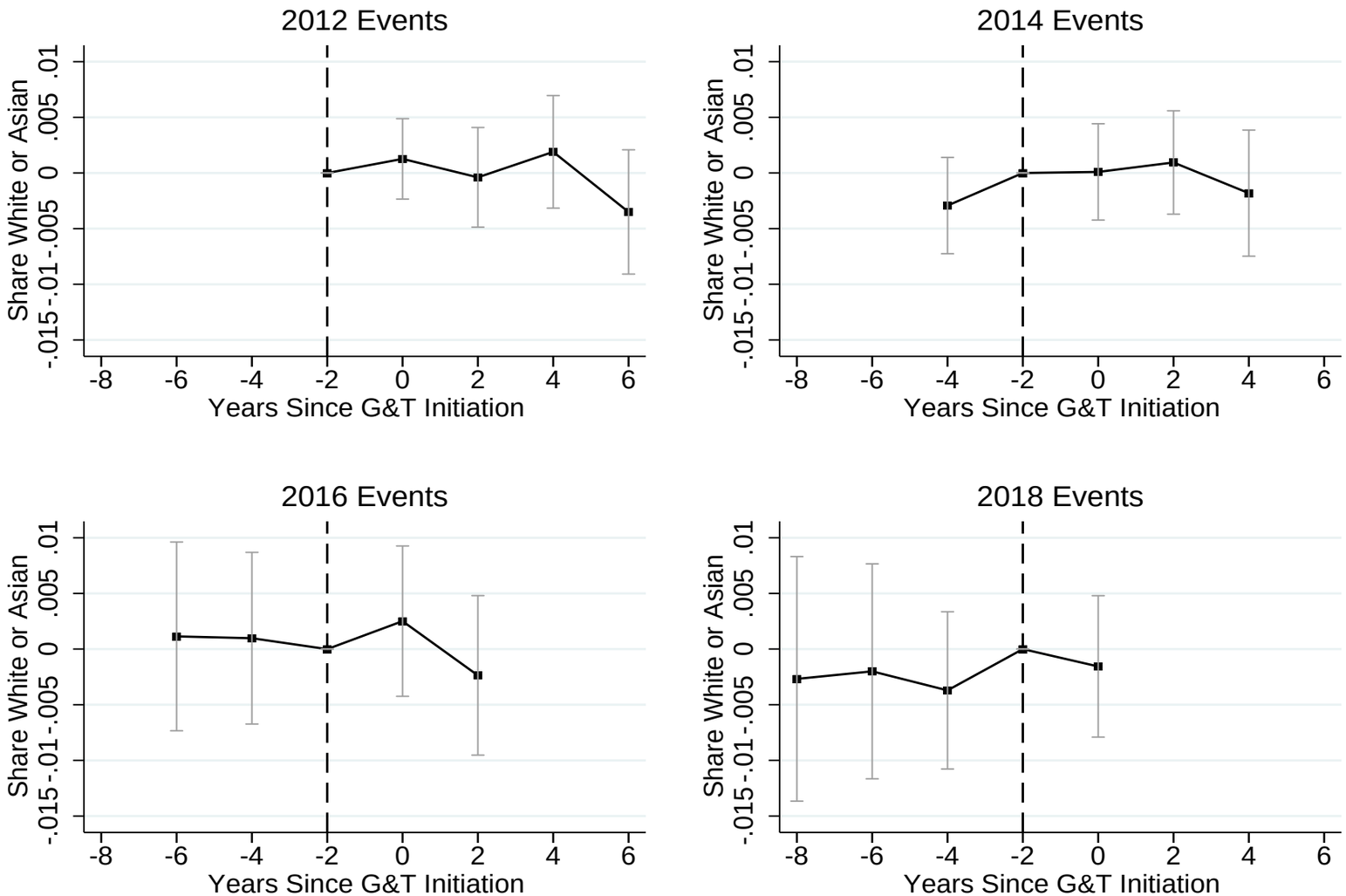
Notes: Table entries report mean values of the indicated segregation index within the indicated set of school districts. The reported between school figures use indices that sum over all of the schools in each district, while the reported between schools and G&T program figures use indices that sum over all school-by-G&T program combinations in each district. See Section 3.2 of the text for a detailed discussion. The mean values in Panel A give each school district equal weight, while the mean values in Panel B weight by total enrollment. The sample in Columns 1 and 2 includes all district-years, the sample in Columns 3 and 4 includes district-years with at least one school that had a G&T program, and the sample in Columns 5 and 6 includes district-years where there were at least 35,000 total students and where the share of under-represented minority students was between 10% and 90%.

Figure 1: Dynamic Enrollment Effects of Discontinuing Gifted & Talented Programs



Notes: Figures plot the coefficients on event time indicators when the share of white or Asian students is regressed onto event time, school fixed effects, and year by district fixed effects. The sample in each figure consists of schools that discontinued a G&T program in the indicated year and "never-treated" schools that maintained a G&T program throughout the study period. Event time -2, the last observation prior to G&T program discontinuation, serves as the reference category. Each school is given equal weight. Vertical bands indicate 95% confidence intervals constructed using standard errors clustered at the school level.

Figure 2: Dynamic Enrollment Effects of Initiating Gifted & Talented Programs



Notes: Figures plot the coefficients on event time indicators when the share of white or Asian students is regressed onto event time, school fixed effects, and year by district fixed effects. The sample in each figure consists of schools that initiated a G&T program in the indicated year and "never-treated" schools that never had a G&T program during the study period. Event time -2, the last observation prior to G&T program initiation, serves as the reference category. Each school is given equal weight. Vertical bands indicate 95% confidence intervals constructed using standard errors clustered at the school level.