Res High Educ
https://doi.org/10.1007/s11162-018-9493-3

Racial and Ethnic Heterogeneity in the Effect of MESA on AP STEM Coursework and College STEM Major Aspirations

Steven Elías Alvarado[1] · **Paul Muniz**[1]

Received: 30 January 2017
© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract Previous research suggests that racial and ethnic disparities in postsecondary STEM outcomes are rooted much earlier in the educational pipeline. One possible remedy to these disparities is participation in early STEM enrichment programs. We examine the impact of MESA, which is an early program that targets socioeconomically disadvantaged students, on outcomes that may lead students down the path to STEM. We analyze three waves of restricted nationally-representative data from the High School Longitudinal Study that trace the STEM progress of more than 25,000 students throughout high school and into their postsecondary careers. Propensity score matching models reveal that MESA participation increases students' odds of taking AP STEM courses in high school and their aspirations for declaring a STEM major in college. However, these effects are driven primarily by black and white students, respectively. Latino and Asian students remain largely unaffected. A formal sensitivity analysis concludes that these findings are moderately robust to unobserved confounding. The results are also robust to alternative matching schemes. Collectively, the findings suggest that MESA may improve black students' high school STEM engagement but may have little impact on black and Latino students' STEM outcomes in college.

Keywords STEM · AP coursework · STEM aspirations/expectations · Propensity score matching

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11162-018-9493-3) contains supplementary material, which is available to authorized users.

✉ Steven Elías Alvarado
alvarado@cornell.edu

[1] Department of Sociology, Cornell University, 336 Uris Hall, Ithaca, NY 14853, USA

Published online: 20 January 2018

Introduction

Educators, researchers, and policy makers have stressed the importance of enhancing students' skills and persistence in science, technology, engineering, and mathematics (STEM) fields. Meanwhile, black and Latino secondary and postsecondary students are still achieving and persisting in these subject areas at much lower rates compared to their white and Asian counterparts (Adams 2014; Catsambis 1995; Dalton et al. 2007; Muller et al. 2001; NSF 2014; Press 2013; Provasnik et al. 2012; Simpson and Oliver 1990; Zhou 2005). Interest in these racial and ethnic disparities has driven much research at the postsecondary level. However, students' commitment to STEM tends to solidify much earlier in their educational careers (Tai et al. 2006) and prior work (e.g., Morgan 1996) argues that the approaching end of high school encourages students to start thinking carefully about their postsecondary STEM plans (Ormerod and Duckworth 1975; Royal Society 2006). Accordingly, this paper examines whether the antecedents to postsecondary STEM disparities might be observed and potentially addressed during high school through an analysis of exposure to a STEM-focused enrichment program.

Data suggests that incoming black and Latino high school students are just as likely as their white and Asian peers to state that they expect to be in a STEM occupation by age 30, net of sociodemographic factors (authors' calculations of High School Longitudinal Study data). This basic finding highlights an overarching question for research in this area: *what can be done to attenuate the underrepresentation of black and Latino students in postsecondary STEM fields?* Previous research findings suggest that interventions aiming to circumvent future disparities might be most beneficial when implemented as early as possible along the STEM pipeline. However, research into the tangible outcomes associated with such programs is limited (Leggon and Pearson 2006; Stake and Mares 2001; Wai et al. 2010).

STEM enrichment programs (SEPs) that provide educational and motivational resources to underrepresented students are promising, yet under-investigated. These programs are important because they offer guidance, mentorship, enrichment activities, and other resources that can nurture the STEM aspirations of students from underrepresented backgrounds and enhance teachers' abilities to promote STEM engagement (Kelly and Zhang 2016; Oakes 1990). However, estimates of the actual impact of SEPs on black and Latino students' educational attainment and STEM participation are rare (Ormerod and Duckworth 1975; Stake and Mares 2001; Tai et al. 2006; Xie and Killewald 2012). Our study contributes to the literature on STEM education by examining the impact of participation in an SEP in high school on two key outcomes: (1) taking any advanced placement (AP) STEM class in high school and (2) aspiring or expecting to declare a STEM major at the outset of college.[1]

We focus our analysis on one particular SEP—Mathematics Engineering Science Achievement (MESA). MESA utilizes a combination of enrichment activities, academic support, and industry involvement to aid first-generation, low income, and socioeconomically disadvantaged students on their path toward STEM fields in college and in the labor market. MESA "engages thousands of educationally disadvantaged students so they excel in math and science and graduate [college] with math-based degrees" (http://mesausa.org).

[1] Although the outcomes under study here are likely to invoke an image of a college-ready student, these analyses are not systematically conditioned on college-readiness or any other indicator that would limit the analysis only to students who are going to enroll (or are enrolled) in postsecondary institutions. We assess the impact of MESA on STEM outcomes for all students.

The program not only aims to better prepare disadvantaged students by encouraging and assisting their participation in STEM prior to college, but it also encourages and guides them to pursue STEM programs once they get to college.[2] MESA is an ideal candidate for analysis because of its longstanding commitment to the development of programs that focus specifically on STEM and to the explicit goal of placing disadvantaged students on track for STEM degree completion. By using restricted longitudinal data that, to our knowledge, have never been used to study enrichment programs and STEM outcomes, this study contributes a fresh understanding of the potential for an SEP to effectively address postsecondary and labor market inequalities in STEM.

Background

Racial and Ethnic Disparities in STEM Participation

The transition to an increasingly technical U.S. economy has been paralleled by the persistence of a racial and ethnic gap in STEM achievement and participation that consistently disfavors black and Latino students (Dalton et al. 2007; Davidson 2012). Racial and ethnic disparities in science and math performance emerge during childhood and worsen over time as students advance through their educational careers (Wai et al. 2010). By the tenth grade, black and Latino students are more likely than their peers to filter into low educational tracks and less likely to enroll in STEM courses (Dalton et al. 2007; Wigfield and Eccles 2000). By the twelfth grade, the expected science and math competency of black and Latino students is comparable to that of white or Asian middle school students (Muller et al. 2001). Structural and social psychological factors constitute meaningful drivers of these and related disparities and include between and within school segregation (Coleman et al. 1966; Lucas 1999; Orfield et al. 2014), which could impact the number and kind of STEM resources to which students are exposed. Further, the positive association between the magnitude of stereotype threat and level of course difficulty could impact disparities (Correll 2004; Owens and Massey 2011; Steele and Aronson 1995) where socially conditioned ideas about academic inferiority limit STEM achievement and attainment for black and Latino students. Black–white disparities in mathematics course taking, for example, are most severe in integrated high schools where black students are in the minority (Kelly 2009) and are primarily explained by lower levels of achievement and less rigorous course taking prior to entering high school.

Such disparities persist at the postsecondary level, where black and Latino students continue to be significantly underrepresented among graduates receiving baccalaureate and advanced STEM degrees (Chen 2009; Dalton et al. 2007; Lewis et al. 2009; Muller et al. 2001; Provasnik et al. 2012; Wai et al. 2010). While black and Latino students are similarly likely as their peers to select STEM majors at the outset of college, white and Asian students are considerably more likely to finish them: about 44% of white students and 40% of Asian students who initially declare a STEM major complete their desired programs, compared to only 32 and 33% of black and Latino students, respectively (Chen 2009). The

[2] The MESA program originated in California in 1970 and has since expanded to ten states that have integrated it into their set of services for disadvantaged students. The program operates in 140 elementary schools, 470 middle schools, 359 high schools, 37 two-year institutions, and 28 four-year institutions. These MESA participation statistics are current as of 2015 and, therefore, may not necessarily represent participation at the time when HSLS respondents participated in MESA.

disparity in STEM baccalaureate degree attainment is notable because of the impacts of a STEM degree on other inequality outcomes. For example, according to research, STEM graduates earn $800,000 more over their lifetimes compared to social science graduates (Kim et al. 2015). At the Ph.D. level, black and Latino students represent less than 10% of STEM doctoral recipients (Yoder 2012), despite representing 11.3% of all doctoral students (Aud et al. 2012). Though black and Latino students currently represent a higher proportion of new degree awardees than in past years, this increase is marginal and partly due to a demographic shift where they are increasingly represented in population cohorts reaching the age where such degrees are typically awarded (Leggon and Pearson 2006).

At the occupational level, black and Latino employees comprise just 5 and 6% of STEM workers respectively, relative to the roughly 12 and 16% that they represent in the larger labor force population (Bureau of Labor Statistics 2013; NSF and NCSES 2013). Additionally, underrepresented minorities who overcome the odds and obtain STEM jobs earn less on average than their similarly educated white and Asian coworkers, net of educational attainment (Riegle-Crumb et al. 2006; Yonezawa et al. 2002). While black and Latino professionals earn less on average than their white and Asian counterparts, they are likely still benefiting from a wage and prestige premium compared to blacks and Latinos outside of STEM. Other things equal, these premiums make STEM occupations more attractive than non-STEM occupations. The hiring of blacks and Latinos into STEM occupations may also have positive externalities such as dispelling lingering doubt as to their competence that may indirectly lead to increases in hiring over time. Despite these benefits, however, the overarching trends highlight the paucity of participation in STEM by blacks and Latinos that begins in the early stages of the life course and compounds over time to produce disparities in postsecondary and labor market outcomes.

STEM Enrichment Programs, AP STEM Classes, and College Major Selection

Theoretically, programs targeting black and Latino students early on to enhance their STEM participation may stem the tide of racial and ethnic disparities in this important segment of the labor market. To frame our study, we borrow from Heckman (2006), Summers and Hrabowski (2006), and others to argue that the early investment of educational resources could make inroads in closing racial and ethnic STEM achievement and attainment gaps. Creating the social and academic environment where students feel supported and where they receive resources to guide them on their path through secondary and postsecondary STEM education is one of the ways that SEP participation could increase STEM success for black and Latino students. Yet, although some SEPs have been operating for decades, the absence of comprehensive data that tracks the educational outcomes of students who have participated in SEPs makes it difficult to assess whether these programs are positively influencing the STEM trajectories of black and Latino students (Stake and Mares 2001). While scholars have established a connection between SEPs and desirable learning outcomes, there is no clear link between these enrichment programs and an increased likelihood of black and Latino students to pursue advanced STEM education in college. Enrichment programs that motivate and guide underrepresented minority students over long periods of time throughout middle and high school might help to reduce racial and ethnic disparities in STEM and prepare black and Latino students to take STEM courses in high school.

Leading scholars (Heckman 2006; Summers and Hrabowski 2006) have supported early educational intervention programs in general as well as those that focus on STEM

specifically as tools to address racial and ethnic disparities in STEM education. Nevertheless, comprehensive program evaluations are lacking, especially with regard to the outcomes of interest in this study. No research has thoroughly explored the role of SEP participation on AP STEM course taking or on encouraging (or preparing) students to major in STEM at the postsecondary level (Hoepner 2010). However, research suggests that black and Latino students continue to face disadvantages that affect their performance in the most advanced math classes in high school (Riegle-Crumb and Grodsky 2010). It may be likely that SEP participation could lead to a closing of the achievement gap in AP and other advanced STEM courses in high school. Enhancing black and Latino students' success in AP STEM courses could not only provide them with the training necessary for college STEM courses, but could also enhance their confidence and self-efficacy that could last beyond college and follow them into the labor force. Similarly, SEP participation may directly or indirectly enhance students' aspirations to major in STEM in college. Most of the assessments that we do have are cross-sectional and based on self-reported data, indicating the need for longitudinal analyses and examinations of the effects of SEP participation on educational outcomes (Jackson 2003; Stake and Mares 2001). Exceptions do exist, however, and include both qualitative and quantitative studies that are largely concerned with the bidirectional relationship between SEP participation and science and math identity and self-efficacy (Afterschool Alliance 2011; Wang 2013). The link between SEP participation and downstream academic outcomes, however, has yet to be sufficiently explored.

Research shows that black and Latino students' likelihood to expect to pursue STEM careers at the primary and secondary levels of education is similar to that of their peers when controlling for differences in access to educational resources (Oakes 1990; Rakow and Walker 1985). Further, black and Latino college freshmen are *more likely* than white students to select STEM degree programs net of prior achievement (Staniec 2004). A potential weak link in this line of thinking, however, is our inadequate understanding of whether the specific types of knowledge and exposure gained through SEP participation translate to stronger STEM orientations for black and Latino students, which is an essential component for justifying SEP implementation given research demonstrating that prior self-concepts such as science identity and self-efficacy can fully mediate the relationship between SEP participation and future commitment to related fields (Chemers et al. 2011). MESA and other SEPs encourage commitment to STEM education, but the question remains to what extent SEPs have been effective in engaging underrepresented groups.

Research has established that, on average, black and Latino students experience systematically lower curricular tracking levels, lower expected math and science competencies toward the end of high school, and higher attrition rates from STEM degree programs at the collegiate level. Given this context, two potentially meaningful and empirically measurable outcomes associated with students' continued SEP participation are the likelihoods for students to enroll in high-level STEM courses in high school and to expect or aspire to declare a STEM major net of program participation. MESA's outcome-oriented, longitudinal approach to mitigating representation gaps in college and beyond makes it an ideal candidate for such an exploration.

How MESA Works

The central goal of MESA is to encourage students from disadvantaged backgrounds to pursue educational pathways that lead to careers in science, engineering, and other high-skilled, math-based fields (http://mesausa.org/). To achieve these ends, MESA has

established various initiatives that assist students as early as elementary school and as late as during college. The specific MESA initiative that targets students before and during high school is the MESA Schools Program (MSP). MSP provides support to students—primarily from disadvantaged backgrounds—in order to enhance their science and math abilities and to bolster college competitiveness. MSP partners with teachers, school administrators, district officials, schools, and industry professionals to provide students with meaningful STEM enrichment opportunities.

Participating students are selected through a collaborative process involving teachers at participating schools and MESA representatives. Once selected, participating students gain access to individualized and college-focused academic planning tools, study support, science and math competitions with other MESA students on the local to national levels, college and career planning assistance, and options for in-school or out-of-school school class periods that focus specifically on MESA projects. Further, MESA provides teachers from participating schools with career development workshops that offer hands on training in science and math education. Collectively, the components of MSP operate across several levels and are unified by the goal of bolstering STEM learning outcomes and persistence among disadvantaged students.

Predictors of MESA participation include a host of characteristics that are tied to educational investments. These variables could include student characteristics such as interest in STEM, high achievement in math and science, planning for the future; family characteristics such as family income, parents' education, parents' STEM education; and high school characteristics such as school socioeconomic status, opportunities for STEM course participation, STEM climate, and partnerships with SEP programs. These could all possibly impact students' participation in SEPs by shaping their awareness of and interest in supplementary educational programs that could lead to enhanced STEM outcomes.

MESA, Academic Performance, and STEM Major Aspirations

Prior work, while scarce, does show that SEP participation increases students' confidence in their STEM abilities (Afterschool Alliance 2011; Stake and Mares 2001). Likewise, the presence of numerous policies and programs to increase students' curiosity in STEM demonstrates that their impact has been recognized (Bottia et al. 2015). The potential for MESA to carry out this function is vital, as black and Latino students must navigate through the educational pipeline while carrying the weight of negative societal pressures that suggest their academic inferiority (Owens and Massey 2011; Steele and Aronson 1995). In addition to providing essential educational resources and boosting the STEM-specific confidence of students from disadvantaged backgrounds, participation in MESA might also reduce or reverse the impact of these negative stereotypes, encourage continued STEM course taking, and propel these students into STEM majors in college at higher rates.

On the surface, the implementation of SEPs seems like a reasonable response to racial and ethnic gaps in STEM achievement and representation. Although the combination of a strong research base and the implementation of informed policy measures has reduced gender gaps in STEM fields over the last quarter century (Muller et al. 2001), similar efforts have not had as much success in reducing racial and ethnic gaps. We find it reasonable to assume that with the development of a stronger understanding of the impact that MESA may have on racial and ethnic disparities and the execution of strategic policy changes, racial and ethnic inequalities might be similarly reduced in the classroom and beyond.

Especially poignant is reaching black and Latino students early in the educational pipeline so that interest in STEM can be nurtured and developed (Wang 2013). Indeed, interest or confidence in STEM as early as middle school can have an impact on students' postsecondary STEM success (Hinojosa et al. 2016; Moakler and Kim 2014). Therefore, early intervention programs that address racial and ethnic disparities in STEM are critical (Summers and Hrabowski 2006). However, most evaluations of such programs are cross-sectional, highlighting the need for longitudinal analyses and examinations of the effects of early SEP participation on later educational success (Jackson 2003; Stake and Mares 2001).

Prior research suggests that enhancing black and Latino students' access to resources through SEPs may help to close the racial and ethnic STEM gap. For example, research shows that black and Latino students' likelihood to expect to pursue STEM careers at the primary and secondary levels of education is similar to that of their white peers when controlling for differences in access to educational resources, e.g. school-level expenditures per student that systematically disfavor black and Latino students (Oakes 1990; Walker and Rakow 1985). MESA might therefore provide one way to balance STEM knowledge and interest between black and Latino students on the one hand and white and Asian students on the other. Further, black and Latino college freshmen are *more* likely than white students to select STEM degree programs when controlling for prior achievement (Staniec 2004; Tyson et al. 2007). Having programs like MESA available may act to increase the development of these aspirations and expectations among black and Latino students. Providing guidance and resources to develop an interest in STEM and to translate that interest into behaviors, such as taking STEM courses in high school, might further boost black and Latino students' self-confidence and likelihood of pursuing STEM at the postsecondary level.

Current Study

This study is motivated by the potential to improve our understanding of whether MESA participation improves STEM outcomes for underrepresented minority students. This leads us to our research questions:

(1) Does participation in MESA increase the odds of students taking AP STEM courses?
(2) Does participation in MESA increase the odds of students planning to major in a STEM field in college?
(3) Do race and ethnicity moderate these associations?

The current study allows us the opportunity to explore how participation in an SEP can affect key academic outcomes related to future STEM participation and to inform research and policy by using novel data and rigorous methodological techniques. Across programs, SEPs vary considerably in terms of curricula, intensity, duration, and levels of implementation, making comparisons difficult even though SEPs are generally unified by the goals of STEM persistence and achievement. Even within MESA, program implementation can vary across participating schools, e.g., some schools hold MESA periods during school, while others hold them before or after school. Given these details, we do not claim that all SEPs behave like MESA or that every participating MESA student or school will experience identical program implementation. Instead, we provide evidence that shows how implementing a set of supports like MESA's—several of which (e.g., study help, college and career guidance) are fairly standard across major SEP programs—may affect students.

Data

The High School Longitudinal Study of 2009 (HSLS:09) is an ongoing nationally representative survey of approximately 25,210 students nested within 944 public and private high schools who were in the ninth grade in the fall of 2009. We use the restricted HSLS:09 transcript data that is especially useful in answering the questions we have raised because of its rich data on science and math enrichment programs, its explicit focus on math and science achievement and participation in high school, and the tracking of STEM postsecondary and occupational orientations. We use data from the baseline wave that was completed in the fall of 2009 when students were in the first semester of ninth grade through the 2013 update that was fielded in the summer and fall of 2013 immediately following on-time high school graduation and, in many cases, the first semester of college. The 2013 update was designed to capture information on students' high school completion, postsecondary plans, applications and acceptances to postsecondary institutions, education and work plans, financial aid applications and offers, choice of postsecondary institution, and employment experiences. The HSLS is arguably the ideal data set to study STEM outcomes as well as any racial and ethnic heterogeneity in these outcomes. The HSLS also provides key methodological advantages in terms of both the longitudinal design and national representativeness. As with nearly all observational data sets, the HSLS contains missing data. We followed the procedures for filling in missing data following the maximum likelihood estimation techniques outlined in Allison (2002). Missing data ranged from 0% (9th grade math score) to 9.96% (school/district offers incentives to attract full time high school science teachers). Rather than drop cases with missing values, we imputed missing data using the ice command in Stata (Allison 2002; von Hippel 2007).

STEM Enrichment Program

The treatment variable is a binary indicator for whether or not students participated in MESA in the fall of 9th grade. Our analysis is not a formal program evaluation of MESA. That is, we do not assess the dosage or timing of MESA on students' outcomes.[3] However, we are able to assess the impact of MESA, operationalized as whether the student indicated that they were currently participating in MESA in the fall of 9th grade, on key subsequent STEM outcomes. We should note that while enrichment programs, such as MESA, are sometimes nested within schools, this is not always the case. Nevertheless, all of our models control for whether students' schools partner with any enrichment programs so as to account for between-school variation in access to STEM enrichment and planning opportunities (Legewie and DiPrete 2014).

Because MESA targets students that come from socioeconomically disadvantaged backgrounds and because the focus of much policy intervention is on the doubly disadvantaged population of underrepresented minority students from disadvantaged backgrounds, it is important to focus particular attention on the distribution of MESA participation by both race and ethnicity and household income. To this end, Fig. 1 demonstrates that (1) black, Latino, and white, students from the most socioeconomically disadvantaged households (i.e., net household incomes of 0–35 thousand dollars) are most likely to participate in MESA; (2) while there is a gradual reduction in MESA participation

[3] The counterfactual group could include students who participated in other SEPs. Given MESA's explicit focus on STEM, however, we distinguish it among all of the other possible SEPs because of the program's specialized attention on getting underrepresented students into the STEM college pipeline.

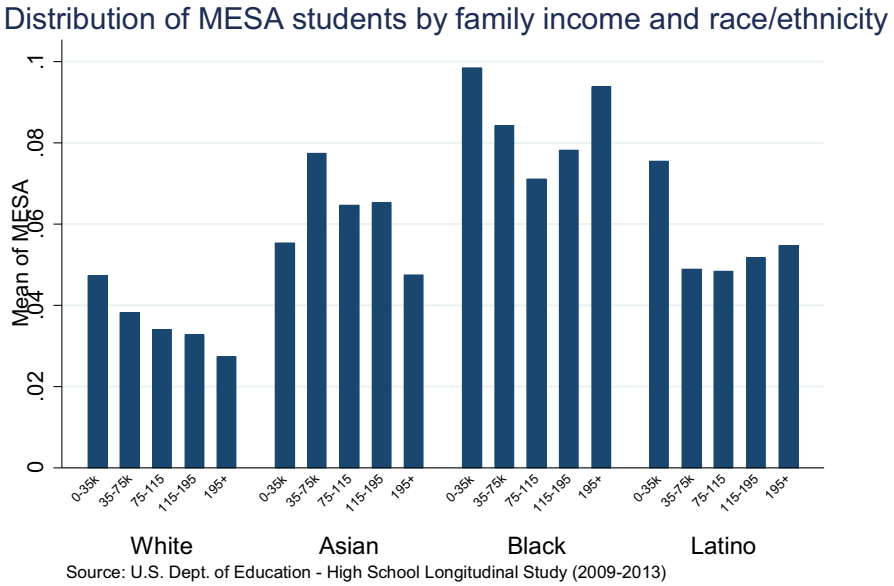

Fig. 1 MESA participation by race and ethnicity and household income

as income increases for white students, Asian, black, and Latino students demonstrate considerable participation throughout the socioeconomic distribution; and (3) black students from every socioeconomic category have relatively pronounced levels of MESA participation. What is more, black students at the highest end of the economic distribution are more likely to participate in MESA compared to black students from any other point in the distribution, save for those at the bottom. Online Appendix A demonstrates that the differences in the zero-order associations between MESA participation and race and ethnicity depicted in Fig. 1 are statistically significant (whites as reference category).

Key STEM Outcomes

The two outcomes we analyze represent students' accomplishments and plans up through the fall of 2013 (i.e., the fall immediately following on-time high school graduation) and therefore represent the vast majority of students' high school experiences as well as their initial plans for college. The first outcome, taking any AP STEM course, indicates whether students have taken an AP math course, AP statistics course, an AP science course, AP biology, AP chemistry, AP physics, or AP computer science. Thus, AP STEM courses are meant to represent students' commitments to STEM vis à vis their enrollment in the most rigorous STEM courses. The second outcome, planning to major in a STEM field in college, represents students' aspirations and/or expectations for their major field of concentration in college. This was asked of students in the follow-up after college and is included in the transcript data wave. The specific wording of the variable is: "What field of study or program [will/were] you [be] considering?" and this question was asked only to those who were attending a postsecondary institution as of November 1, 2013. Together, these two outcomes yield substantive measures of students' investments in STEM, their

future plans for STEM participation in college, and provide an indirect indication of their labor market trajectories.

Covariates

Table 1 provides means and standard deviations for all independent and dependent variables broken down by race and ethnicity. Because scholars have found that socioeconomic advantage leads to higher rates of persistence in advanced high school courses (Crosnoe and Schneider 2010) we control for resources associated with family socioeconomic background by accounting for net household income, poverty status, parent's education, and the number of persons in the household. We account for students' academic interest and achievement with data on whether either parent received any degree in STEM, whether science, math, or computer science is the student's stated favorite subject, whether the student took an advanced math or science course in the 8th grade, students' 8th grade science or math course grade, and 9th grade math achievement. We also control for whether students have an educational or career plan and whether students took a science or math course in the fall of ninth grade. Further, we control for parents' educational expectations for their children.

Finally, because schools' resources and hierarchical climates may negatively impact disadvantaged students' participation in advanced courses and college preparation, we include a vector of variables that address the socioeconomic and academic climate of the school. For example, we adjust for the percent of students receiving free or reduced priced lunch, the percent of students taking AP courses, the number of certified full-time math and science teachers, whether the school offers advanced science and math courses, whether the school requires completion of specific math or science courses for graduation, whether the school partners with MESA or similar enrichment program, a scale measuring if the school has a pro-science climate, and whether the school has a program to encourage black and Latino students to engage with math or science.[4] The final covariate addresses between-school variation in climate regarding racial and ethnic participation in STEM, which previous scholars have found to be influential for gender gaps in STEM participation (Legewie and DiPrete 2014). Finally, we adjust for whether the school uses a tracking policy to place students in 9th grade courses (administrator report).

Methods

We invoke a counterfactual causal framework wherein the effect is formally defined as the difference in outcome between the scenario in which an individual receives some treatment and the counterfactual scenario in which a similar individual receives a different treatment

[4] Pro-science climate includes whether the school has any of the following features: special focus on math or science, holds math or science fairs/workshops/competitions, offers pre-high school summer reading/math instruction for struggling 9th graders, sponsors a math or science after school program, pairs students with mentors in math or science, brings in guest speakers to talk about math or science, takes students on math- or science-relevant field trips, tells students about math/science contests/websites/blogs/other programs, requires teacher professional development in how students learn math/science, requires teacher professional development in increasing interest in math/science, raises student math/science interest/achievement in another way, offers College Board AP courses on-site, offers incentives to attract full-time high school math or science teachers, and on-site AP offerings for Calculus AB, Calculus BC, Computer Science, Computer Science A, Computer Science B, Advanced Biology, Advanced Chemistry, or Advanced Physics.

Table 1 Descriptive statistics of HSLS sample

Variable	Whites (N = 14,520)		Asians (N = 2490)		Blacks (N = 3140)		Latinos (N = 3520)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Treatment								
MESA	0.04	0.19	0.06	0.25	0.09	0.28	0.06	0.24
Outcomes								
Took an AP STEM course by Spring 2012	0.15	0.36	0.27	0.45	0.10	0.30	0.10	0.30
Plans to major in STEM	0.25	0.44	0.33	0.47	0.22	0.41	0.23	0.42
Students characteristics								
Favorite subject is science, math, or computer science	0.27	0.44	0.33	0.47	0.28	0.45	0.26	0.44
Advanced science course in 8th grade	0.05	0.22	0.07	0.25	0.06	0.24	0.06	0.23
A or B in most advanced science course in 8th grade	0.80	0.40	0.84	0.37	0.69	0.46	0.69	0.46
Advanced math course in 8th grade	0.76	0.42	0.82	0.39	0.74	0.44	0.71	0.45
A or B in most advanced math course in 8th grade	0.76	0.43	0.83	0.37	0.67	0.47	0.69	0.46
Educational/career plan in fall of 9th grade	0.60	0.49	0.67	0.47	0.71	0.46	0.62	0.49
Student took science fall of 9th grade	0.85	0.36	0.85	0.36	0.80	0.40	0.80	0.40
Student took math fall of 9th grade	0.91	0.28	0.91	0.29	0.88	0.33	0.88	0.33
Math score in fall of 9th grade	51.78	9.68	55.76	10.99	47.85	9.65	48.99	9.62
Family background								
Family income								
0–34k	0.22	0.41	0.25	0.43	0.38	0.49	0.39	0.49
35–74k	0.32	0.47	0.31	0.46	0.31	0.46	0.30	0.46
75–114k	0.22	0.41	0.20	0.40	0.17	0.37	0.16	0.37
115–194k	0.16	0.36	0.15	0.36	0.10	0.30	0.10	0.30
195 + k	0.08	0.28	0.09	0.29	0.05	0.22	0.05	0.22
In poverty	0.17	0.37	0.21	0.41	0.30	0.46	0.32	0.47
Parents' education								
Less than high school	0.03	0.18	0.06	0.23	0.05	0.22	0.16	0.36
HS	0.35	0.48	0.27	0.45	0.40	0.49	0.37	0.48
Some college	0.42	0.49	0.39	0.49	0.38	0.49	0.34	0.47
BA+	0.20	0.40	0.28	0.45	0.17	0.37	0.13	0.34
Parents' educational expectations for child								
Less than high school	0.00	0.07	0.00	0.04	0.00	0.06	0.00	0.07
HS	0.09	0.28	0.05	0.22	0.10	0.30	0.09	0.29
Some college	0.46	0.50	0.37	0.48	0.39	0.49	0.41	0.49
BA+	0.45	0.50	0.57	0.49	0.51	0.50	0.49	0.50
Number of household members	4.32	1.31	4.42	1.37	4.28	1.47	4.52	1.43
At least one parent's highest degree was STEM	0.14	0.35	0.25	0.43	0.11	0.32	0.10	0.31

Table 1 continued

Variable	Whites (N = 14,520)		Asians (N = 2490)		Blacks (N = 3140)		Latinos (N = 3520)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
High school characteristics								
Percent of students receiving reduced/free lunch	29.78	22.71	28.96	24.78	40.12	28.01	40.19	28.03
Percent of students enrolled in advanced placement courses	15.94	13.37	20.14	14.92	15.93	13.18	16.79	13.73
Number of full-time certified math teachers	9.01	6.07	12.23	7.15	10.12	6.82	11.04	7.50
Number of full-time certified science teachers	8.23	5.58	11.38	6.87	8.97	5.95	9.91	6.79
School offers advanced STEM courses	0.78	0.42	0.79	0.41	0.71	0.45	0.78	0.42
School requires specific math courses	0.90	0.29	0.91	0.29	0.93	0.26	0.91	0.28
School requires specific science courses	0.87	0.34	0.87	0.34	0.88	0.32	0.86	0.35
School partners with MESA or similar enrichment program	0.21	0.41	0.24	0.43	0.25	0.43	0.24	0.43
School has a pro STEM climate	2.87	1.40	3.48	1.35	3.01	1.41	3.12	1.41
School uses a tracking policy	0.46	0.50	0.46	0.50	0.46	0.50	0.48	0.50
Program that encourages underrepresented students in STEM	0.28	0.45	0.32	0.47	0.35	0.48	0.32	0.47

Source U.S. Department of Education. National Center for Educational Statistics: HSLS 2009–2013

(Morgan and Harding 2006; Morgan and Winship 2014; Shadish et al. 2002). Here, the treatment is participating in MESA in the fall of 9th grade. We create our comparison by estimating students' propensity to participate in MESA conditional on improving the covariate balance using observed characteristics of children, their families, and the schools that they attend.

Specifically, we used propensity score matching (PSM), as developed by Rosenbaum and Rubin (Rosenbaum and Rubin 1983b; Rubin 1974, 1980, 1978), which is widely considered an informative alternative for estimating causal effects in the absence of randomized data (Becker and Ichino 2002; Caliendo and Kopeinig 2008; Stuart and Rubin 2008). We compared students who participated in MESA with the control group of students who did not and estimated the average treatment effect on the treated (ATT). The strength of matching lies in its ability to reduce the role of observed covariates on any remaining differences between students who participated in MESA and students who did not after matching if selection into treatment depends exclusively on observed variables, thereby assuming that treatment assignment is 'ignorable' (D'Agostino 1998). We test the robustness of this assumption in a formal sensitivity analysis. Therefore, the strengths of matching over traditional regression are that matching (1) allows for the explicit assessment of covariate balance between treatment and control groups; (2) it is non-parametric (DiPrete and Gangl 2004); and (3) it allows for the formal testing of confoundedness in treatment effects (Caliendo and Kopeinig 2008; Heckman et al. 1998). Still, PSM alone is

no panacea for unobserved confounding and is limited in that it also relies on observed covariates.

We compiled a list of variables that may impact educational investments and outcomes.[5] Given that the primary purpose of these covariates is to predict selection into the treatment group rather than to determine distinct effects on the outcomes of interest, we opted to include a set of 24 independent variables that may drive MESA selection. Table 1 lists the vector of covariates used for matching save for the outcomes and race and ethnicity. We enter all covariates into the selection model as main effects. We calculate propensity scores using a logit model (Leuven and Sianesi 2003) and we match individuals who had similar propensities to enroll in MESA through kernel matching within a bandwidth of 0.09 (Heckman et al. 1998; Stuart and Rubin 2008). All analyses are restricted to observations that fell in the region of common support to minimize the possibility of bad matches (Caliendo and Kopeinig 2008). We evaluate whether the groups being compared have equal (or sufficiently balanced) distributions of relevant observed variables (Dehejia 2005) by inspecting standardized bias scores.[6]

Unobserved variables that affect both treatment assignment and the outcome threaten our ability to make causal inferences (Stuart and Rubin 2008). As a result, we conduct a formal sensitivity analysis (see Online Appendix B for details of the sensitivity analysis) of our statistically significant ATTs, which allows us to gauge the robustness of our estimates and increase confidence that these estimates represent "real" effects.

Limitations

A key limitation of this paper is that the HSLS does not allow us to formally assess the dosage or cumulative effects of MESA participation. That is, we lack data on whether students participated in MESA for many years and cannot therefore tell if there are differential effects that accumulate over time as students are exposed to MESA resources over time. Similarly, we do not know the multiplicative effect of participating in MESA and some other enrichment program in college such as Summer Bridge, Prefreshman Academic Enrichment Programs, McNair Scholars, Mellon-Mays Scholars, and other similar programs. MESA alone may not produce strong effects on our outcomes for black and Latino students but the combination of participation with other programs may indeed do so. These are opportunities that we hope future research can address with more detailed data.

Another limitation stems from the fact that students may erroneously self-report MESA participation. Online Appendix C demonstrates that students' reporting of MESA participation spanned almost every state, even those not listed on the official MESA website (www.mesausa.org/). Among the possible reasons for this are (1) students are misreporting their MESA participation; (2) students are in reality participating in a MESA program that is not officially sanctioned or under the auspices of the national MESA organization (i.e., offshoots that are providing enrichment like MESA and are being identified by students as

[5] In order to control for heterogeneity in selection into the treatment and outcome that lies between schools, we included school-level variables in the matching model. This adjustment allows us to partially account for the clustered nature of the data and accurately specify the propensity score model (Arpino and Mealli 2008).

[6] According to the Stata help file for the command pstest, "The standardized percent bias is the percent difference of the sample means in the treated and non-treated (full or matched) sub-samples as a percentage of the square root of the average of the sample variances in the treated and non-treated groups (formula from Rosenbaum and Rubin 1985).".

MESA by name but may not be officially sanctioned or listed on mesausa.org); (3) migration to a non-MESA state after participating at some point between kindergarten and 11th grade; or 4) some other unknown reason. This is all speculation that can neither be substantiated nor disproven with the existing data. Nevertheless, as an ancillary robustness check, we re-ran our PSM models using only students in those states listed on mesausa.org (see Online Appendix D). These results support our main results in Table 2 almost exactly. The similarity in results when we limit our sample to students in MESA states gives us confidence in our main results and conclusions.

Further, we must limit our claims to the assumptions inherit in any study that uses observational data—which even extends to studies that formally test the robustness of treatment effects such as ours. One such assumption is that the Rosenbaum bounds technique may be too conservative given that it does not simultaneously model the impact of hidden bias on both selection into the treatment and the outcome. Other sensitivity approaches that also model how an unobserved confounder impacts outcomes might refine our understanding of the impact of hidden bias (Ichino et al. 2008). Further, the Rosenbaum bounds technique cannot assess the impact of an array of unobserved confounders on estimated treatment effects. Still, Rosenbaum bounds have the appeal of mimicking randomized experiments by treating the impact of an unobserved confounder on the outcome as irrelevant. A second limitation is the assumption that the unobserved variable for which we test is binary: a continuous unobserved confounder could potentially impact our findings. Additionally, while the techniques employed here can lead to unbiased estimates of causal effects, we may not always clearly know when it has done so. Still, the empirical results and subsequent sensitivity analysis lead us to the conclusion that our results are plausibly robust to an unobserved confounder.

Results

Descriptive Statistics

We begin with results from the descriptive analysis in Table 1. All of the variables in Table 1 are dichotomous (i.e., 0 or 1 values) save for number of household members, student's ninth grade math score, school percent on free/reduced priced lunch, percent of students enrolled in AP courses, the number of certified math and science teachers, and the scale for school pro-science climate.

The proportion of students who participated in MESA varies by race and ethnicity. Among all students combined, the average MESA participation is 5% (authors' calculation of HSLS data). However, participation varies from 4% among white students, to 6% among Asian and Latino students, and 9% among black students.[7] While we cannot know the exact reason for the varying participation rates in the HSLS sample across racial and ethnic groups, we can speculate that these students may have passed through at least one of the MESA elementary, middle, or high schools and therefore had the opportunity to become involved in MESA. Moreover, given that the HSLS has a specific focus on STEM, it is likely that the sampling frame called for sampling from schools that were likely to

[7] We would like to note that findings for Southeast Asian American students may differ—especially given likely differences in family socioeconomic status. However, we cannot pinpoint any such differences since we did not disaggregate the data sufficiently in this regard.

Table 2 Treatment effects of MESA participation on AP STEM coursetaking and STEM major (bandwidth = 0.03)

	(1) All students		(2) Whites		(3) Asians		(4) Blacks		(5) Latinos	
	AP STEM	STEM major	AP STEM	STEM major	AP STEM	STEM major	AP STEM	STEM major	AP STEM	STEM major
MESA	**0.04**	**0.04**	0.03	**0.07**	0.02	0.00	**0.07**	0.04	0.03	0.02
Standard error	0.01	0.01	0.02	0.02	0.04	0.04	0.02	0.03	0.02	0.03
T-statistic	3.31	3.33	1.96	3.61	0.43	0.01	3.11	1.53	1.34	0.57
On support untreated	23,663	22,663	13,979	13,979	2326	2326	2865	2865	4493	4493
On support treated	1265	1265	545	544	161	160	274	273	285	285
Bias before matching	10.71	10.33	10.08	9.94	9.93	10.31	10.43	9.43	9.28	9.04
Bias after matching	6.28	6.12	7.30	7.09	3.62	3.84	2.66	2.72	5.57	5.61

Covariates: student: favorite subject in school is science, math, or computer science, advanced math and science courses in 8th grade, received an A or B grade in math and science courses in 8th grade, has an educational or career plan, took math or science course in fall of 9th grade, 9th grade math score. Family: family income, family in poverty, parents' highest degree earned, parent's educational expectations for child, number of household members, at least one parent's highest degree was in STEM. School: percent receiving reduced/free lunch, percent enrolled in AP courses, number of full time certified math or science teachers, school offers advanced STEM courses, school requires specific math or science courses to graduate, school partners with MESA or similar enrichment program, school has a pro-STEM climate, school uses a tracking policy, school has a program that encourages underrepresented students in STEM

Source U.S. Department of Education. National Center for Educational Statistics: HSLS 2009–2013

Bold ATTs denote statistically significant effects ($p < 0.05$)

have MESA and other STEM-oriented enrichment programs or students that had at least some exposure to MESA.

In terms of family background, White and Asian students are more advantaged than black and Latino students by every measure including income, rates of poverty, and parent's education. White and Asian families are much more likely to have at least one parent with a STEM degree compared to black and Latino families, reflecting historical disparities in STEM attainment by race and ethnicity.

At the school-level, there seems to be equality in terms of resources and organizational structure by race and ethnicity (e.g., STEM enrichment program offerings, science and math requirements for graduation, school pro-STEM climate, and tracking policies). However, Asian students attend schools with a higher percent of students enrolled in AP courses than any other group. Also, black and Latino students are more likely than white and Asian students to attend schools whose pupils come from socioeconomically disadvantaged families, as measured by the percent of the students receiving free or reduced priced lunches, while also being more likely than white students to attend schools that offer STEM enrichment programs that specifically target black and Latino students.

Logistic Regression

To address our central research questions, we began by conducting an analysis of the association between our treatment, MESA, and students' outcomes without covariate adjustment. These traditional logistic regression models allow us to examine the full sample rather than restricting the sample to respondents in the region of common support and therefore provide a more comprehensive, albeit more unbalanced, depiction of results compared to the PSM models. Here, our focus is on whether separate racial and ethnic groups experience associations between MESA participation and the two outcomes, not to formally assess either moderating or causal effects.

Figure 2 renders the results from unadjusted models where we regress high school AP STEM coursework and college STEM major aspirations on MESA participation for all students combined and for each racial and ethnic group separately. Among the most immediately noticeable findings are that (1) MESA participation appears to have a positive impact on AP STEM coursework and on STEM major aspirations among all students (the odds of taking an AP STEM course increase by 19% ($e^{0.17} = 1.19$) and the odds of aspiring toward a STEM major increase by 23% ($e^{0.21} = 1.23$)); (2) MESA appears to have uneven impacts on these outcomes by race and ethnicity, suggesting that race and ethnicity may act as a moderator for these outcomes; (3) black students' odds of taking an AP STEM course increases by 70% due to MESA participation $\left(e^{0.53} = 1.70\right)$; and (4) white students' odds of aspiring to major in STEM in college increases by 43% due to MESA participation $\left(e^{0.36} = 1.43\right)$. Somewhat surprisingly given the nature of the programs, MESA does not appear to significantly impact AP STEM coursework or STEM major aspirations among Latinos.

Next, we present fully adjusted models of the association between MESA participation and the two outcomes of interest. Like the bivariate models, these models also use more comprehensive data than the PSM models by including students beyond the region of common support. Additionally, these models improve observed covariate balance, but not in the specific manner that PSM does so. Figure 3 renders the results of these models and suggests that MESA participation increases the odds of all students taking any AP STEM course by 55% ($e^{0.44} = 1.55$) while also suggesting that the odds for white students increase by 49% ($e^{0.40} = 1.49$) and the odds for black students increase by 110% ($e^{0.74} = 2.10$). Curiously, these results further suggest that the unadjusted association between MESA and AP STEM coursework for whites could have been suppressed. Figure 3 also demonstrates that MESA participation increases the odds of all students to aspire to major in a STEM field in college by 34% ($e^{0.29} = 1.34$). However, the group-specific analysis reveals that MESA only affects STEM major aspiration for white students [by 54% [($e^{0.43} = 1.54$)]. As in the unadjusted models, MESA participation does not appear to have a statistically significant impact on black or Latino students' odds of aspiring to major in college. Online Appendix E summarizes coefficients and standard errors for the analyses summarized in Figs. 2 and 3.

The finding of racial and ethnic disparities for AP STEM coursework may suggest that MESA could act as a remedy for previous research findings for the underrepresentation of black students in advanced courses (Riegle-Crumb and Grodsky 2010). Overall, these regression-based findings offer suggestive evidence that the impact of MESA on STEM outcomes may differ by racial and ethnic subgroup.

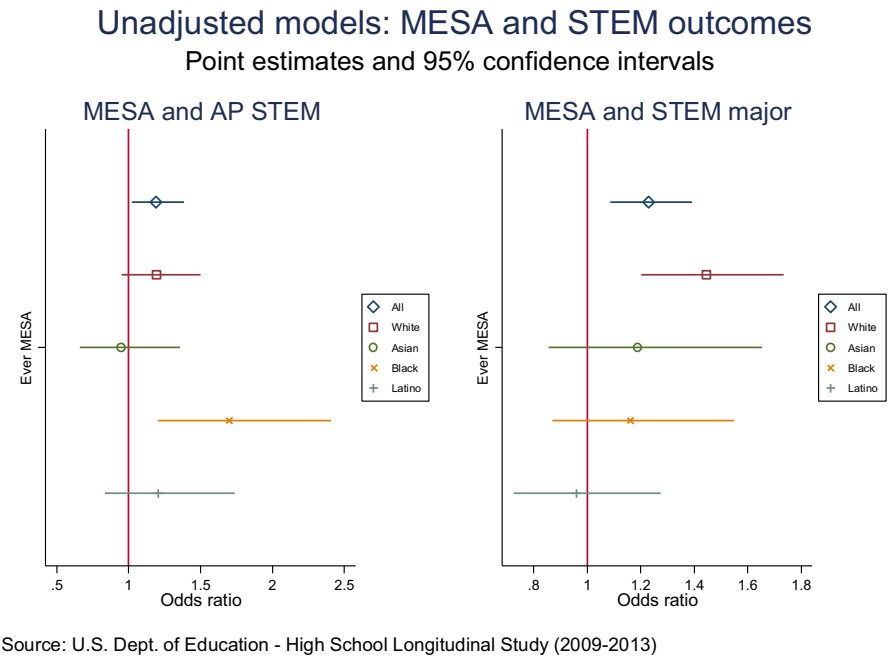

Fig. 2 Unadjusted impact of MESA on high school AP STEM coursework and STEM major aspirations, by race and ethnicity

Propensity Score Matching

Table 2 summarizes the findings from PSM models that we ran separately for all students combined and by racial and ethnic group separately. We maximize the similarity among matched pairs of treated and untreated respondents by limiting the proximity of matches to a kernel bandwidth of 0.03. In addition to ATTs we also report standard errors, T-statistics, sample sizes for treatment and control groups in the region of common support, as well as pre and post-matching bias. However, here we only report the overall standardized imbalance, or bias, before and after matching. Statistically significant ($p < 0.05$) treatment effects are in bold.

The reduction in covariate bias indicates that the PSM model is successfully creating a more balanced set of comparison groups than with traditional regression models, the latter of which can be assessed by examining pre-matching bias scores. That is, the PSM technique has very likely reduced the differences between treatment and control samples as evidenced by the reduction in the pre and post-matching bias scores. Our post-matching bias ranged from 2.72 (blacks) to 7.30 (whites) and we should note that levels of post-matching bias that approach 5% are generally considered sufficient (Caliendo and Kopeinig 2008). For further detail, Online Appendices F and G demonstrate the pre- and post-matching standardized bias for each of the covariates in our PSM models (where zero bias is the ideal), clearly demonstrating migration toward zero bias after matching for the AP STEM and the STEM major outcomes, respectively.

In the first column of Table 2, we consider the MESA treatment effect (vs. not participating in MESA) on the likelihood of taking any AP STEM course among all students.

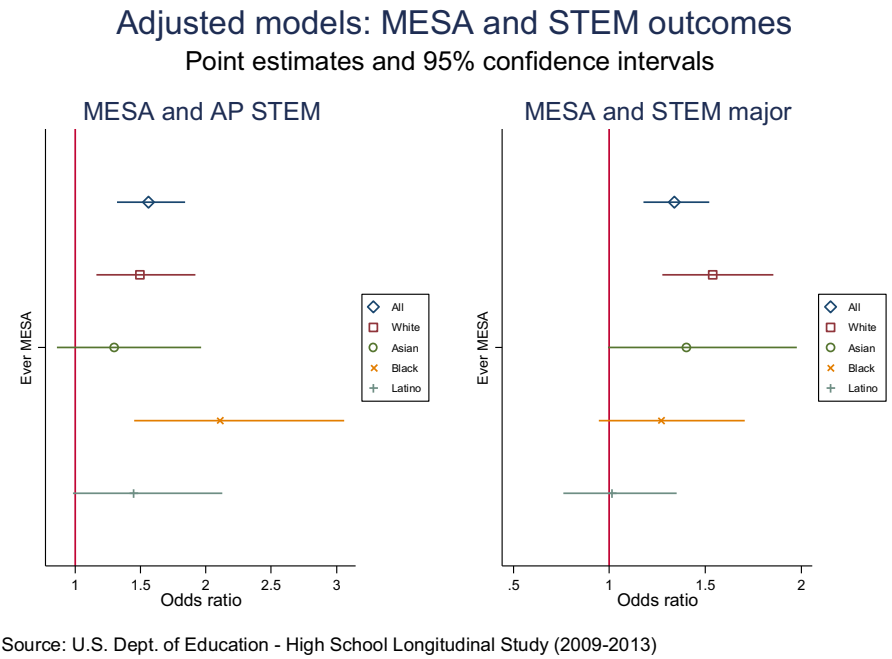

Fig. 3 Adjusted impact of MESA on high school AP STEM coursework and STEM major aspirations, by race and ethnicity

We find that MESA participation increases all students' probability of taking an AP STEM course by 4 percentage points. Scanning across the row, we find that black students drive this effect. That is, MESA participation increases the probability of taking an AP STEM course by 7 percentage points for black students. The treatment effect for whites, however, is moderately statistically significant (T = 1.96) and suggests that MESA participation increases their probability of taking an AP STEM course by 3 percentage points. Statistical significance here indicates that a treatment effect is statistically different from zero.

In the second column of Table 2, we consider the MESA treatment effect on respondents' likelihood of stating that they aspire and/or expect to major in a STEM field in college. Again, we find that MESA participation increases the likelihood that students, in general, will state that they intend to major in STEM in college. MESA participation increases the probability of aspiring to major in STEM by 4 percentage points for all students. Scanning across the row, we find that white students primarily drive this effect—their probability of aspiring to major in STEM increases by 7 percentage points due to MESA participation. No other racial and ethnic group shows even marginally significant effects of MESA participation on STEM major expectations.

In summary, the evidence from the PSM models suggests that MESA participation has positive effects on STEM outcomes that are limited to white and black students. MESA appears to increase the probability of taking AP courses for black students and increase the probability of aspiring towards a STEM major for white students. Moreover, the findings from the unadjusted models in Fig. 2 are not spurious to observed covariate imbalance. Indeed, the group-separate analyses reveal that MESA only affects AP STEM coursework for black students and MESA only affects STEM major aspirations for white students.

Mantel–Haenszel Bounds Sensitivity Analyses

Like traditional regression, the PSM strategy continues to suffer from the 'hidden bias' problem stemming from the inability to control for unobserved heterogeneity in selection into the treatment. One partial solution to this issue is to test how robust the treatment effects are to the bias that stems from an unobserved variable. Although the sensitivity analysis is not a remedy for the problems of unobserved heterogeneity, we will test the robustness of our results to incremental erosions of the ignorability assumption.

Our sensitivity analysis allows us to assess how large an unobserved confounder, U, and its associated selection bias, must be in order to undermine our results, adjusting for observed controls. Here, we report the range of gammas where the statistically significant ATT became statistically insignificant (i.e., the 'kill zones') due to the unobserved confounder, U. The gammas (Γ) are presented as odds ratios ranging between 1.00 and 2.00 in increments of 0.05. To be clear, Mantel–Haenszel bounds assume that the unobserved confounder is perfectly correlated with the outcome, suggesting that these 'kill zones' represent *conservative* bounds on our treatment effect. That is, they represent scenarios where the effect of U may be much stronger than we might expect a priori (Rosenbaum and Harris 2001).

In Model 1, the results demonstrate that U undermines the statistically significant ATT for the effect of MESA on taking AP STEM courses among all students when U is between 1.00 and 1.05, net of controls. That is, our ATT is no longer statistically significant when U causes the odds ratio of treatment assignment to differ between treatment and control cases by a factor of about 1.05 for all students after including observed covariates in the model. Furthermore, the results demonstrate that U undermines the statistically significant ATT for the effect of MESA on taking AP STEM courses among black students when U is between 1.05 and 1.10, net of controls. In Model 2, the critical level of Γ at which we would have to question our conclusion of a positive effect of MESA participation on planning to major in STEM among all students combined is between 1.05 and 1.10, net of observed covariates. Among white students, a hidden bias of between 1.20 and 1.25 would be necessary to render our results of a positive effect of MESA participation spurious, net of observed covariates.

Online Appendix H summarizes the results from separate selection models that predict MESA participation based on observed covariates. Comparing the results for the bounds of the 'killer' confounder in Model 1 of Table 3 with the selection effects of observed covariates in Appendix H, an unobserved confounder would have to be as strong as or stronger than being in poverty (1.11) to render our result for black students spurious. It is important to recall, however, that this theoretical confounder would have to mimic the impact of poverty *after* already including poverty in the model as a control, which we have done.

Examining the sensitivity results in Model 2 in Table 3, the unobserved confounder would have to be as strong as high achievement in 8th grade math (1.29) or much stronger than poverty (1.10) to undermine the treatment effect for SEP on taking an AP STEM course among white students. It is important to take into account that we have already included these observed variables in our models. Therefore, one must first justify, in a theoretical sense, the inclusion of some unobserved confounder that has a larger impact on selection into the treatment than those variables that we have already included when calling our results into question.

Table 3 Sensitivity analysis: Magnitude of the unobserved binary confounder's effect on selection into the treatment that renders the ATT null ($p < 0.05$)

	(1) Took an AP STEM course by Spring 2012 Γ (odds ratios)	(2) Plans to major in STEM
All students	1.00–1.05	1.05–1.10
Whites	–	1.20–1.25
Asians	–	–
Blacks	1.05–1.10	–
Latinos	–	–

Mantel–Haenszel bounds for treatment effects of college bound friend on high school performance

Binary unobserved confounder effect on selection into the treatment (Γ) is in the metric of odds ratios in increments of 0.05 between 1.00 and 2.00

Source U.S. Department of Education. National Center for Educational Statistics: HSLS 2009–2013

The strength of some as of yet unmeasured hidden variable rendering our qualitative conclusion for whites spurious must be greater than that for any other racial and ethnic group in any other model, net of observed covariates. That is, the treatment effect for MESA is most robust among white students. However, comparing each of the 'killer' confounders with observed covariates that we have already included in the propensity score matching model gives us confidence that all of our results may be robust to unobserved bias. That is, it is difficult to imagine a variable that we have not yet included that is (1) as strong as or stronger than the observed covariates and (2) can affect selection to the point of undermining our treatment effects. We should also note that in order to undermine our qualitative conclusions, any unobserved confounders must not only impact selection into the treatment on the order of some of the most theoretically relevant variables that we have already included in our propensity model but must also be able to almost *perfectly* predict the relative outcomes of the matched treatment and control variables (Rosenbaum and Harris 2001). While we do not absolutely rule out the possibility that an unobserved confounder could be lingering, we are confident that our results are robust given current data and methodological limitations.

To be clear, we do not claim to fully meet the assumption of 'ignorability.' However, we feel that we have exhausted the resources at hand to assess the robustness of the matching estimator to potential endogeneity bias and we conclude that any plausible confounder must be able to act very strongly (and independently) on the decision to participate in an SEP program.

Kernel Bandwidth Robustness Check

Finally, as a second method of testing the robustness of our treatment effects, we supplement the results in Table 2 with results from models that increase the number of individuals in the control group we use to construct the counterfactual outcome vis à vis the weighted averages in kernel matching. Online Appendix I summarizes the treatment effects after expanding the bandwidth to 0.09, yielding a smoother estimated density function, improving fit, and decreasing the variance between the estimated and the true underlying density function (Caliendo and Kopeinig 2008). Online Appendix I demonstrates that

increasing the bandwidth parameter by a factor of three has no real overall impact on the estimated treatment effects, despite slightly increasing the standardized post-matching bias due to utilizing a larger share of respondents as matches.

Discussion and Conclusion

This paper is the first to examine whether participating in MESA affects students' AP STEM course taking behaviors and postsecondary participation by using previously untapped restricted and nationally representative longitudinal data from the High School Longitudinal Study (2009–2013). Prior research has suggested that educational enrichment programs, especially ones that intervene in children's lives early on, may aid in mitigating racial and ethnic disparities in educational outcomes, but gaps in the research on the impact of MESA on STEM participation remain (Kaushal et al. 2011). Racial and ethnic disparities in STEM participation that emerge at the end of middle school and the beginning of high school place black and Latino students on a tenuous path that may inhibit their educational and economic attainment. Overall, our findings imply that MESA may be succeeding at improving minority students' participation in AP STEM educational preparation, at least for black students, but are falling short of improving the STEM aspirations of minority students. In particular, black students who participated in MESA were more likely to take AP STEM classes in high school compared to black students who did not. However, other than AP STEM for black students, participation in MESA did not result in statistically significant treatment effects for Asian, Black, or Latino students.

Still, one may argue that the AP STEM courses represent a more tangible investment in STEM than a single aspirations question, potentially increasing the positive implications of the MESA finding among black students. Further, AP STEM participation could increase both the training that is necessary to succeed in STEM courses in college, which could ultimately lead to increased confidence in actually majoring in a STEM field and in entering the STEM labor force.

The evidence we presented in this paper suggests that racial and ethnic inequality may persist in students' STEM outcomes despite interventions, such as MESA, in the academic careers of students. Participation in MESA represents an important early injection of academic support and guidance while the outcomes represent key high school and college behaviors that should lead students on a path to be competitive in their STEM postsecondary careers and in the labor market. Although the findings do not suggest overwhelming support for MESA's effect on black and Latino students on these outcomes, MESA may still have impacts on their actual participation in STEM courses in college, as it does in high school. Our findings suggest mixed results for MESA participation that depend on the outcome and on the sub-group under analysis.

Among the possible mechanisms through which MESA participation may affect AP course taking (for black students) and college STEM aspirations (for white students) are (1) an increase in social capital and motivation that accrue due to interaction with mentors and peers who are trained and focused on academic success, in general, and STEM, in particular (Owens and Massey 2011; Steele and Aronson 1995; Wang 2013); (2) access to educational resources and equipment that foster interest in scientific inquiry (Hewson et al. 2001; Oakes 1990; Walker and Rakow 1985); and (3) increased parental interest and participation in the educational and STEM experiences and outcomes of their children (Moakler and Kim 2014; Simpkins et al. 2015). The current analysis, however, is limited in

its ability to tease out which of these (or other) mechanisms drive the treatment effects we observe for all students combined and separately for different racial and ethnic groups.

MESA has a stated focus of increasing underrepresented students' high school performance and college attainment in STEM. Our descriptive findings suggest that these programs are by and large targeting the intended population of students because black students participate the most and white students do so the least, with Asians and Latinos in between. Moreover, the participation of black students spans the full socioeconomic range.

If MESA participation is not having the intended impact of increasing all underrepresented students' high school performance in advanced courses, then we are left with a series of questions that future research may investigate. One, what are other outcomes that SEP participation can affect? Second, who are these programs really trying to reach? The white students that we found benefited from MESA participation may comprise a segment of the targeted population if they are low income or whose parents did not attend college. Still, on average, white students come from higher SES families than black and Latino students. One could therefore argue based on Fig. 1 that MESA appears to be reaching the intended groups, without excluding other students who want to take advantage of it. Moreover, the data and methods of this analysis may not be suitable to identify the full range of effects that MESA may have on STEM outcomes. For example, it is entirely possible that MESA affects STEM major aspirations indirectly—such as through first affecting AP STEM course taking. Furthermore, MESA may be part of a larger set of interventions that could positively affect major aspirations and other outcomes that our research has not yet studied.

Nevertheless, the findings from the current paper provide evidence for both celebrating and questioning the effectiveness of MESA on improving the advanced STEM coursework and STEM major aspirations of black and Latino students. Selection on observed characteristics like ability, SES, and school resources, it appears, plays a large role in the ability of participation in MESA to alter the academic outcomes of black and especially Latino students, despite the explicit effort to target and develop the academic abilities of these students. Still, the positive treatment effect on taking AP STEM courses among black students is noteworthy not only because this represents an important stepping stone to postsecondary enrollment but also because it signals a commitment to education through advanced coursework. Improving the academic outcomes of black and Latino students may necessitate tighter linkages between program designs and high school curricula, linkages between families and programs to enhance social closure, and culturally relevant designs that could make them more attractive to black and Latino students. MESA participation is one of many possible remedies to black and Latino students' relatively weak STEM outcomes and—despite this paper's evidence of mixed results—the program still has the potential to positively impact the STEM outcomes of black and Latino students in ways that are beyond the scope of the current analysis. We leave it to future research to replicate and reanalyze these processes with new and improved data.

Acknowledgements The authors would like to thank their undergraduate research assistant, Summer Lopez Colorado, for her work on developing this paper, Dr. Anna R. Haskins and Clara M. Elpi for their careful reviews of the manuscript, and the anonymous reviewers for their helpful comments. In addition, the authors thank seminar participants at the Cornell Center for the Study of Inequality for their helpful comments and suggestions.

References

Adams, S. (2014). *The best jobs for 2014*. New York: Forbes.

Afterschool Alliance. (2011). *STEM learning in afterschool: an analysis of impact and outcomes*. Washington, DC: Afterschool Alliance.

Allison, P. (2002). *Missing data*. Thousand Oaks: Sage Publications.

Arpino, B., & Mealli, F. (2008). The specification of the propensity score in multilvel observational studies. *Computational Statistics & Data Analysis, 55,* 1770–1780.

Aud, S., Hussar, W., Johnson, F., Kena, G., Roth, E., Manning, E., et al. (2012). *The condition of education 2012 (NCES 2012-045).* Washington, DC: U.S. Department of Education, National Center for Education Statistic.

Becker, S. O., & Ichino, A. (2002). Estimation of average treatment effects based on propensity scores. *STATA Journal, 2,* 358–377.

Bottia, M. C., Stearns, E., Mickelson, R. A., Moller, S., & Parker, A. D. (2015). The relationships among high school STEM learning experiences and students' intent to declare and declaration of a STEM major in college. *Teachers College Record, 117,* 1–46.

Bureau of Labor Statistics (2013). Labor force characteristics by race and ethnicity, 2012. BLS Reports.

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys, 22,* 31–72.

Catsambis, S. (1995). Gender, race, ethnicity, and science-education in the middle grades. *Journal of Research in Science Teaching, 32,* 243–257.

Chemers, M., Zurbriggen, E., Syed, M., Goza, B., & Bearman, S. (2011). The role of efficacy and identity in science career commitment among underrepresented minority students. *Journal of Social Issues, 67*, 469–491.

Chen, X. (2009). Students who study science, technology, engineering, and mathematics (STEM) in post-secondary education. In Stats in Brief (NCES 2009-161). National Center for Educational Statistics.

Coleman, J. S., Campbell, E. Q., Hobson, C., McPartland, J., Mood, A., Weinfeld, F. D., et al. (1966). *Equality of educational opportunity*. Washington, DC: Government Printing Office.

Correll, S. J. (2004). Constraints into preferences: Gender, status, and emerging career aspirations. *American Sociological Review, 69,* 93–113.

Crosnoe, R., & Schneider, B. (2010). Social capital, information, and socioeconomic disparities in math course work. *The American Journal of Education, 117,* 79–107.

D'Agostino, R. B. (1998). Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Statistics in Medicine, 17,* 2265–2281.

Dalton, B., Ingels, S. J., Downing, J., Bozick, R., & Owings, J. (2007). *Advanced mathematics and science coursetaking in the spring high school senior classes of 1982, 1992, and 2004*. Washington, DC: National Center for Educational Statistics, Institute of Education Sciences, U.S. Department of Education.

Davidson, A. (2012). *Making it in America*. Boston, MA: The Atlantic.

Dehejia, R. (2005). Practical propensity score matching: A reply to Smith and Todd. *Journal of Economic Theory, 125,* 355–364.

DiPrete, T. A., & Gangl, M. (2004). *Assessing bias in the estimation of causal effects: Rosenbaum bounds on matching estimators and instrumental variables estimation with imperfect instruments* (pp. 271–310). Weinheim: Wiley-VCH Verlag Gmbh.

Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science, 312,* 1900–1902.

Heckman, J. J., Ichimura, H., & Todd, P. (1998). Matching as an econometric evaluation estimator. *Review of Economic Studies, 65,* 261–294.

Hewson, P. W., Kahle, J. B., Scantlebury, K., & Davies, D. (2001). Equitable science education in urban middle schools: Do reform efforts make a difference?*. *Journal of Research in Science Teaching, 38,* 1130–1144.

Hinojosa, T., Rapaport, A., Jaciw, A., LiCalsi, C., & Zacamy, J. (2016). *Exploring the foundations of the future STEM workforce: K–12 indicators of postsecondary STEM success*. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest.

Hoepner, C. C. (2010). Advanced placement math and science courses: Influential factors and predictors for success in college STEM majors. Proquest Dissertation Publication.

Ichino, A., Mealli, F., & Nannicini, T. (2008). From temporary help jobs to permanent employment: What can we learn from matching estimators and their sensitivity? *Journal of Applied Economics, 23,* 305–327.

Jackson, S. A. (2003). *The quiet crisis: Falling short in producing american scientific and technical talent*. San Diego: Building Engineering & Science Talent (BEST).

Kaushal, N., Magnuson, K., & Waldfogel, J. (2011). How is family income related to investments in children's schooling. In G. J. Duncan, & R. J. Murnane (Eds.) (pp. 187–205). New York: Russell Sage Foundation.

Kelly, S. (2009). The black–white gap in mathematics course taking. *Sociology of Education, 82,* 47–69.
Kelly, S., & Zhang, Y. (2016). Teacher support and engagement in math and science: Evidence from the high school longitudinal study. *High School Journal, 99,* 141–165.
Kim, C., Tamborini, C. R., & Sakamoto, A. (2015). Field of study in college and lifetime earnings in the United States. *Sociology of Education, 88,* 320–339.
Legewie, J., & DiPrete, T. A. (2014). The high school environment and the gender gap in science and engineering. *Sociology of Education, 87,* 259–280.
Leggon, C. B., & Pearson, W. (2006). Assessing programs to improve minority participation in STEM fields: What we know and what we need to know. In R. Ehrenberg, & C. V. Kuh (Eds.). Ithaca, NY: Cornell University Press.
Leuven, E., & Sianesi, B. (2003). PSMATCH2: Stata module to perform full mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing.
Lewis, J., Menzies, H., Najera, E., & Page, R. (2009). Rethinking trends in minority participation in the sciences. *Science & Education, 93,* 961–977.
Lucas, S. R. (1999). *Tracking inequality*. New York: Teachers College Press.
Moakler, M. W., & Kim, M. M. (2014). College major choice in STEM: Revisiting confidence and demographic factors. *Career Development Quarterly, 62,* 128–142.
Morgan, S. L. (1996). Trends in black–white differences in educational expectations: 1980–1992. *Sociology of Education, 69,* 308–319.
Morgan, S. L., & Harding, D. J. (2006). Matching estimators of causal effects—Prospects and pitfalls in theory and practice. *Sociological Methods & Research, 35,* 3–60.
Morgan, S. L., & Winship, C. (2014). *Counterfactuals and causal inference: Methods and principles for social research* (2nd ed.). Cambridge: Cambridge University Press.
Muller, P., Stage, F. K., & Kinzie, J. (2001). Science achievement growth trajectories: Understanding factors related to gender and racial-ethnic differences in precollege science achievement. *American Educational Research Journal, 38,* 981–1012.
NSF. (2014). Science and engineering indicators 2014. National Science Foundation (NSB 14-01).
NSF and NCSES. (2013). *Women, minorities, and persons with disabilities in science and engineering: 2013*. Alexandria, VA: National Science Foundation.
Oakes, J. (1990). *Opportunities, achievement, and choice—Women and minority-students in science and mathematics* (p. 16). Educ: Review of Research in Education.
Orfield, G., Frankenberg, E., Ee, J., & Kuscera, J. (2014). Brown at 60: Great progress, a long retreat and an uncertain future.
Ormerod, M. B., & Duckworth, D. (1975). *Pupils' attitudes to science*. Slough: NFER.
Owens, J., & Massey, D. S. (2011). Stereotype threat and college academic performance: A latent variables approach. *Social Science Research, 40,* 150–166.
Press, W. H. (2013). What's so special about science (and how much should we spend on it?). *Science, 342,* 817–822.
Provasnik, S., Kastberg, D., Ferraro, D., Lemanski, N., Roey, S., & Jenkins, F. (2012). *Highlights from TIMSS 2011: Mathematics and science achievement of US fourth-and eighth-grade students in an international context*. Washington, DC: National Center for Education Statistics.
Rakow, S. J., & Walker, C. L. (1985). The status of Hispanic American-students in science—Achievement and exposure. *Science & Education, 69,* 557–565.
Riegle-Crumb, C., Farkas, G., & Muller, C. (2006). The role of gender and friendship in advanced course taking. *Sociology of Education, 79,* 206–228.
Riegle-Crumb, C., & Grodsky, E. (2010). Racial-ethnic differences at the intersection of math course-taking and achievement. *Sociology of Education, 83,* 248–270.
Rosenbaum, P. (2002). *Observational studies*. New York: Springer.
Rosenbaum, J. E., & Harris, L. E. (2001). Low-income families in their new neighborhoods. *Journal of Family Issues, 22,* 183–210.
Rosenbaum, P., & Rubin, D. B. (1983a). Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society, Series B, 45,* 212–218.
Rosenbaum, P., & Rubin, D. B. (1983b). The Central role of the propensity score in observational studies for causal effects. *Biometrika, 70,* 41–55.
Rosenbaum, P., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician, 39*, 33–38.
Royal Society. (2006). *Taking a leading role—scientists survey*. London: The Royal Society.
Rubin, D. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology, 66,* 688–701.

Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *Annals of Statistics, 6,* 34–58.

Rubin, D. (1980). Discussion of "randomization analysis of experimental data in the fisher randomization test" by Basu. *Journal of American Statistical Association, 75,* 591–593.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized inference*. Boston: Houghton Mifflin.

Simpkins, S. D., Price, C. D., & Garcia, K. (2015). Parental support and high school students' motivation in biology, chemistry, and physics: Understanding differences among latino and caucasian boys and girls. *Journal of Research in Science Teaching, 52,* 1386–1407.

Simpson, R. D., & Oliver, J. S. (1990). A Summary of major influences on attitude toward and achievement in science among adolescent students. *Science & Education, 74,* 1–18.

Stake, J. E., & Mares, K. R. (2001). Science enrichment programs for gifted high school girls and boys: Predictors of program impact on science confidence and motivation. *Journal of Research in Science Teaching, 38,* 1065–1088.

Staniec, J. F. O. (2004). The effects of race, sex, and expected returns on the choice of college major. *Eastern Economic Journal, 30,* 549–562.

Steele, C., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69,* 797.

Stuart, E., & Rubin, D. B. (2008). Best practices in quasi-experimental designs: Matching methods for causal inference. In J. Osborne (Ed.) (pp. 155–176). Thousand Oaks: Sage Publications.

Summers, M. F., & Hrabowski, F. A. (2006). Preparing minority scientists and engineers. *Institutions, 17,* 18.

Tai, R. H., Liu, C. Q., Maltese, A. V., & Fan, X. T. (2006). Planning early for careers in science. *Science, 312,* 1143–1144.

Tyson, W., Lee, R., Borman, K. M., & Hanson, M. A. (2007). Science, technology, engineering, and mathematics (STEM) pathways: High school science and math coursework and postsecondary degree attainment. *Journal of Education for Students Placed at Risk, 12,* 243–270.

von Hippel, P. T. (2007). Regression with missing Y's: An improved strategy for analyzing multiply imputed data. *Sociological Methodology, 37,* 83–117.

Wai, J., Lubinski, D., Benbow, C. P., & Steiger, J. H. (2010). Accomplishment in science, technology, engineering, and mathematics (STEM) and its relation to STEM educational dose: A 25-year longitudinal study. *Journal of Educational Psychology, 102,* 860.

Walker, C. L., & Rakow, S. J. (1985). The status of Hispanic American-students in Science—attitudes. *Hispanic Journal of Behavioral Sciences, 7,* 225–245.

Wang, X. (2013). Why students choose STEM Majors: Motivation, high school learning, and postsecondary context of support. *American Educational Research Journal, 50,* 1081–1121.

Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology, 25,* 68–81.

Xie, Y., & Killewald, A. (2012). *Is American Science in Decline?*. Cambridge: Cambridge Harvard University Press.

Yoder, B. L. (2012). *Engineering by the Numbers*. St. Joseph, MI: American Society of Agricultural and Biological Engineers.

Yonezawa, S., Wells, A. S., & Serna, I. (2002). Choosing tracks:"Freedom of Choice" in detracking schools. *American Educational Research Journal, 39,* 37–67.

Zhou, X. (2005). The institutional logic of occupational prestige ranking: Reconceptualization and reanalyses. *American Journal of Sociology, 111,* 90–140.