COMPETITON, COURSEWORK AND CAREERS: UNDERSTANDING DIVERSITY GAPS IN TECHNOLOGY

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL OF EDUCATION

AND THE COMMITTEE ON GRADUATE STUDIES

OF STANFORD UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

JUNE PARK JOHN

JUNE 2018

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Martin Carnoy, Primary Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Christine Wotipka, Co-Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Prashant Loyalka

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Muriel Niederle

Approved for the Stanford University Committee on Graduate Studies.

Patricia J. Gumport, Vice Provost for Graduate Education

This signature page was generated electronically upon submission of this dissertation in electronic format. An original signed hard copy of the signature page is on file in University Archives.

Acknowledgments

I would first like to thank my dissertation committee—Professors Martin Carnoy, Christine Min Wotipka, Prashant Loyalka and Muriel Niederle. They have been instrumental in shaping my scholarship and moving my dissertation forward. Professor Martin Carnoy has been a model of someone who always questions and learns more as a result. I am humbled by the time he spends helping students and knows that he cares for me not only as a scholar but as a person. Professor Christine Min Wotipka has had incisive comments and advice for me throughout the past five years. I greatly respect her as a person and her commitment to equity. Professor Prashant Loyalka has always been gracious with his time. I have learned about navigating through academia as well as what direct developing country research looks like through his example. Professor Muriel Niederle's work inspired my second dissertation paper. I have greatly benefited from her time and expertise. I would also like to thank Professor Roy Pea for serving as my dissertation chair.

For the dissertation, I am grateful to the many people who helped me plan and execute my studies. Professors Doug Bernheim, Al Roth and Charlie Sprenger were influential in the direction of the research design of my second paper. I am also deeply grateful to Dr. K.J. John and Mary John for providing access to the schools in the study, as well as the principals and the "penolong kanan" who gave me permission to enter their schools and conduct these studies during classroom time. I could not have run the study without the hard work of Kimberly Gan and Sheng Wei Chiam, who provided research assistance with flexibility and great attitudes. I also benefited from Erin Fahle's comments and the community of CEPA and behavioral economic seminar participants in writing this paper. For the third paper, I am so thankful for Perry Weirich at the Texas

Education Agency and his prompt replies to my data requests. I heavily leaned on Professor sean reardon as I began conducting the analyses for this paper and appreciate his generosity with his time.

During my time at Stanford, I have greatly benefited from many people. I have appreciated the friendships, advice, and support of other students, including Erin Fahle, Jenna Finch, Jing Liu, Graciela Perez, Sade Bonilla, Jamie Johnston, Lindsay Fox, Mana Nakagawa and other students who shared their knowledge and experience, especially participants in the CEPA and IEI seminars. Sade, in addition to your triple difference wisdom, your family and hospitality have enabled me to complete this last stretch. I am also thankful for Professor Eric Bettinger for his time and advice through the years. I thank Laura Wentworth and Bryan Twarek for allowing me to participate in real-world issues pertinent to my dissertation. Outside of Stanford, I have cherished the practical and spiritual support of my friends, particularly at the Palo Alto Vineyard Church.

I could not have started on this path without the sacrifice and perseverance of my family. My parents, Hyunsoo and Sungja Park, had all three children graduate from college although they did not have the opportunity to do so. They instilled in me a fundamental value for education which has obviously lasted through many years. My siblings, Jean and Jake, have helped me grow in character and have patiently supported this long academic journey. I am grateful for the family I married into, stemming from K.J. and Doreen John, who have played crucial roles in my research and have rooted for me for many years. I am so thankful for Johann John, who has endured this all and more with me. You have more confidence in me than I do and encourage me towards

excellence in so many areas of life. James Park John motivated me to finish my second paper before he was born and is a daily joy. I thank God who enabled this all.

The research reported here was supported by the Freeman Spogli Institute of International Studies through the Mentored Global Research Fellowship and the Institute of Education Sciences, U.S. Department of Education, through Grants R305B090016 and R305B140009 to the Board of Trustees of the Leland Stanford Junior University. The opinions expressed are those of the author and do not represent views of the Institutes or the U.S. Department of Education.

Table of Contents

List of Tables

Paper 1: Is Expanding Universities' STEM Programs Enough? The Case of Computer Science Education, Employment, Gender, and Race/Ethnicity in Silicon Valley, 1980-2015

Paper 2: Gender differences and the effect of facing harder competition

Paper 3: The (unintended) effects of allowing Computer Science to count as a mathematics graduation requirement in Texas

List of Figures

Paper 1: Is expanding universities' STEM programs enough? The case of computer science education, employment, gender, and race/ethnicity in Silicon Valley, 1980-2015

Paper 2: Gender differences and the effect of facing harder competition

Paper 3: The (unintended) effects of allowing Computer Science to count as a mathematics graduation requirement in Texas

Introduction

Gender and racial gaps ("diversity gaps") in technology[1] are a persistent problem, starting early and continuing in the labor force. As technology plays an increasingly important role inside and outside the classroom, it is important to address diversity gaps in this rapidly expanding and influential field.

Many studies have examined the extent of race and/or gender gaps within Science, Technology, Engineering & Math (STEM) fields. For example, women comprise 48% of the U.S. workforce but just 24% of STEM workers (Beede et al., 2011). Similarly, the representation of Blacks and Hispanics in the STEM workforce is about half of their representation in the overall workforce (Landivar, 2013). Even Asians, who have become increasingly represented in STEM, appear to encounter a "glass ceiling" in upper management and executive levels (U.S. Equal Employment Opportunity Commission, 2016).

There is some evidence that these diversity gaps may be particularly severe in technology. Although the proportion of females earning degrees in STEM has risen in most STEM fields, computer science remains one of the STEM fields in which the proportion of women is lowest (National Science Foundation, 2017). Furthermore, the gender wage gap is smaller in STEM compared to non-STEM jobs, but Computer & Math has the highest gender wage gap among the STEM fields (Beede et al., 2011). In addition, while there has been increasing representation of Blacks and Hispanics in STEM degrees (NSF), these minorities have relatively low representation in computer occupations compared to in other STEM occupations (Landivar, 2013).

[1] This dissertation focuses primarily on computer science within the broader category of technology.

Many factors appear to contribute to these diversity gaps. I adapt Byrnes & Miller's (2006) framework to categorize these factors into internal factors which are situated within external factors that contribute to future decisions to pursue STEM and more specifically technology.

Internal factors include demographics such as race and gender, but also beliefs and behavioral characteristics of individuals. In addition, demographic characteristics may interact with these other characteristics. Research shows that there are racial/gender differences in beliefs such as confidence or interest in STEM, which could contribute to differential persistence in STEM (Cech et al., 2011; Riegle-Crumb et al., 2010; Rosson et al., 2011).

A growing body of research has begun to explore whether individual behavioral characteristics may play a significant role in participation in STEM. For example, competitive behavior is associated with choosing a science track in secondary school (Buser, Niederle & Osterbeek, 2014). Thus, competitive behavior is a promising area of research to understand diversity gaps in STEM. Policies can then address these personal characteristics to spur greater participation from females and minorities in STEM.

Along with internal characteristics associated with technology attitudes and achievement, external factors may contribute to the observed diversity gaps. The internal factors described earlier exist within these external factors, which include access, stereotypes, role models, and physical/psychosocial support. The third paper of this dissertation focuses on access, yet it is important to recognize the potential impact of these other external factors on student participation in STEM. These factors create environments that could be beneficial or detrimental to students' STEM decisions. For example, cultural beliefs, stereotypes, role models, physical portrayals and perceived similarity to others can influence students' beliefs in STEM (Cheryan & Plaut, 2010; Cheryan et al., 2011; Cheryan et al., 2015; Correll, 2011).

Access to STEM is a critical component of addressing diversity gaps. Student access to STEM courses in secondary school varies by race and socio-economic status (U.S. Department of Education Office for Civil Rights, 2016). Opportunities to take AP CS courses are especially limited; for example, only 2,100 of the 42,000 high schools in the US were certified to teach the AP computer science course in 2011 (Exploring Computer Science, 2017). Qualitative evidence indicates that schools that offer AP CS courses are unequally distributed by race/socioeconomic status (Margolis et al., 2003).

This framework shows the complexity of the factors associated with decisions to pursue STEM. Internal factors may interact with each other and with the external environment they are situated in, and external factors may help shape some of the internal factors such as confidence. This dissertation will address different areas of this framework, with two papers specifically focusing on computer science. The first paper describes differences by race and gender in higher education and employment in computer science (i.e. outcomes of the internal and external factors that influence STEM participation). The second paper explores an internal factor: how gender differences in competition play a role in differences in STEM choices. The third paper analyzes a policy that attempts to overcome an external factor, barriers to taking STEM courses in high school.

Together, these papers provide different levels of analysis of gender and racial inequalities, including regional trends, individual decision-making and state-level policy.

I use different datasets and methods for each of these papers, which provide insights into several important issues around diversity in technology and education.

The first paper provides an overview of the racial and gender trends in computer science higher education and the technology labor force in the Silicon Valley, exploring whether pipeline and wage arguments can help explain these trends. I use publicly available data from the census, Integrated Postsecondary Education Data System (IPEDS) and the Open Doors surveys (Open Doors) for these analyses. While there has been dramatic demographic change in technology, racial and gender disparities still remain. Main findings suggest that different barriers exist for different race-gender groups. White females' share of degree completions in computer science declined from 1985 to 2015; this trend is mirrored in their declining share among programmers, which supports a pipeline argument that there are too few white females majoring in STEM. Patterns associated with other groups do not necessarily align with the pipeline argument. The most salient example is that Hispanic males have become an increasingly large proportion of degree completers in computer science, yet their representation in the programmer labor force has declined.

The second paper uses an experiment to measure gender differences in a competitive STEM situation using a behavioral economic framework. Results confirm that males are more competitive than females, which is one explanation for gender differences in STEM; however, males appear to lower their performance when facing harder competition while females are unaffected by the level of competition. These somewhat surprising results show that females may not necessarily be disadvantaged in more competitive STEM settings, even though they may not choose into them. This suggests that STEM policies that encourage or compel participation may not be detrimental to female achievement.

The third paper examines the effects of recent K-12 CS state-level policy. Allowing CS to count towards an academic high school graduation requirement has become an accepted policy in the majority of states, yet it is unclear what, if any, effects this policy has on student enrollment in CS courses. It is important to examine whether there are differences by gender or race to see whether this policy improves or contributes to existing diversity gaps in CS. I use a triple difference identification strategy, leveraging the cohort-based adoption of Texas' CS graduation policy and concurrent trends in enrollment in an unaffected subject (AP Psychology). Enrollment in AP CS appears unaffected for most student groups, and negative for White males and females. In addition, I conduct spillover analyses on mathematics courses, to see whether math course enrollment goes down (negative spillover) or goes up (positive spillover) as a result of the CS policy. Although there are negative spillovers for certain groups in AP Calculus BC and AP Statistics, there are positive spillover effects for certain Latina and Black females in Pre-calculus but not in AP Calculus AB, AP Calculus BC or AP Statistics. These results underscore that this policy does not have the intended effects of increasing participation in CS.

References

Beede, D.N., Julian, T.A., Langdon, D., McKittrick, G., Khan, B. & Doms, M.E. (2011). "Women in STEM: A Gender Gap to Innovation." SSRN Scholarly Paper. Rochester, NY: Social Science Research Network.

Buser, T., Niederle, M., & Oosterbeek, H. (2014). Gender, Competitiveness, and Career Choices. *The Quarterly Journal of Economics*, *129*(3), 1409–1447.

Byrnes, J.P. & Miller, D.C. (2007). The relative importance of predictors of math and science achievement: An opportunity–propensity analysis. *Contemporary Educational Psychology, 32*, 599-629.

Cech, E., Rubineau, B., Silbey, S. & Seron, C. (2011). "Professional Role Confidence and Gendered Persistence in Engineering." *American Sociological Review, 76* (5), 641–666.

Cheryan, S., Master, A. & Meltzoff, A.N. (2015). Cultural stereotypes as gatekeepers: increasing girls' interest in computer science and engineering by diversifying stereotypes. *Frontiers in Psychology*, Vol. 6 Art. 49.

Cheryan, S. & Plaut, V. (2010). Explaining Underrepresentation: A Theory of Precluded Interest. *Sex Roles, 63,* 475-488.

Cheryan, S., Siy, J.O., Vichayapai, M., Drury, B.J. & Kim, S. (2011). Do Female and Male Role Models Who Embody STEM Stereotypes Hinder Women's Anticipated Success in STEM? *Social Psychological and Personality Science 2*(6), 656-664.

Correll, S. (2001). Gender and the Career Choice Process: The Role of Biased Self-Assessments. *American Journal of Sociology, 106*(6), 1691-1730.

Exploring computer science. (2017). CS Education Statistics. Retrieved May 10, 2017, from http://www.exploringcs.org/resources/cs-statistics

Landivar, L. (2013). Disparities in STEM Employment by Sex, Race, and Hispanic Origin. American Community Survey Reports. Retrieved from https://www.census.gov/prod/2013pubs/acs-24.pdf

Margolis, J., Holme, J. J., Estrella, R., Goode, J., Nao, K., & Stumme, S. (2003). The computer science pipeline in urban schools: Access to what? For whom. *IEEE Technology and Society Magazine*.

National Science Foundation, National Center for Science and Engineering Statistics. (2017). Women, minorities, and persons with disabilities in science and engineering: 2017. Retrieved from www.nsf.gov/statistics/wmpd/

Riegle-Crumb, C. & Grodsky, E. (2010). Racial-ethnic differences at the intersection of math course-taking and achievement. *Sociology of Education, 83*(3), 248-270.

Rosson, M., Carroll, J.M. & Sinha, h. (2011). Orientation of Undergraduates Toward Careers in the Computer and Information Sciences: Gender, Self-Efficacy and Social Support. *ACM Transactions on Computing Education*, Vol. 11, No. 3, Article 14

U.S. Department of Education Office for Civil Rights (2016). 2013-2014 Civil Rights Data Collection: A first look. Retrieved from https://www2.ed.gov/about/offices/list/ocr/docs/2013-14-first-look.pdf

U.S. Equal Employment Opportunity Commission. (2016, May). Diversity in high tech. Retrieved from https://www.eeoc.gov/eeoc/statistics/reports/hightech/upload/diversity-in-high-tech-report.pdf

Paper 1: Is expanding universities' STEM programs enough? The case of computer science education, employment, gender, and race/ethnicity in Silicon Valley, 1980-2015

1. Introduction

Science, Technology, Engineering & Math (STEM) higher education is central to the worldwide debate on whether training more women and disadvantaged minorities in STEM majors can help close gaps in job opportunities and income.[1] Given the rapidly growing technology sector in developed and some developing countries, graduating more women and underrepresented minority STEM from universities could be productive in equalizing gender wage differences and increase minority social mobility. Certain fields, such as software development, seem to offer especially great possibilities for higher wages and resulting social mobility (Xue and Larson, 2015).

However, such a scenario implicitly assumes that regardless of gender or race/ethnicity, STEM graduates have equal access to higher paying high technology jobs and are therefore similarly motivated to take university education leading to STEM professions. This paper investigates this assumption by examining race-ethnic/gender differences in degree attainment and the employment and wages of college graduates over the past 35 years in one prototypical example of high-tech industry: Silicon Valley, California. Although Silicon Valley has its own peculiarities (Saxenian, 1994), its labor market practices are representative of practices in the high technology industry globally (Benner, 2008).

The six counties in the Bay Area surrounding San Francisco employ over 400,000 (330,000 manager and professional) workers in the technology industry. There have been dramatic changes in the racial and gender composition of this sector's workforce since

[1] For example, see Burke and Mathis, 2007 and Marginson et al, 2013. For the U.S., see, Xie et al, 2015. UNESCO has also focuses on gender gaps in access to STEM jobs and job mobility: http://www.unesco.org/new/en/natural-sciences/priority-areas/gender-and-science/improving-measurement-of-gender-equality-in-stem.

1980, and significant increases in the percent with advanced degrees. In the 1980s, technology companies seemed to have shifted in hiring white females in professional and management jobs (Carnoy, 1996). But after 1990, this shift ended and was superseded by large relative increases in Asian male and (less so) Asian female employment, mainly non-US citizens. Hispanics (the largest minority group in California) and blacks have made little or no inroads into the industry despite, in the case of Hispanics males, significant increases in the number achieving computer science degrees.

A growing body of research suggests that while gender and race disparities have been widely documented in STEM (Beede et al., 2011; Landivar, 2013; U.S. Equal Employment Opportunity Commission, 2016), these gaps may be particularly severe in computer science, one of the major disciplines in STEM. Although the proportion of females earning degrees in STEM has risen in most STEM fields, computer science remains one of the STEM disciplines with the lowest proportion of women (National Science Foundation, 2017). Furthermore, the gender wage gap is smaller in STEM compared to non-STEM jobs, but the computer and mathematics job category has the highest gender wage gap among the STEM fields (Beede et al., 2011). In addition, while there has been increasing representation of blacks and Hispanics in STEM degrees (National Science Foundation, 2017), these minorities have relatively lower representation in computer occupations than in other STEM occupations, specifically the largest of the STEM occupations, software developer ("programmers") (Landivar, 2013).

This paper provides a broad overview of race and gender employment patterns in the technology industry from 1980 to 2015. It then analyzes "gaps" between the supply of potential and actual programmers in the labor force by gender and race/ethnicity as a heuristic for understanding whether higher education or other factors, such as industry hiring preferences or employment/wage discrimination, may be responsible for these gaps.

Although we cannot provide a causal analysis of race/gender gaps in the technology industry, we explore two common explanations for such differences—the supply of potential programmers (higher education computer science graduates) and trends in the wages of employed programmers. We analyze the programmer pipeline as a possible factor for demographic differences in the programmer workforce. Many papers have used or critiqued a "leaky pipeline" metaphor for STEM workers (Metcalf, 2010, for a review). Whether the pipeline is the most appropriate metaphor, gender and racial disparities have been observed in the supply of STEM labor, "explained" by factors such as attitudes and exposure to technology in junior high and high school (Google Inc. & Gallup Inc., 2016; Riegle-Crumb et al., 2011; Quinn & Cooc, 2015) and enrollment and persistence in STEM courses in higher education (Katz et al., 2003; Griffith, 2010). Higher education is the most common transition into the professional workforce and thus a critical juncture for understanding its demographics. Keeping this in mind, we examine race and gender trends in students who major in computer science, which we use as a proxy for the numbers of potential programmers by race and gender. We also examine wage differences as a factor in observed demographic employment trends.

Our results show that professional and management employment in the technology industry is becoming increasingly male—contrary to trends in the 1980s—and increasingly more highly educated (advanced degrees), increasingly Asian, and increasingly non-citizen, and is characterized by pervasively low representation of

Hispanics and blacks. These trends are even more dramatic among programmers. Our analysis suggests that the masculinization and the continued low representation of California's largest minority group, Hispanics, and of blacks may have different explanations. Although there is some evidence of gender wage gaps, the computer science major in U.S. higher education is also becoming increasingly masculinized, suggesting that gender employment trends could reflect pipeline issues. To the contrary, the supply of Hispanic male computer science majors has increased significantly, but not their employment in the Valley's programmer labor force. This suggests that low employment of Hispanics may be due to employer preferences, not pipeline effects.

Section 2 describes the data used in the analyses. We then build our analysis of the demographic changes that took place in employment in the technology industry. First, we undertake a broad overview of the demographic trends in the Silicon Valley technology industry with comparisons to other major industries in the same region to show how the technology industry has systematically less female employment than other industries (Section 3). Second, we analyze demographic trends in computer science higher education graduates and in programmers employed in the Valley's labor force to show whether pipeline effects may or may not influence employments in this key technology occupation (Section 4). Third, in Section 5, we analyze wage trends for white versus Asian male and female programmers—citizens and non-citizens— to better understand how higher education supply interacts with possible employer preferences and wage patterns in affecting employment patterns. Section 6 discusses the significance of these findings.

2. Data

To construct the dataset of race and gender in computer science higher education, we combined data from the Integrated Postsecondary Education Data System (IPEDS) with data from the Open Doors surveys (Open Doors). We use IPEDS completion data for computer science degrees, including race and gender data, from 1985-2015 (earliest available data is 1985) for California[2] and include national data for context. We then combine these data with country of origin data from the Open Doors surveys for non-resident alien students whose race is not identified. This combination creates degree completion numbers by race and gender for computer science undergraduate and graduate degrees.

The labor force analyses use microdata from the 5% sample in 1980, 1990 and 2000 U.S Censuses and the 1% samples of the 2010 and 2015 American Community Surveys[3]. To examine the technology sector, we limit the sample population to the geographic region most salient in technology: Silicon Valley.[4] In addition, we limit the sample to full-time full-year (FTFY) workers in the labor force, who are defined as individuals who usually work 35 hours a week or more and worked at least 50 weeks in the previous year.[5] All analyses use individual weights.

In the introductory set of analyses, we estimate patterns and trends across major (exclusive but not exhaustive) industries in the Silicon Valley: manufacturing, high

---

[2] We recognize that the Silicon Valley technology labor market may be a national market, but restrict the higher education analysis to California for comparability.
[3] The long form of the population census ceased in 2000.
[4] This includes respondents in the following counties: Alameda, Contra Costa, San Francisco, San Mateo, Santa Clara and Santa Cruz. The Silicon Valley is not an official government designation and thus we use an inclusive geographic region in our analyses.
[5] The National Center for Education Statistics uses these definitions of full-time and full-year.

services, and technology[6]. We also categorize types of occupations into managers and professionals,[7] leaving out other occupation categories. In the second stage, we restrict our analyses to one specific occupation, software developers[8] ("programmers"). Although there have been several re-classifications of technology occupations in the census over time, the definition of the programmer occupation has remained stable since the census began recording information on technology professions in 1970, and it is easily comparable across years (Beckhusan, 2016). Programmers are part of the professional occupation category, and they span across industries.

Research suggests that gender wage gaps are due largely to differences between occupations or industries rather than within. We therefore use this narrow occupation to minimize potential differences between occupations and obtain a more conservative estimate of any wage gaps (Petersen & Morgan, 1995). Because of low numbers of observations in wage data for Hispanics and blacks, our analysis of the wage data is restricted to whites and Asians. Hourly wages from the U.S. Census and American Community Surveys are restricted to positive wages (i.e. reported wages of 0 are dropped) of full-time, full-year workers and are constructed by dividing the annual

[6] We define an industry as belonging to the technology industry if the industry is listed as "Computers and related equipment" (#322), "Radio, TV, and communication equipment" (#341), "Electrical machinery, equipment, and supplies, nec" (#342), "Guided missiles, space vehicles, and parts" (#362), "Scientific and controlling instruments" (#371), "Computer and data processing services" (#732), "Engineering, architectural, and surveying services" (#882) or "Research, development, and testing services" (#891) in the harmonized industry variable (ind1990). Manufacturing industries were industries with the codes 100-392 in the harmonized industry variable (ind1990), excluding those in the computer category. High services industries were industries with the codes 700-712, 721,732, and 812-893 in the harmonized industry variable (ind1990), excluding those in the computer industry.

[7] Occupations are categorized as Manager with the codes 004-022 in the harmonized occupation category (occ1990). These do not include management-related occupation such as accountants or HR specialists and include executives (there were too few executives to be a separate category). Occupations are categorized as Professionals with the codes 043-200 (Professional Specialty list), 229 (programmers), and 23-37 (Management-Related occupations) in the harmonized occupation category (occ1990). All other occupations are categorized as "Other" in these analyses (includes occupations such as cook, bookkeeper, waiter, office clerk, etc.).

[8] We use the harmonized occ1990 occupation category of 229 (programmers) which is defined as computer software developers and computer scientists/analysts (occ1990).

income from work in that occupation by the number of weeks worked per year and number of hours worked per week.[9]

The absolute percentages of workers employed by race and gender over time provide one important perspective on the demographic trends in the technology industry. However, these percentages do not account for the overall representation of each demographic group within higher education or within the entire labor force. We calculate ratios for each race-gender group of the group's representation in the occupation or completing a CS degree relative to its representation in the labor force or in the entire spectrum of higher education, based on the ratio of representation measure constructed by Lewis and colleagues (Lewis et al, 2009). A ratio of 1 indicates equal representation, greater than 1 indicates overrepresentation and less than 1 indicates underrepresentation.

3. Employment trends in the technology industry

The Silicon Valley context

In 2015, 2% of the national full-time, full-year labor force was in Silicon Valley, yet 7.5% of the technology labor force and 10% of programmers were employed there. In the U.S., as in other developed countries, the technology labor force is becoming increasingly highly educated--the proportion with advanced degrees is growing rapidly, more so than in other industries. Because the technology industry is such an important employer in Silicon Valley, the labor force in the region is more highly educated than the national average. There are also significant demographic differences between the national and Silicon Valley labor force, in part because of the overall demographics of the Bay Area, which is more Hispanic and Asian than the rest of the country. From 1980 to 2015,

[9] 2015 data uses intervalled wage data unlike the other census years, and thus the average of these intervals is used as the wages for 2015.

Silicon Valley's white and black labor force across all industries declined more than nationally, but Hispanics and Asians increased much more rapidly, reaching 50% of all workers, compared to 30% nationwide.

In addition, the percentage of foreign (non-citizen) workers is particularly high in Silicon Valley. Nationally, the percentage of foreign workers has been increasing, from about 3% in 1980 to 8.5% in 2015, and is even higher in the technology industry (11%). Meanwhile, in Silicon Valley, the percentage of foreign workers was 7% in 1980 and increased to 18% in 2015. Foreign workers comprised nearly a quarter (24%) of the Valley's technology labor force in 2015. Although the countries of origin for technology workers were more broadly distributed in 1980, by 2015, more than 60% came from India (46%) and China (17%). Since foreign technology workers are likely to enter the U.S. labor force through the U.S. university system—especially through graduate STEM education (Carnoy, 1998), the availability of large numbers of post-baccalaureate foreign student graduates may be a key factor in explaining both the attractiveness of hiring foreign (mainly Asian) workers, and the rapid increase in the proportion of advanced degrees among technology professionals and managers. We examine the possible implications of and for Asian-white wage gaps in the programmer occupation.

The technology industry

The three largest employment sectors in Silicon Valley in 1980-2015 were manufacturing, high services, and technology. Together they represented over half of the labor force working in all industries in the Valley during this period. Technology's proportion of total employment increased from 14% in 1980 to 18% in 2015. Managers and professionals proliferated in all three industries, but this proliferation was especially salient in the technology industry, where the number of managers and professionals approximately quadrupled from 1980 to 2015. By 2015, managers made up almost a quarter (24%) and professionals, 58%, of workers in the technology industry.

Several demographic trends mark all three of the Valley's major industries. First, the managerial and professional labor force has become more educated. This was especially true in the technology industry—almost all managers (89%) and professionals (91%) in technology held at least a bachelor's degree, and almost half of managers (45%) and professionals (48%) had graduate degrees. Second, managers and professionals overall in the Valley became more Asian and female. Yet, in contrast to the other two industries, the technology labor force became even more Asian and even less white (Table 1). It also remained less female. In 1980, white males were the largest group in both occupation categories. They represented 75% of managers and 69% of professionals in technology, but these percentages decreased each decade, to 38% of managers and 29% of professionals in 2015.

[Table 1 about here]

The next largest groups of managers and professionals in 1980 were white females and Asian males, followed by Hispanic males and Asian females. But these four groups followed different growth trajectories throughout the next several decades. Asian males and females rapidly increased their proportion in the tech labor force, and Hispanic males increased their participation, but only slightly. On the other hand, white females reversed course: the increase in white female representation among managers and professionals in the 1980s did not continue and began to decline in the 1990s.

The *ratios of representation*—the percentage of a given race/ethnic/gender group of workers in the technology sector compared to its percentage in the overall labor force—shows a consistent narrative with the distribution of race and gender groups in technology (Table 2). Only white males, Asian males and Asian females were overrepresented relative to their representation in the workforce during any of the years of analysis. White male managers and professionals were over-represented among managers and professionals (1.56 and 1.44, respectively) in 1980. This figure fluctuated somewhat in the next 35 years, but ended up higher for white male managers and lower for white male professionals in 2015. Meanwhile, the ratio of representation for Asian males and females among managers and professionals steadily increased, and white female, Hispanics and blacks remained underrepresented throughout the period.

[Table 2 about here]

These two methods of describing the racial and gender composition show a dramatic change in the technology labor force over the past several decades. The percentage of managers and professionals who were white males declined, while the percentages and ratios of representation of Asians greatly increased, even more so among professionals than among managers. Although their absolute numbers increased from 1980-2015, Hispanics and blacks had a consistently low presence as managers or professionals in technology, especially when considering their overall representation in the labor force.

4. Education and employment of programmers

Undergraduate and graduate degrees in computer science

One possible explanation for these changes in high tech employment is trends in the racial and gender demographics of STEM higher education programs. This section discusses these dynamics for undergraduate and graduate computer science (CS) degree completions in California, the main region supplying the pool of potential programmer candidates in Silicon Valley. The analyses provide a description of the potential programmer pipeline by race and gender immediately prior to entry in the labor force. Although the potential technology labor force is not limited to California, it is likely that the demographic trends in the state's CS programs heavily influence workforce composition in Silicon Valley.

A key factor explaining the increase in Asian employment in the technology industry is the major role that foreign-born graduates have come to play in the supply of high-tech labor. Although international students have remained a relatively low percentage of total degrees earned in the U.S., up from 2.5% in the 1980 to 4.8% in 2015 (Institute of International Education, Inc., 2015), they are increasingly concentrated in the STEM fields, particularly at the graduate level. By 2015, the proportion of graduate degrees in, for example, computer science, earned by foreign-born students was 56% nationally and 60% in California (Appendix A-1).

The countries of origin for these international students have become increasingly concentrated in East and South Asia. The percentage of international students from Asian countries has increased from 29% of all international students in 1980 to 64% by 2015 (Appendix A-2). Just several countries make up the bulk of these students. The most current Open Doors data indicates that over half of all international students (51%) are from China, India and South Korea (Institute of International Education, Inc., 2015).

Trends for undergraduate and graduate international students are similar, but there has been and continues to be a higher concentration of international Asian students in graduate programs (Table 3). In 1986, 37% of international undergraduates were from Asian countries while 55% of international graduate students were from Asian countries. These percentages increased dramatically by 2015, when 60% of international undergraduates and 72% of international graduate students were from Asian countries.

[Table 3 about here]

These data give an overview of the international nature of the higher education population which is then reflected in the U.S. labor force. The potential labor force in computer science includes a sizeable proportion of Asian non-citizens.

We use these data on the countries of origin of international students to construct the race and gender percentages over time for those who completed a degree in computer science in California (Figure 1). We impute race from the countries of origin to calculate the distribution of race among international students and assign race to foreign students, creating overall race-gender percentages of degree completions instead of by citizenship status. These percentages probably represent a lower bound for Asians, since Asians are more likely to pursue CS than other fields. The percentages of different race-gender groups show a more distributed demographic for degree completions than the labor force percentages. Furthermore, there are differences between the racial and gender distributions of bachelor's degree and graduate degree completions, which may reflect the more international population of those who obtain graduate degrees.

[Figure 1 about here]

In California, the number of computer science bachelor's degrees was 2,957 in 1985 and declined through 1995, rose through 2005, then declined sharply before rebounding to 5,518 in 2015 (Table 4). The number of graduate degrees in computer science showed a steadier increase, starting at 764 in 1985, with a small reduction in 2010 before increasing to 2,868 in 2015 (Table 4). These patterns were similar at the national level (Appendix A-1). However, California had greater representation of Asians and Hispanics compared to nationally, particularly at the undergraduate level. Whites represented 63% and Asians, 24% of computer science undergraduate degrees in California in 1985. The percentage of whites nearly halved while the percentage of Asians increased to 32% by 2015, although the percentage of Asian females decreased during this time. The percentage of Hispanics increased from 4% in 1980 to over 17% in 2015 and the percentage of black males increased from 3% to 4%, and the percentage of black females declined.

The racial distribution at the graduate level in California began and remained less white and more Asian than at the undergraduate level. The percentages of Asian males and females almost doubled from 1985 to 2015, and of Hispanic males more than doubled, but from a much smaller base.

In addition, there appears to be a growing gender gap in computer science higher education at the undergraduate level. The percentage of CS undergraduate degrees completed by females fell by nearly half, from 31% in 1985 to 14% in 2015, whereas the percentage in graduate degrees increased slightly, from 23% to 26%, the upward tick driven entirely by an increase of foreign-born females. The ratio of males to females receiving CS undergraduate degrees more than doubled for every racial group from 1985 to 2015 (Table 4). At the graduate level, the male to female ratio increased for each racial group except Asians, but by relatively small amounts.

[Table 4 about here]

The ratios of representation at the undergraduate level underscore the stark contrast between genders beginning in 2005 (Table 5). There was some fluctuation in earlier years, but males of all races were overrepresented whereas females of all races were underrepresented from 2005 to 2015. White males completing computer science undergraduate degrees have been increasingly overrepresented relative to their proportions of overall undergraduate degree completions since 1985, and Asian males' representation fluctuated more but ended higher in 2015 than in 1985. Unlike in the labor force, black males became overrepresented in 1990 and Hispanic males were overrepresented in 1990 and 2005 onwards.

[Table 5 about here]

The ratios of representation in graduate degree completions were similar to the undergraduate trends. One of the main differences is that Asian females were overrepresented throughout 1985-2015. Females from all other races were underrepresented throughout this time period, while males were generally overrepresented. These figures show clear and growing gender gaps in the supplies of higher education graduates in CS, while the representation of Hispanics and blacks increased from 1985-2015.

Programmer workforce

Programmers are a key occupation among technology workers. The number of programmers in Silicon Valley increased dramatically in 35 years, from about 10,000 in 1980 to over 140,000 in 2015. Programmers have always had high levels of education, and the fraction of programmers with graduate degrees increased rapidly, especially after 1990. In 1980, educational attainment was primarily split between those with some college (31%), an undergraduate degree (33%) or a graduate degree (26%). By 1990, half of programmers had undergraduate degrees and a fifth had graduate degrees, and by 2015, programmers were almost all college educated, and roughly equal percentages of programmers held undergraduate (46%) and graduate (48%) degrees. Education levels were similar between genders. However, even in 1980, Asians had much higher levels of education than whites, Hispanics, or blacks, and by 2015, more than half of Asian programmers had graduate degrees (Appendix A-3).

In 1980, the percentage of foreign-born who were working as programmers was generally no higher than their proportion in the overall technology industry back in the 1980s, but a key characteristic of Asians employed as programmers is the enormous increase after 1990 in those who were non-citizens. Of those with undergraduate degrees, beginning in 2000, about 40-50 percent of Asian males and 30-40% of Asian females were non-citizens—these percentages contrast sharply with white male undergraduate degree programmers, who, even in recent years, have been more than 85-90% U.S. citizens or white female programmers, who have been almost entirely US citizens. Of programmers with graduate degrees, about 25-30 percent of white males and 16-25% of white females and more than 50 percent of Asians have been non-citizens since 2000. As we have noted, the proportion of programmers with graduate degrees reached almost half in 2015, which suggests that programmers are being increasingly drawn from a non-citizen, largely graduate degree labor pool (Figure 2). All races except Hispanics experienced an increase in the percentage of foreign-born programmers from 1980 to 2015 (Appendix A-4).

[Figure 2 about here]

Driven in part by the increase in foreign-born Asians in the occupation, a major demographic shift took place in the programmer workforce (Table 6). In 1980, over three-quarters of programmers (77%) were white and 16% were Asian. By 2010, Asians represented 58% of programmers and Whites 35% of programmers, and a high and increasing fraction of these were non-citizens.

[Table 6 about here]

When race and gender groups are examined together, there is a clear transition from white males to Asian males as the dominant group of programmers, with the relative decline in the female programmer labor force also shifting from white to Asian females, and in both female groups, increasingly to non-citizens. The percentage of white females decreased to just 4%, whereas the percentage and the percentage of Asian females more than doubled (14%) during this time. Meanwhile, the low percentages of Hispanic and black males and females declined even further. The male to female ratio increased for every race from 1980 to 2015, but it varied across race.

Another way to measure these trends is the ratios of representation for each race-gender category. These also show the increasing white and Asian, largely male, trend among programmers (Table 7). The ratio of representation increased for white and Asian males but decreased for every other group during this time. In 1980, Asian males were the most overrepresented group (1.81), followed by Asian females (1.66) and white males (1.17). All other groups were underrepresented. The overrepresentation of white and Asian males increased while the overrepresentation of Asian females decreased through 2015.

[Table 7 about here]

Comparisons between higher education and the labor force

The trends in computer science degree completions appear different from employment trends among programmers. Notably, the percentages of both undergraduate and graduate degree completions for Hispanics and blacks were generally higher than their corresponding percentages in the labor force (i.e. programmers with only undergraduate degrees and programmers with graduate degrees), and the gaps between their representation in degrees and in the labor force increased from 1990 to 2015.

In an illustrative exercise, we compare the percentages of race gender groups who obtain degrees in CS (Table 4) and their corresponding percentages in the younger programmer labor force. We restrict the labor force to younger workers (30 years and younger), although similar results hold for other age ranges or when using lagged data (i.e. labor market data from 5 years after higher education data). Hispanic males made up 5% of undergraduate CS degree completions and 3% of employed programmers with only undergraduate degrees in 1990 (Table 8). In 2015, the percentage of Hispanic males rose to 14% of CS undergraduate degree completions, yet represented only 7% of employed programmers with only undergraduate CS degrees in 2015. Similarly, Hispanic males made up about 2% of graduate degree completions and 3% of programmers with graduate degrees in 1990 and increased to 5% of graduate degree completions yet dropped to just over 1% of programmers with graduate degrees in 2015. These percentages are smaller for Hispanic females and blacks, yet generally follow the same pattern of representing a higher proportion of degrees than programmers employed in the technology labor force.

[Table 8 about here]

The differences between the percentages of degree completions in computer science and the labor force suggest that there may be differential rates of CS degree completers entering the programmer occupation. In general, Asians make up a higher proportion of the programmer labor force than their proportion of computer science degrees, whereas Hispanics and blacks represent a larger proportion of degree completions than their proportions in the labor force. Whites have had decreasing representation in both degrees and in the labor force, yet have varied between greater or less representation in undergraduate CS degrees than in the employed programmer labor force. White females have higher representation among younger programmers than among undergraduate CS degree completions. At the graduate level, white females have been a larger proportion of CS degree completers than their proportion in the programmer labor force. These trends indicate that for certain groups, such as Hispanic males, or white females with graduate degrees, getting degrees in computer science does not seem to increase their corresponding proportion into the programmer labor force, and that other groups, such as white females with undergraduate degrees, are declining as a proportion of the programmer labor force as least partly because they are not maintaining their share of undergraduate degrees in computer science.

There are other possible explanations for the observed gaps between the completion percentage and the labor force percentage. For example, the supply of potential programmers extends beyond state or national borders, those who complete CS degrees may not work as programmers, or programmers may not necessarily complete degrees in computer science (Stackoverflow, 2015). There may be distinctions in degree quality that are not reflected in the number of degree completions. However, our analyses suggest that Hispanic males may face barriers to the programmer occupation after higher education, whether these are internal (e.g. choosing not to go into programming based on perceived job difficulty) or external (e.g. facing discrimination in hiring).

5. Programmer wages

As we have shown, except for a dip in employment for whites in 2000-2010, the number of white and Asian programmers generally increased steadily from 1980-2015. These increases were accompanied by increases in wages, but with relatively flat wages during the 2000-2010 period (Table 9).[10] We also showed that, as measured by the numbers of CS degree completers, the supply of potential white and Asian programmers increased from 1985-2015. The fact that wages kept increasing despite increases in supply suggests an increase in demand for programmers in most years.

[Table 9 about here]

Even so, wage patterns appear to differ by race/ethnicity, gender, and education level (Figure 3). At the undergraduate level, white males consistently have had the highest average hourly wage, and, on average, Asian males earn less. The gap between white and Asian males appears to be relatively steady, although the difference is only statistically significant in 1980 and 2000. Meanwhile, white and Asian females appear to have similar wages during this period, although white females have slightly higher (but not statistically significantly different) wages most years. Since white female undergraduate CS degree completers was the only group with decreased number of CS degrees, hence a declining potential supply of programmers, this could have contributed

[10] Unfortunately, because of the small sample sizes for Hispanics and blacks, we are forced to restrict the wage analyses to white and Asian full-time full-year workers. We also limit the sample to 25-44 year olds, include only positive wages, and separate analyses into programmers with undergraduate degrees only and programmers with graduate degrees--this to provide less biased wage comparisons.

to the seemingly higher wages of white females than Asian females. Finally, although the differences were not always significant in prior years, the gender gaps for both whites and Asians became larger and statistically significant in 2015.

[Figure 3 about here]

These overall wage differences suggest that somewhat lower wages for non-citizens in recent years may be one factor that promotes hiring more non-citizen programmers and yet are high enough to create an increasing supply of foreign-born workers with CS degrees. At the same time, if women are paid lower wages, this should increase the demand for women workers, but the wages may not be high enough compared to alternatives to convince U.S. citizen women to major in computer science to prepare for jobs as programmers. U.S. citizen white women may also be less inclined to join a male-dominated field even if wages were somewhat higher for female programmers than alternative work in other fields (See Appendices A-5 and A-6 for more detail).

We also estimate programmer wages for non-U.S. citizens and U.S. citizens, Asian and white for 25-44 year olds. The results suggest that for white males and Asian females, the wages are essentially the same for citizens and foreign workers with either undergraduate or graduate degrees throughout this period (Figures 4a and 4b). This is also true for Asian males with an undergraduate degree. However, for Asian males with a graduate degree, after 2000, non-U.S. citizens earn somewhat less than their citizen counterparts, and, in 2015, considerably less than U.S. citizen whites with graduate degrees. Generally, then, there is some limited support for the notion that firms in the

Valley are hiring non-citizen Asian programmers because they are paid lower wages than U.S. citizen whites, but this evidence is not very strong.

[Figures 4a and 4b about here]

To the contrary, gender wage gaps seem incongruous with representation in the labor force. Females have lower wages than males, yet males greatly outnumber their female counterparts. Although there may be other unobserved factors involved, higher wages for males may indicate greater demand for males over females. Overall, then, the relationships between wage and labor force representation suggest that wage differences provide neither a consistent nor an especially strong explanation for differences in employment of race/gender groups in the programmer labor force, and that we need to look elsewhere to understand these employment patterns.

6. Discussion

In this paper, we have tried to draw insights into whether increasing women's and minority groups' university STEM education can contribute significantly to their employment in high wage high tech jobs—specifically, in rapidly expanding programmer jobs. If universities paid more attention to attracting more women and minorities into STEM majors would this change employment patterns in high tech industry?

Our analysis of graduates by race and gender in computer science shows that the answer may vary by group. White females' share of degree completions in computer science declined from 1985 to 2015, which supports the argument that their rapidly declining share in programming jobs is largely a result of too few white females majoring in STEM. This argument is also supported by the increasing proportion of Asian females in programmer jobs in this period—suggesting that the overall lower presence of females may not be exclusively due to gender bias in hiring. That said, gender gaps among programmers appear to be increasing across all races, including Asians. Further, bachelor degree women earn much lower wages than male programmers of similar age and with the same level of education.

Other groups also do not necessarily align with the pipeline argument. The most salient example is that Hispanic males have become an increasingly large proportion of degree completers in computer science, yet their representation in the programmer labor force has declined. It appears that certain groups such as Hispanic males may face barriers to working as programmers that other groups, such as Asians do not. Lower wages for Asian males compared to whites may partly explain Asians rapidly increasing presence among programmers, and why much of the increase in Asian employment has been in non-U.S. citizen programmers, most with graduate degrees, where the wage gap in recent years has been highest.[11]

Thus, these findings tend to confirm that inducing more women to study computer science and enter the programmer occupation could increase their average wages relative to males in the overall economy. Yet, it may not be easy to induce more women into computer science, since programming and high tech more generally has become increasingly male-dominated. Although many analyses assume that the main problem for young women entering the STEM pipeline—especially into engineering and computer science majors—is math and physical science aversion (for example, Xie et al, 2015) the work climate for women in high tech may be an even more important factor (U.S Equal Employment Commission, 2016).

[11] As mentioned above, non-US citizens may continue to major in CS and accept lower pay than their U.S. citizen counterparts because they face much lower wages should they return home. A more complex question is why US citizen Asian males with undergraduate degrees receive lower wages than their white male counterparts.

Our results also suggest that inducing more disadvantaged minorities to consider computer science may not be the main barrier to employment. Hispanic males are increasingly preparing to enter the computer science labor force by obtaining computer science degrees, thus it appears that they face external barriers to entering the labor force.

Therefore, policies designed to attract white females may need to focus on getting them to major in computer science, but our results suggest that attracting white females into the CS major may also require changes in the culture of the technology industry, especially conveying a serious commitment to hiring professional females, raising wages, and perhaps creating a more female-friendly culture in the industry. Policies designed to employ disadvantaged minority males much more clearly need to focus either on inducing more of them to apply for programming or other tech jobs or convincing employers to hire more disadvantaged minorities with CS degrees. Whether either of these policies can be successful in overcoming these barriers is anyone's guess.

The broader lesson for countries that seek to increase the economic and social mobility of women and disadvantaged groups through higher education policies that increase these groups' CS and other high tech related STEM graduates, is the importance of considering the hiring practices of the technology sector itself. If the technology industry "prefers" certain categories of employees, increasing the supply of graduates who do not fit those categories will not likely result in much change. The other side of this coin is that if the industry is known to be "unfriendly" to certain groups, talented individuals from these groups will be much less likely to respond to incentives to enter programming or other technology professions.

References

Altonji, J. & Blank, R. (1999). “Race and Gender in the Labor Market.” In *Handbook of Labor Economics*, 3143–3259. Elsevier, 1999. http://econpapers.repec.org/bookchap/eeelabchp/3-48.htm.

American Institute for Economic Research. (2014). H-1B Visas: No Impact on Wages | AIER. Retrieved August 3, 2017, from https://www.aier.org/research/h-1b-visas-no-impact-wages.

Beckhusan, J. 2016. American Community Survey Reports. Occupations in Information Technology.

Beede, D.N., Julian, T.A., Langdon, D., McKittrick, G., Khan, B. & Doms, M.E. (2011). “Women in STEM: A Gender Gap to Innovation.” SSRN Scholarly Paper. Rochester, NY: Social Science Research Network.

Benner, C. (2008). *Work in the new economy: Flexible labor markets in Silicon Valley* (Vol. 9). John Wiley & Sons.

Burke, R. & Mattis, M., eds. (2007). Women and minorities in science, technology, engineering, and mathematics: Upping the numbers. Northampton, MA: Edward Elgar.

Carnoy, M. (1998). The Globalization of Innovation, Nationalist Competition, and Internationalization of Scientific Training. *Competition and Change*, 3: 237-262.

Carnoy, M. and Gong, W. (1996). Women and Minority Gains in a Rapidly Changing Local Labor Market: The San Francisco Bay Area in the 1980s. Economics of Education Review, 15(3), 273–87.

Desilver, D. (2015). Growth from Asia drives surge in U.S. foreign students | Pew Research Center. Retrieved August 10, 2017, from http://www.pewresearch.org/fact-tank/2015/06/18/growth-from-asia-drives-surge-in-u-s-foreign-students/.

Fiegerman, S. (2017). “Labor Department Goes after Big Tech for Discrimination”. *CNN Tech*, April 10, 2017. http://money.cnn.com/2017/04/10/technology/labor-department-tech/index.html.

Griffith, A. L. (2010). Persistence of women and minorities in STEM field majors: Is it the school that matters? Economics of Education Review, 29(6), 911–922.

Goldin, C., & Katz, L. F. (2007). The Race between Education and Technology: The Evolution of U.S. Educational Wage Differentials, 1890 to 2005 (Working Paper No. 12984). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w12984.

Google Inc. & Gallup Inc. (2016). Diversity Gaps in Computer Science: Exploring the Underrepresentation of Girls, Blacks and Hispanics. Retrieved from http://goo.gl/PG34aH.

Hyun, J. (2006). *Breaking the Bamboo Ceiling: Career Strategies for Asians* (Reprint edition). New York: HarperBusiness.

Institute of International Education, Inc. (2015). Open doors 2015 "Fast Facts". Retrieved 7 September 2017 from h ttp://www.iie.org/~/media/Files/Corporate/Open- Doors/Fast-Facts/Fast-Facts-2015.pdf?la=en.

Katz, S., Aronis, J., Allbritton, D., Wilson, C., & Soffa, M. L. (2003). Gender and race in predicting achievement in computer science. *IEEE Technology and Society Magazine*, *22*(3), 20–27.

Landivar, L. (2013). Disparities in STEM Employment by Sex, Race, and Hispanic Origin. American Community Survey Reports. https://www.census.gov/prod/2013pubs/acs-24.pdf.

Lewis, J.L., Menzies, H., Najera, E.I. & Page, R.N. (2009). Rethinking trends in minority participation in the sciences, *Science Education*, *93*(6), 961–977.

Marginson, S., Tytler, R., Freeman, B. & Roberts, K. (2013). STEM: Country comparisons. Report for the Australian Council of Learned Academies. Retrieved from http://dro.deakin.edu.au/view/DU:30059041.

Matloff, N. (2013). Are foreign students the 'best and brightest'? Economic Policy Institute Briefing Paper, February 28. http://www.epi.org/publication/bp356-foreign-students-best-brightest-immigration-policy.

Metcalf, H. (2010). Stuck in the Pipeline: A Critical Review of STEM Workforce Literature. InterActions: UCLA Journal of Education and Information Studies, 6(2). Retrieved from http://escholarship.org/uc/item/6zf09176.

National Academy of Sciences, National Academy of Engineering, and Institute of Medicine, Rising above the gathering storm: energizing and employing America for a brighter economic future (Washington, DC: The National Academies Press, 2007), National Science Foundation (2017). “Women, Minorities, and Persons with Disabilities in Science and Engineering: 2017,” National Center for Science and Engineering Statistics. www.nsf.gov/statistics/wmpd/.

Petersen, T., & Morgan, L. A. (1995). Separate and Unequal: Occupation-Establishment Sex Segregation and the Gender Wage Gap. *American Journal of Sociology*, *101*(2), 329–365.

Quinn, D. M., & Cooc, N. (2015). Science Achievement Gaps by Gender and Race/Ethnicity in Elementary and Middle School: Trends and Predictors. Educational Researcher, 44(6), 336–346.

Riegle-Crumb, C., Moore, C., & Ramos-Wada, A. (2011). Who wants to have a career in science or math? exploring adolescents’ future aspirations by gender and race/ethnicity. Science Education, 95(3), 458–476.

Salzman, H., Kuehn, D., & Lowell, B. L. (2013). Guestworkers in the high-skill U.S. labor market: An analysis of supply, employment, and wage trends. Retrieved August 3, 2017, from http://www.epi.org/publication/bp359-guestworkers-high-skill-labor-market-analysis/.

Saxenian, A. (1994). *Regional advantage: Cultural and competition in Silicon Valley and Route 128*. Cambridge, MA: Harvard University Press.

Smith, M. (2016). Computer science for all. Retrieved January 13, 2017, from https://www.whitehouse.gov/blog/2016/01/30/computer-science-all.

U.S. Equal Employment Opportunity Commission (2016). *Diversity in high tech*. Retrieved from https://www.eeoc.gov/eeoc/statistics/reports/hightech/upload/diversity-in-high-tech-report.pdf.

Xie, Y., Fang, M., & Shauman, K. (2015). STEM education. *Annual review of sociology*, *41*, 331-357.

Xue, Y. and Larson, R. (2015). STEM crisis or STEM surplus? Yes and yes. *Monthly Labor Review*, U.S. Bureau of Labor Statistics, https://doi.org/10.21916/mlr.2015.14.

Tables

Table 1. Race and gender percentages by occupation and industry

Manufacturing	Manager					Professional				
	1980	1990	2000	2010	2015	1980	1990	2000	2010	2015
White Male	75.34	61.2	53.47	50.58	38.75	61.35	49.73	41.7	35.9	31.9
White Female	13.99	21.18	21.25	18.03	16.22	20.66	22.96	22.5	12.4	17.65
Asian Male	2.77	6.25	8.44	12.89	20.53	6.64	11.35	14.77	21.52	22.21
Asian Female	1.19	2.91	4.55	7.17	11.08	2.49	6.12	8.87	18.18	17.76
Hispanic Male	4.19	3.92	4.33	4.9	2.92	3.14	4.06	3.88	4.57	4.98
Hispanic Female	0.71	2.36	2.55	2.77	3.47	2.4	1.68	2.38	2.55	1.7
Black Male	1.26	1.03	0.92	0.62	1.14	1.57	2.89	1.36	0.41	0.68
Black Female	0.32	0.71	1.35	0.58	1.44	1.11	0.82	1.3	0.47	0.38
Total	*25,300*	*31,430*	*26,226*	*28,299*	*31,152*	*21,680*	*34,318*	*31,413*	*36,588*	*45,997*

High Services	Manager					Professional				
	1980	1990	2000	2010	2015	1980	1990	2000	2010	2015
White Male	48.16	36.38	33.22	27.42	27.58	46.4	39.33	33.4	24.25	24.37
White Female	34.64	40.78	36.91	30.19	29.2	33.89	35.28	31.43	30.73	28.55
Asian Male	3.15	4.72	5.2	7.94	8.54	4.22	5.52	8.65	11.87	11.73
Asian Female	2.2	4.75	6.86	10.94	12.7	4.32	7.31	10.11	15.37	16.68
Hispanic Male	3.39	2.58	3.63	5.95	4.79	2.59	2.41	3.13	4.21	3.22
Hispanic Female	2.25	3.99	4.77	7.38	7.84	2.37	3.73	4.19	6.07	6.6
Black Male	2.96	2.26	1.87	3.24	2.43	2.22	2.18	2.4	1.77	1.99
Black Female	2.63	4.09	4.6	3.68	3.32	3.32	3.78	3.4	3.03	3.61
Total	*41,860*	*63,881*	*79,598*	*107,248*	*122,287*	*108,880*	*165,937*	*212,168*	*299,836*	*342,862*

	Technology									
	Manager					Professional				
	1980	1990	2000	2010	2015	1980	1990	2000	2010	2015
White Male	74.73	58.2	49.27	43.92	38.39	68.74	55.4	41.8	32.27	29.04
White Female	14.9	23.85	20.76	11.87	13.35	12.48	17.9	12.92	8.43	7.27
Asian Male	3.79	8.25	15.03	24.18	24.92	9.91	14.5	25.63	36.13	39
Asian Female	1.07	2.6	5.36	10.41	13.58	2.21	5.02	10.39	13.38	14.73
Hispanic Male	2.8	2.3	2.74	3.32	3.82	3.25	2.74	3.47	3.75	4.08
Hispanic Female	0.99	1.56	2.35	2.59	2.55	0.97	1.16	1.29	1.02	1
Black Male	0.99	1.54	1.2	0.97	0.48	1.4	2.05	1.27	1.52	1.59
Black Female	0.33	1.24	0.77	0.65	0.48	0.55	0.97	0.79	0.71	0.5
Total	*24,300*	*45,127*	*74,089*	*76,895*	*97,912*	*61,540*	*103,301*	*178,010*	*183,742*	*240,501*

Note: Only full-time (at least 35 hours of work a week) full-year (at least 50 weeks in previous year) workers in labor force (16+ years old). Percentages may not add up to 100% due to other races (left out of table).

Table 2. Ratio of representation of manager/professional occupation categories within the technology industry to the overall labor force

	Manager					Professional				
	1980	**1990**	**2000**	**2010**	**2015**	**1980**	**1990**	**2000**	**2010**	**2015**
White Male	1.56	1.44	1.46	1.73	1.60	1.44	1.37	1.24	1.27	1.21
White Female	0.59	0.97	0.99	0.67	0.83	0.50	0.73	0.61	0.47	0.45
Asian Male	0.71	0.97	1.26	1.54	1.48	1.87	1.70	2.14	2.31	2.32
Asian Female	0.29	0.40	0.60	0.82	1.02	0.60	0.78	1.16	1.05	1.11
Hispanic Male	0.43	0.29	0.30	0.27	0.29	0.50	0.35	0.38	0.30	0.31
Hispanic Female	0.27	0.33	0.42	0.32	0.31	0.26	0.24	0.23	0.13	0.12
Black Male	0.25	0.42	0.40	0.37	0.18	0.35	0.56	0.42	0.58	0.60
Black Female	0.10	0.38	0.26	0.23	0.20	0.17	0.30	0.26	0.26	0.21

Note: Only full-time (at least 35 hours of work a week) full-year (at least 50 weeks in previous year) workers in labor force (16+ years old).

Table 3. Percentage of international students from region of origin, by academic level

	Africa		Asia		Europe		Latin America		Middle East		North America		Oceania		World Total #	
Year	**UG**	**G**	**UG**	**G**	**UG**	**G**	**UG**	**G**	**UG**	**G**	**UG**	**G**	**UG**	**G**	**UG**	**G**
1985-1986	11.3	9.5	37.2	54.8	9.2	11.1	17.7	8.0	17.5	11.4	5.8	4.2	1.3	1.0	149,200	132,430
1989-1990	8.0	5.4	42.7	64.9	12.4	11.0	17.2	6.9	12.3	7.0	6.0	4.1	1.4	0.7	137,560	169,820
1994-1995	5.2	4.0	52.2	64.8	14.7	13.2	13.2	7.2	7.3	5.8	6.1	4.3	1.2	0.7	228,184	195,166
1999-2000	8.2	3.8	47.0	62.2	16.1	14.2	15.2	8.6	7.1	6.1	5.2	4.5	1.1	0.7	249,786	225,383
2004-2005	9.2	4.2	48.6	65.1	13.1	11.6	16.5	8.1	5.8	5.5	5.7	4.9	1.0	0.6	247,255	269,933
2009-2010	7.3	4.1	56.9	68.6	10.3	8.9	12.5	7.3	6.9	6.5	5.1	4.1	0.9	0.5	274,431	293,885
2014-2015	4.7	2.7	59.9	71.8	9.1	7.9	10.2	5.7	12.0	8.4	3.3	2.9	0.8	0.5	398,824	362,228

Notes: For years 1979-80 & 1984-85 Open Doors data do not include breakdown of country of origin by academic level, so 1985-86 data is the earliest year used. Starting in 2009-10, Cyprus & Turkey were re-categorized from Middle East to Europe. However, due to the quality of data from prior to 1995-96 data, it is not possible to re-categorize these countries so Cyprus & Turkey were re-classified as Middle East a 2009-2010 and 2014-2015 in this analysis. Cyprus represents 211 undergrads and 296 graduate students while Turkey represents 3,656 undergraduates and 6,585 graduate students in 2009-10. Cyprus represents 187 undergrads and 155 graduate students while Turkey represents 3,242 undergraduates and 5,357 graduate students in 2014-15.

North America consists of Canada and Bermuda (vast majority is from Canada).

Table 4. Percentage of degree completions in CS, by race and gender (re-coded)

	Bachelor's						
	1985	**1990**	**1995**	**2000**	**2005**	**2010**	**2015**
White Male	44.49	43.68	39.21	30.93	27.78	34.19	30.11
White Female	18.78	11.41	12.01	8.81	5.15	4.9	4.32
Asian Male	14.32	18.05	20.84	27.75	28.11	17.87	24.8
Asian Female	10.13	10.71	10.32	11.63	8.58	4.14	6.82
Hispanic Male	3.01	5.22	5.07	6.02	9.08	11.4	13.97
Hispanic Female	1.35	2.22	1.78	2.4	2.56	2.22	2.56
Black Male	1.72	2.54	2.89	2.83	3.38	3.6	3.73
Black Female	1.16	1.32	1.87	1.54	1.17	1.01	0.64
Unknown	0	4.04	5.53	7.56	13.48	19.20	8.90
total	2957	2798	2479	3506	5585	3594	5518
	Graduate						
	1985	**1990**	**1995**	**2000**	**2005**	**2010**	**2015**
White Male	36.46	41.8	37.3	27.71	23.47	25.25	22.25
White Female	12.48	11.33	8.64	9.9	6.75	6.12	7.09
Asian Male	20.2	26.25	28.94	30.46	34.84	34.64	37.57
Asian Female	8.29	9.52	10.08	17.98	16.92	12.15	16.09
Hispanic Male	1.92	2.4	3.5	4.01	4.98	4.72	5.09
Hispanic Female	1.14	0.99	1.11	2.28	2.02	1.27	1.73
Black Male	1.9	2.15	1.98	2.27	2.6	2.92	3.11
Black Female	0.86	0.88	0.54	1.25	1.07	1.66	1.05
Unknown	0	4.78	7.83	4.13	7.26	10.72	4.85
total	764	1087	1188	1574	2425	2378	2868

Note: for degree completions in CS in California.

Table 5: Ratio of representation of CS degrees to All degrees

	Bachelor's						
Percentage	**1985**	**1990**	**1995**	**2000**	**2005**	**2010**	**2015**
White Male	1.28	1.46	1.54	1.55	1.54	1.95	1.92
White Female	0.50	0.33	0.39	0.33	0.21	0.22	0.22
Asian Male	2.27	2.47	2.21	2.72	2.72	1.71	2.44
Asian Female	1.85	1.48	1.02	0.96	0.67	0.33	0.57
Hispanic Male	0.87	1.35	0.93	0.93	1.42	1.64	1.45
Hispanic Female	0.39	0.50	0.27	0.25	0.24	0.20	0.16
Black Male	0.91	1.58	1.61	1.47	1.99	2.20	1.83
Black Female	0.49	0.60	0.68	0.47	0.38	0.36	0.18

	Graduate						
Percentage	**1985**	**1990**	**1995**	**2000**	**2005**	**2010**	**2015**
White Male	1.02	1.23	1.26	1.10	1.09	1.30	1.20
White Female	0.42	0.37	0.28	0.32	0.25	0.26	0.31
Asian Male	2.65	2.87	2.51	2.66	2.94	2.90	2.99
Asian Female	2.16	1.76	1.23	1.74	1.41	1.00	1.23
Hispanic Male	0.70	0.87	1.09	1.03	1.19	1.07	0.97
Hispanic Female	0.56	0.40	0.32	0.44	0.31	0.17	0.19
Black Male	0.84	1.17	1.08	1.11	1.31	1.42	1.37
Black Female	0.47	0.45	0.22	0.40	0.32	0.51	0.22

Note: Authors calculated numbers in May 2016, constructed from two data sources: IPEDS Completions data (1985, 1990, 1995, 2000, 2005, 2010) and Open Doors data (1985-86, 1989-90, 1995-96, 1999-2000, 2004-2005, 2009-2010) for degree completions in California. Non-citizens in the IPEDS data were re-categorized into these categories using region of origin data from the corresponding Open Doors year. There are “Unknown” racial/gender categories, thus columns do not add up to 100%. The ratios were calculated as the percentage of completions in Computer Science divided by the percentage of completion for all subjects for each race-gender category.

Table 6. Programmers by race and gender percentages

	1980	1990	2000	2010	2015
White Male	56.16	53.99	41.38	31.96	31.51
White Female	20.35	18.38	8.49	3.96	3.74
Asian Male	9.59	11.26	31.5	46.12	44.26
Asian Female	6.07	8.4	11.06	12.92	14.05
Hispanic Male	3.52	3.37	2.58	1.92	2.34
Hispanic Female	0.78	1.04	0.83	0.16	0.5
Black Male	2.15	2.29	1.02	0.18	1.02
Black Female	0.98	0.98	0.67	0.57	0.07
Total	*10,220*	*24,264*	*77,532*	*88,137*	*143,286*

Note: Only full-time (at least 35 hours of work a week) full-year (at least 50 weeks in previous year) workers in labor force (16+ years old).

Table 7: Ratio of representation for programmers compared to the overall labor force

	1980	1990	2000	2010	2015
White Male	1.17	1.34	1.23	1.26	1.32
White Female	0.81	0.75	0.40	0.22	0.23
Asian Male	1.81	1.32	2.63	2.94	2.63
Asian Female	1.66	1.30	1.24	1.01	1.05
Hispanic Male	0.54	0.43	0.28	0.15	0.18
Hispanic Female	0.21	0.22	0.15	0.02	0.06
Black Male	0.54	0.63	0.34	0.07	0.38
Black Female	0.30	0.30	0.22	0.21	0.03

Note: Only full-time (at least 35 hours of work a week) full-year (at least 50 weeks in previous year) workers in labor force (16+ years old).

Table 8. Race and gender percentages of young programmers

	Undergraduate				
	1980	1990	2000	2010	2015
White Male	49.41	44.42	33.85	29.73	29.35
White Female	23.53	24.37	5.32	2.34	5.66
Asian Male	10.59	12.63	43.67	54.41	38.67
Asian Female	10.59	11.02	11.63	5.49	13.6
Hispanic Male	3.53	3.19	2.45	0.61	7.32
Hispanic Female	0	0.61	0	0	0.57
Black Male	1.18	2.67	0.45	0	1.11
Black Female	1.18	1.09	0.14	0	0
Observations	1700	4608	11254	9374	18547
	Graduate				
	1980	1990	2000	2010	2015
White Male	40.62	46.52	25.27	12.2	14.86
White Female	28.13	15.55	2.76	2.49	3.04
Asian Male	12.5	24.01	48.97	68.45	54.25
Asian Female	12.5	11.46	18.35	12.1	18.79
Hispanic Male	3.13	2.46	0.97	0.74	1.22
Hispanic Female	0	0	0	0	0.65
Black Male	0	0	0.05	0	6
Black Female	3.13	0	0	1.03	0
Observations	640	733	6273	7953	14573

Note: Only full-time (at least 35 hours of work a week) full-year (at least 50 weeks in previous year) workers in labor force who are 30 years old and under. This age restriction is to approximate the demographic of degree completers.

Table 9. Wages of programmers, by race and gender

Undergraduate degree

		1980	1990	2000	2010	2015
White Male	Mean	25.77	26.97	39.72	40.00	50.98
	SE	0.60	0.76	1.24	1.80	3.05
	Obs	1,440	4,415	11,324	9,124	14,440
White Female	Mean	23.33	24.40	33.57	30.77	32.40
	SE	1.44	0.55	1.75	3.06	3.17
	Obs	580	2,097	2,321	860	1,667
Asian Male	Mean	21.52	25.36	34.13	36.10	42.54
	SE	1.15	0.88	0.90	1.04	1.86
	Obs	260	1,176	9,594	14,705	20,466
Asian Female	Mean	22.48	22.50	32.24	32.58	29.49
	SE	1.37	1.02	1.08	1.06	1.29
	Obs	140	1,146	3,224	3,730	6,176

Graduate degree

		1980	1990	2000	2010	2015
White Male	Mean	26.76	30.01	40.43	42.03	58.37
	SE	1.02	1.37	1.40	2.17	3.59
	Obs	1,160	2,089	6,222	5,356	10,117
White Female	Mean	18.94	25.25	31.88	32.02	49.59
	SE	1.74	1.21	1.42	2.20	9.00
	Obs	360	435	1,206	411	1,281
Asian Male	Mean	26.03	34.08	38.89	38.90	43.95
	SE	2.52	2.70	0.97	1.41	1.59
	Obs	260	623	10,507	17,510	25,658
Asian Female	Mean	20.52	27.22	34.25	37.58	40.99
	SE	1.67	1.03	0.80	1.22	2.25
	Obs	220	612	3,632	4,871	8,548

Note: Sample for wage analyses are full-time (at least 35 hours of work a week) full-year (at least 50 weeks in previous year) workers in labor force with positive wages using 1999 dollars. The sample is further limited to workers between 25-44 years old.

Figures

Figure 1. Percentage of degree completions in CS in CA, by race and gender

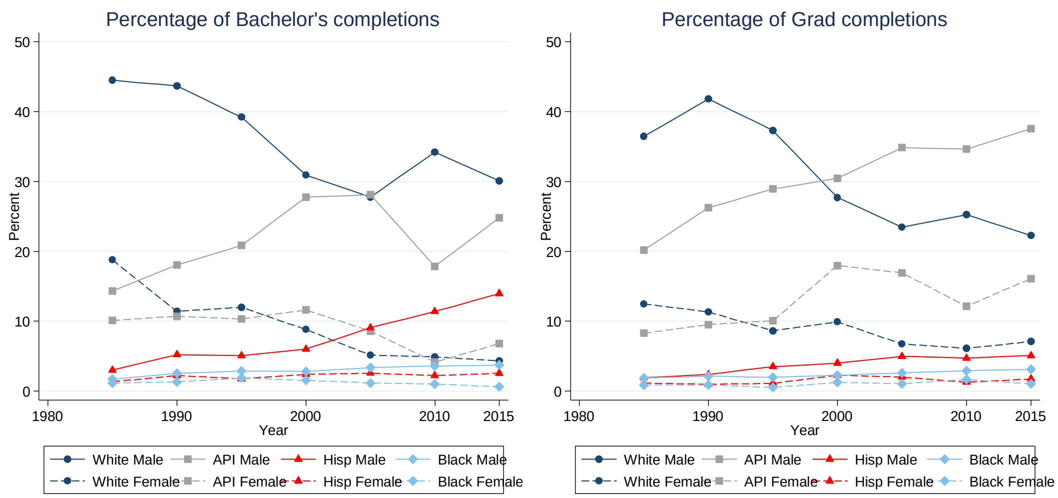
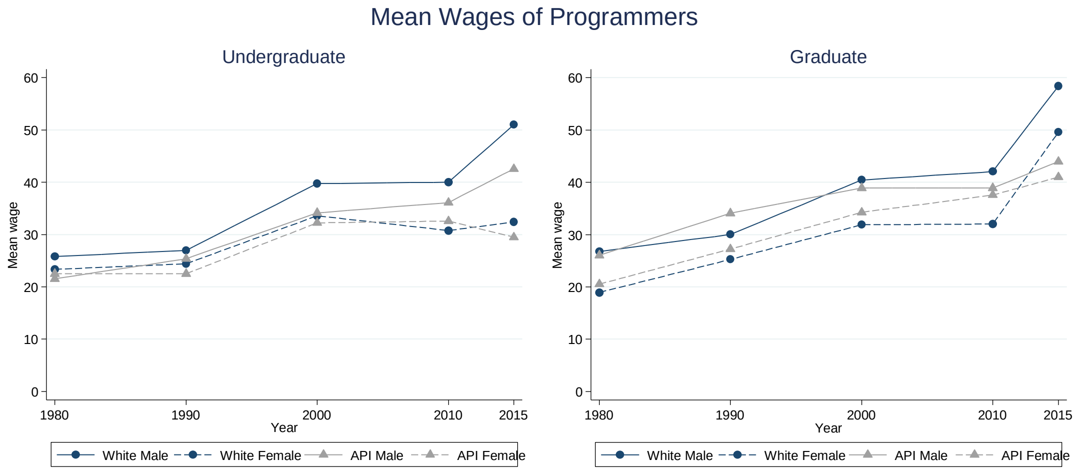


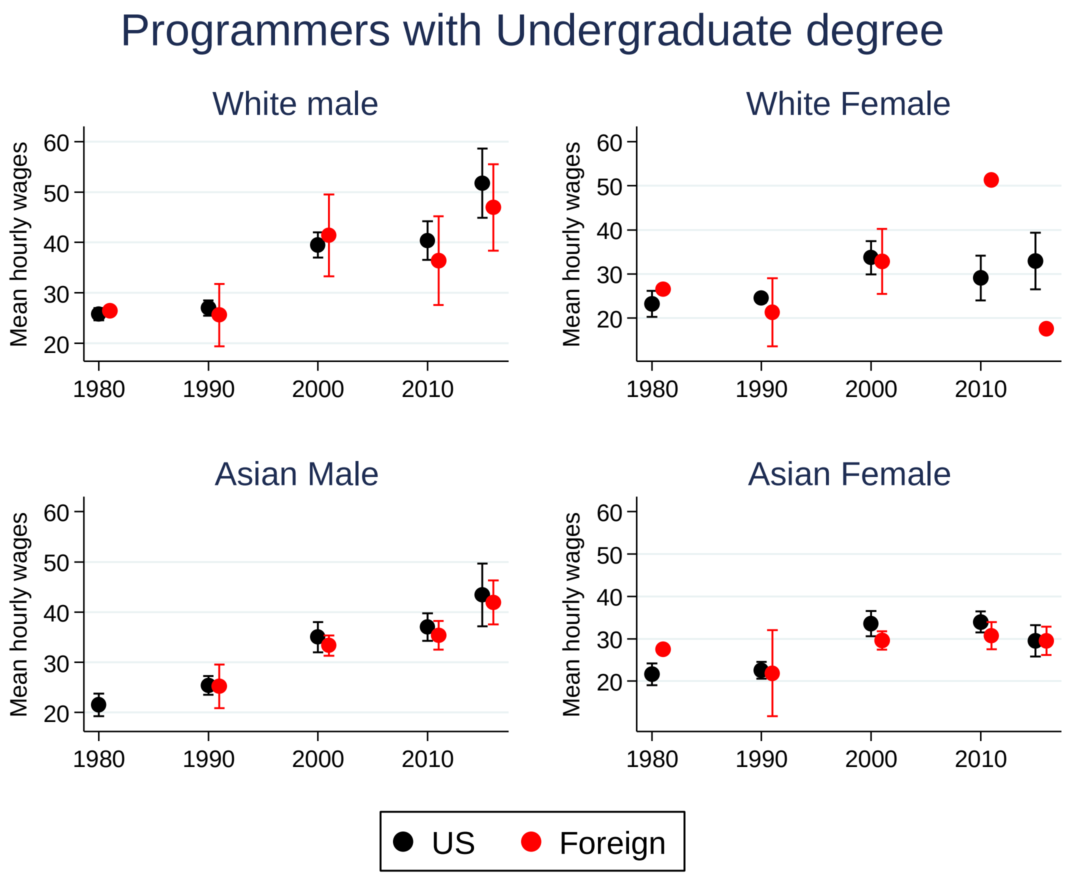
Figure 2. Percent of professionals who are foreign, by degree status.

Figure 3. Mean hourly wages of programmers, by race and gender



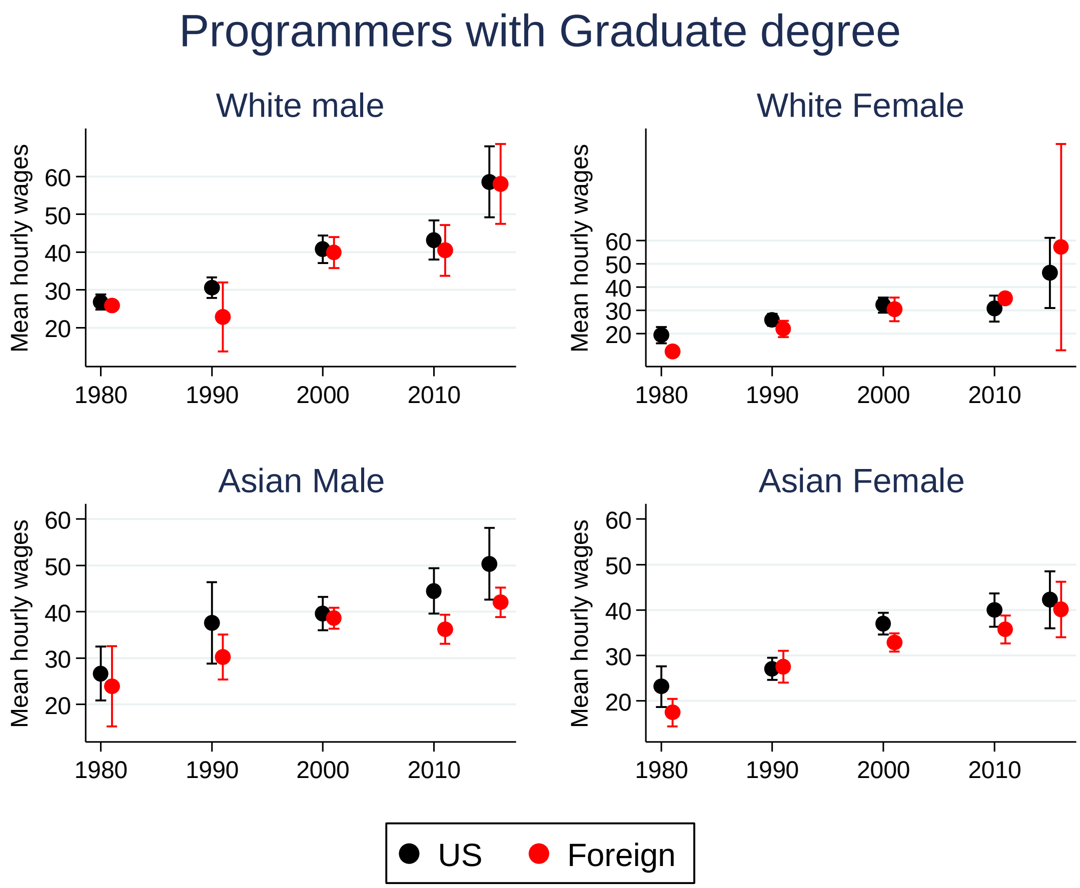
Note: Sample for wage analyses are full-time (at least 35 hours of work a week) full-year (at least 50 weeks in previous year) workers in labor force with positive wages using 1999 dollars. The sample is further limited to workers between 25-44 years old.

Figure 4a. Wages of programmers with undergraduate degrees, by race, gender and citizenship status



Note: Sample for wage analyses are full-time (at least 35 hours of work a week) full-year (at least 50 weeks in previous year) workers in labor force with positive wages using 1999 dollars. The sample is further limited to workers between 25-44 years old.

Figure 4b. Wages of programmers with graduate degrees, by race, gender and citizenship status



Note: Sample for wage analyses are full-time (at least 35 hours of work a week) full-year (at least 50 weeks in previous year) workers in labor force with positive wages. The sample is further limited to workers between 25-44 years old.

Appendix

A-1. Percentage of degree completions in CS, by race and gender
California

	Bachelor's						
	1985	**1990**	**1995**	**2000**	**2005**	**2010**	**2015**
White Male	42.17	41.1	37.27	29.38	25.85	33.28	28.96
White Female	17.75	10.51	11.01	8.19	4.48	4.7	3.99
Asian Male	11.77	14.62	17.39	25.27	24.46	15.64	22.07
Asian Female	9	9.36	8.55	10.64	7.31	3.64	6.03
Hispanic Male	1.79	3.9	4.2	5.22	7.84	10.91	13.5
Hispanic Female	0.81	1.72	1.33	2.08	2.13	2.11	2.43
Black Male	0.95	1.89	2.54	2.4	2.69	3.31	3.52
Black Female	0.81	1.07	1.69	1.37	0.93	0.95	0.58
Foreign Male	6.87	8.04	6.62	5.28	7.52	3.92	4.55
Foreign Female	3.04	3.15	3.39	2.11	2.63	0.86	1.3
Total	2,957	2,798	2,479	3,506	5,585	3,594	5,518
	Graduate						
	1985	**1990**	**1995**	**2000**	**2005**	**2010**	**2015**
White Male	32.07	36.61	30.98	20.14	15.67	17.96	14.05
White Female	10.34	9.57	6.9	5.78	3.42	3.57	3.52
Asian Male	11.52	11.5	11.87	12.01	12.37	9.63	7.71
Asian Female	4.06	4.51	5.39	7.94	7.34	3.41	3.07
Hispanic Male	0.65	0.83	1.6	1.46	2.19	2.06	2.72
Hispanic Female	0.52	0.46	0.59	0.89	0.82	0.34	0.7
Black Male	0.39	0.92	0.93	1.14	1.15	1.43	1.99
Black Female	0.13	0.46	0.25	0.64	0.45	1.14	0.56
Foreign Male	15.84	22.72	26.35	29.67	34.52	36.46	41.6
Foreign Female	7.72	7.73	7.24	16.14	14.72	12.74	18.13
Total	764	1,087	1,188	1,574	2,425	2,378	2,868

National

	Bachelor's						
	1985	1990	1995	2000	2005	2010	2015
White Male	51.6	50.96	47.72	44.76	45.19	50.46	46.7
White Female	28.46	17.95	15.13	13.36	9.33	8.78	8.31
Asian Male	2.96	4.92	6.4	9.77	8.58	5.94	8.38
Asian Female	2.26	2.97	3.16	4.67	3.01	1.63	2.53
Hispanic Male	1.26	2.49	3.54	3.76	4.71	6.27	7.95
Hispanic Female	0.88	1.6	1.74	1.81	1.69	1.47	1.72
Black Male	2.65	3.87	5.02	4.82	6.33	6.82	6.95
Black Female	2.83	4.26	5.09	4.42	4.43	3.18	2.38
Foreign Male	3.69	5.39	6.75	5.44	5.53	3.69	4.38
Foreign Female	1.72	2.12	2.54	2.47	1.79	1.02	1.39
Total	39,121	27,259	24,719	36,565	56,150	40,973	62,023
	Graduate						
	1985	**1990**	**1995**	**2000**	**2005**	**2010**	**2015**
White Male	43.31	36.74	31.41	22.59	24.31	22.98	17.84
White Female	17.28	14.4	9.4	7.73	7.14	6.33	5.49
Asian Male	5.81	6.28	7.59	7.92	7.52	5.53	4.55
Asian Female	2.75	3.27	4.21	5.64	4.45	2.52	2.34
Hispanic Male	0.91	0.95	1.37	1.39	1.92	2.55	2.44
Hispanic Female	0.39	0.26	0.45	0.56	0.77	0.84	0.84
Black Male	1.51	1.46	1.83	2.06	2.84	3.45	3.74
Black Female	0.98	0.99	1.32	1.65	1.98	2.44	2.14
Foreign Male	18.17	21.65	28.77	30.09	29.13	31.83	37.37
Foreign Female	6.04	6.44	9.05	15.71	11.84	12.48	17.23
Total	7,349	10,146	11,294	15,209	20,091	19,961	33,948

A-2. Region of origin of all degrees for foreign students (Undergraduate, Graduate, and other)

Year	Africa	Asia	Europe	Latin America	Middle East	North America	Oceania	TOTAL
1979-1980	12.6%	28.5%	7.9%	14.8%	29.2%	5.4%	1.4%	286,340
1984-1985	11.6%	42.0%	9.7%	14.2%	16.5%	4.7%	1.2%	342,110
1989-1990	6.4%	53.8%	11.9%	12.4%	9.6%	4.8%	1.0%	386,850
1994-1995	4.6%	57.8%	14.3%	10.4%	6.7%	5.2%	1.0%	452,635
1999-2000	5.9%	54.4%	15.2%	12.1%	6.8%	4.7%	0.9%	514,723
2004-2005	6.4%	57.5%	12.7%	12.0%	5.5%	5.1%	0.8%	565,039
2009-2010	5.4%	63.1%	10.8%	9.5%	6.4%	4.1%	0.7%	690,923
2014-2015	3.4%	64.3%	8.2%	8.9%	11.7%	2.8%	0.7%	974,926

Notes: In 2009-10, Cyprus & Turkey were re-categorized from Middle East to Europe. However, due to the quality of data from prior to 1995-96 data, it is not possible to re-categorize these countries so Cyprus & Turkey were re-classified as Middle East in this analysis. Cyprus represents 211 undergrads and 296 graduate students while Turkey represents 3,656 undergrads and 6,585 graduate students in 2009-10. North America consists of Canada and Bermuda (vast majority is from Canada).

A-3. Educational attainment of programmers, by race

	White					Asian				
	1980	**1990**	**2000**	**2010**	**2015**	**1980**	**1990**	**2000**	**2010**	**2015**
Less than HS	0.77	0.46	0.48	0	0	2.5	0.67	0.18	0	0
	0.44	0.21	0.18	0	0	2.45	0.67	0.11	0	0
HS Graduate	8.95	3.66	2.27	3.13	2.92	2.5	2.01	0.68	0.38	0
	1.48	0.66	0.38	1.19	0.84	1.75	1.35	0.24	0.38	0
Some College	31.97	28.81	17.9	13.38	9.61	20	10.25	4.56	1.97	2.34
	2.36	1.66	0.94	2.27	1.45	4.49	2.06	0.55	0.74	0.53
College Graduate	32.99	47.31	50.19	48.71	49.95	40	58.66	45.27	44.41	41.37
	2.36	1.82	1.24	3.12	2.55	5.52	3.42	1.4	2.79	2.09
Graduate Degree	25.32	19.76	29.16	34.79	37.53	35	28.41	49.3	53.24	56.29
	2.22	1.44	1.13	2.9	2.44	5.37	3.12	1.41	2.8	2.1
Total	*7,820*	*17,558*	*38,663*	*31,664*	*50,511*	*1,600*	*4,770*	*32,996*	*52,034*	*83,552*

	Latino					Black				
	1980	**1990**	**2000**	**2010**	**2015**	**1980**	**1990**	**2000**	**2010**	**2015**
Less than HS	0	5.61	3.71	0	0	0	4.29	2.91	0	0
	0	5.36	1.7	0	0	0	3.03	2.1	0	0
HS Graduate	9.09	12.34	6.77	8.41	1.5	25	3.28	0.99	0	0.45
	6.13	4.6	2.51	7.91	1.51	10.83	3.23	0.99	0	0.49
Some College	50	30.65	32.58	23.96	6.48	43.75	35.81	50.11	15.69	0
	10.66	7.21	5.03	10.22	3.72	12.4	9.01	6.99	16.69	0
College Graduate	27.27	45.61	41.13	49.18	66.07	12.5	54.22	35.65	0	37.18
	9.5	7.85	5.15	11.48	8.42	8.27	9.4	6.91	0	18.32
Graduate Degree	13.64	5.79	15.82	18.45	25.95	18.75	2.4	10.33	84.31	62.37
	7.32	2.92	3.74	7.84	8.03	9.76	2.38	3.43	16.69	18.4
Total	*440*	*1,070*	*2,643*	*1,832*	*4,058*	*320*	*793*	*1,307*	*663*	*1,552*

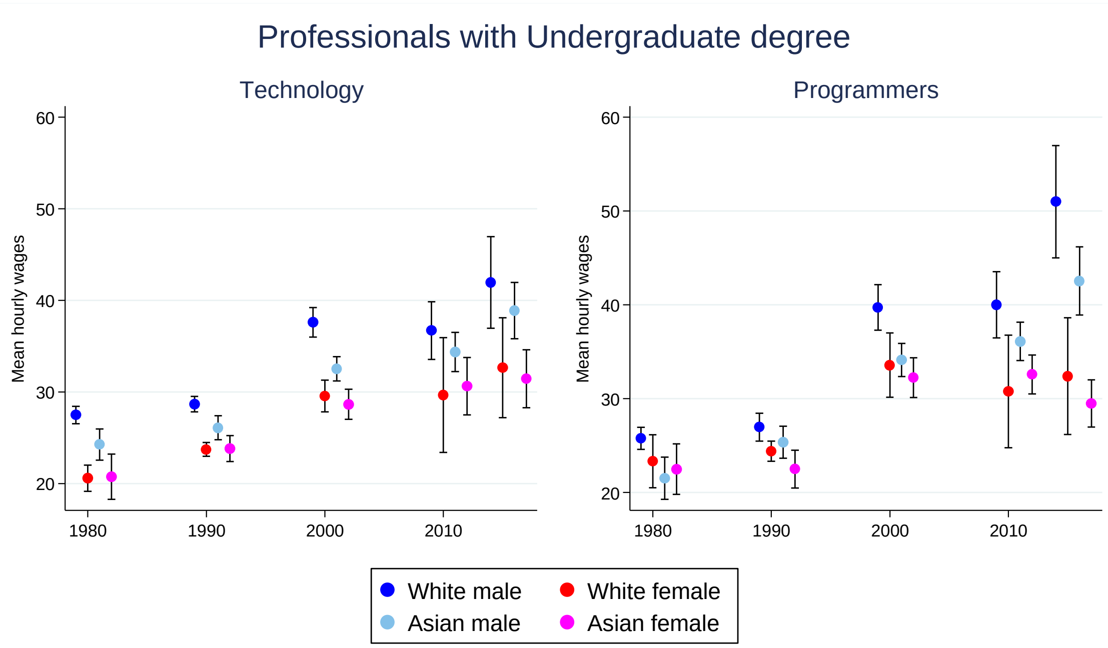
Note: Only full-time (at least 35 hours of work a week) full-year (at least 50 weeks in previous year) workers in labor force (16+ years old).

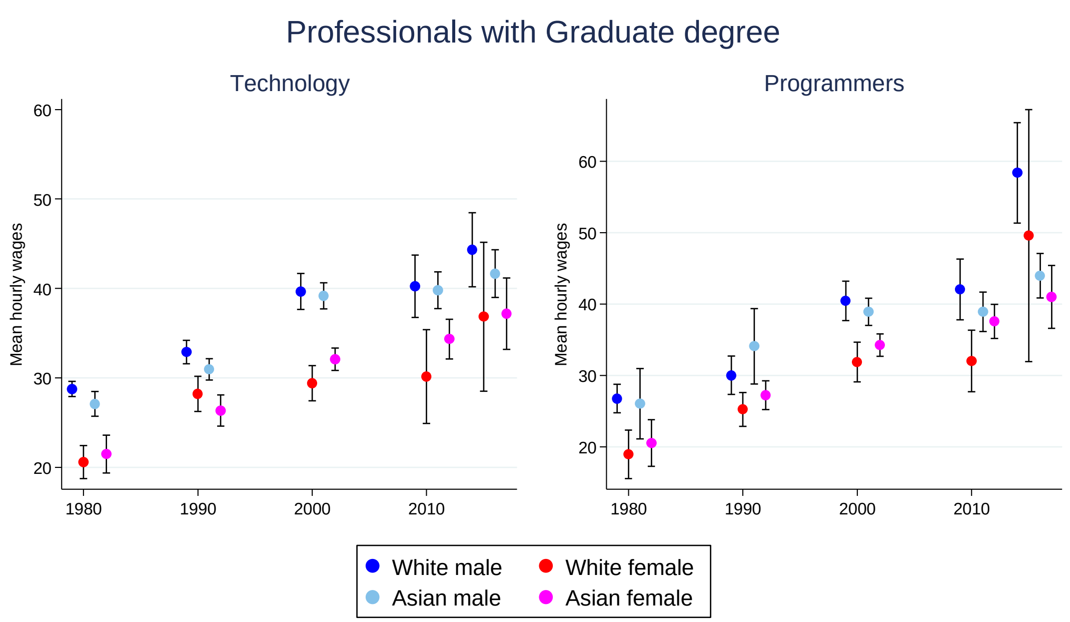
A-4. Percentage of programmers who are foreign workers

	1980	1990	2000	2010	2015
White	3.32	4	14.24	14.42	17.53
Asian	22.5	22.56	53.2	49.56	53.57
Hispanic	13.64	18.41	23.46	22.33	12.64
Black	0	0	11.25	12.37	32.02

Note: Only full-time (at least 35 hours of work a week) full-year (at least 50 weeks in previous year) workers in labor force (16+ years old).

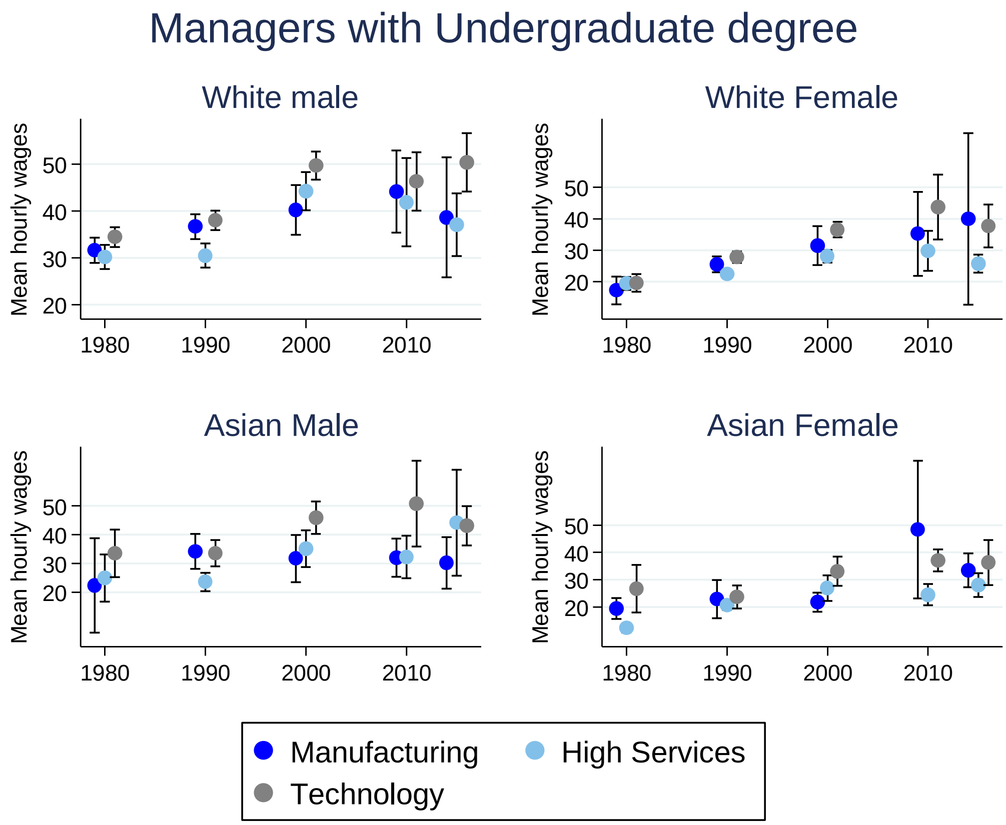
Appendix A-5. Hourly wages in technology industry and among programmers, by degree level



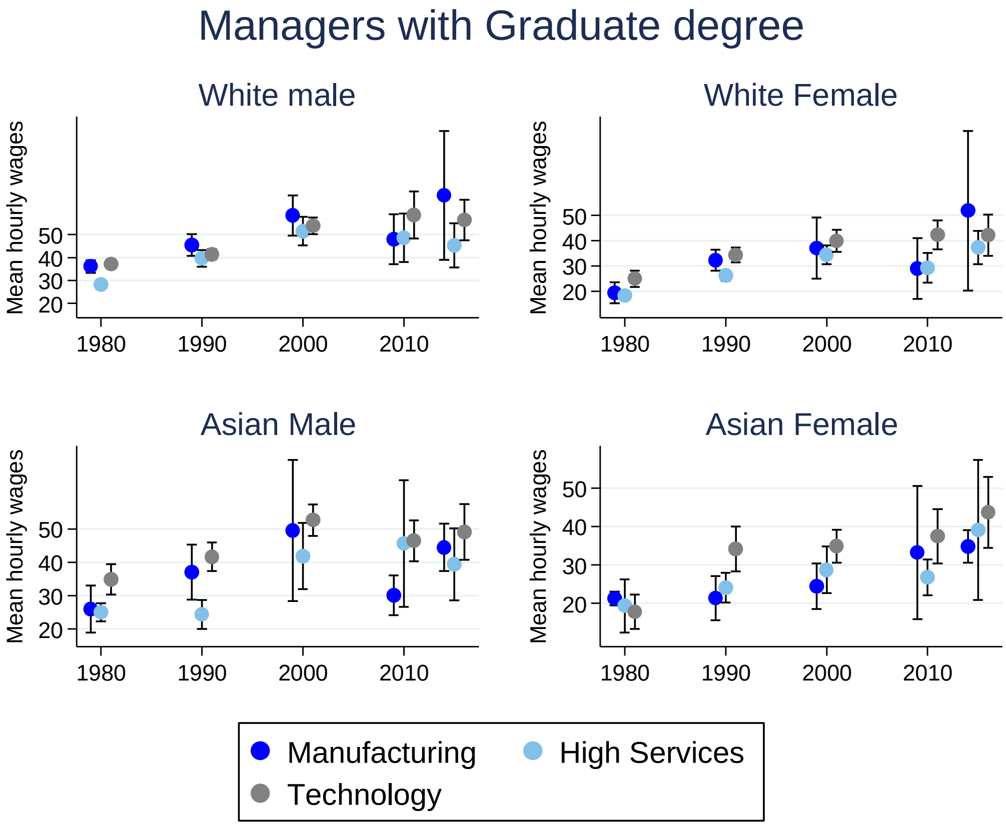


Note: Sample for wage analyses are full-time (at least 35 hours of work a week) full-year (at least 50 weeks in previous year) workers in labor force with positive wages using 1999 dollars. The sample is further limited to workers between 25-44 years old.

Appendix A-6. Wages of managers and professionals across industries, by degree level

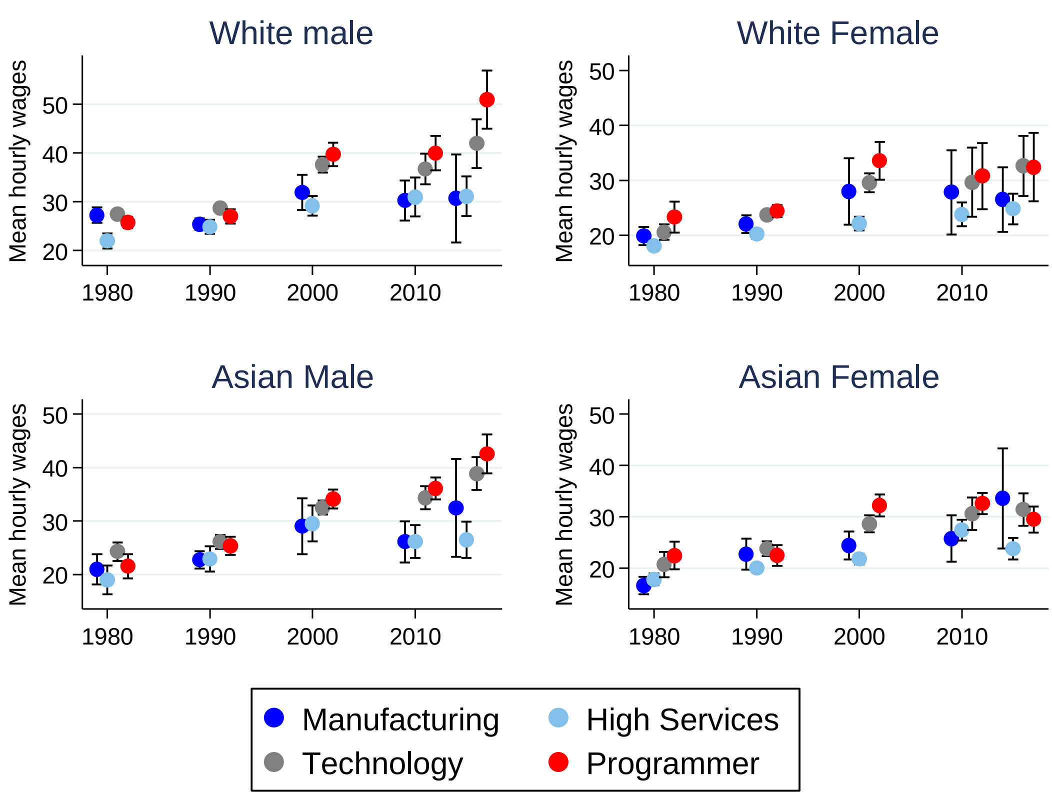


Note: Sample for wage analyses are full-time (at least 35 hours of work a week) full-year (at least 50 weeks in previous year) workers in labor force with positive wages using 1999 dollars. The sample is further limited to workers between 25-44 years old.

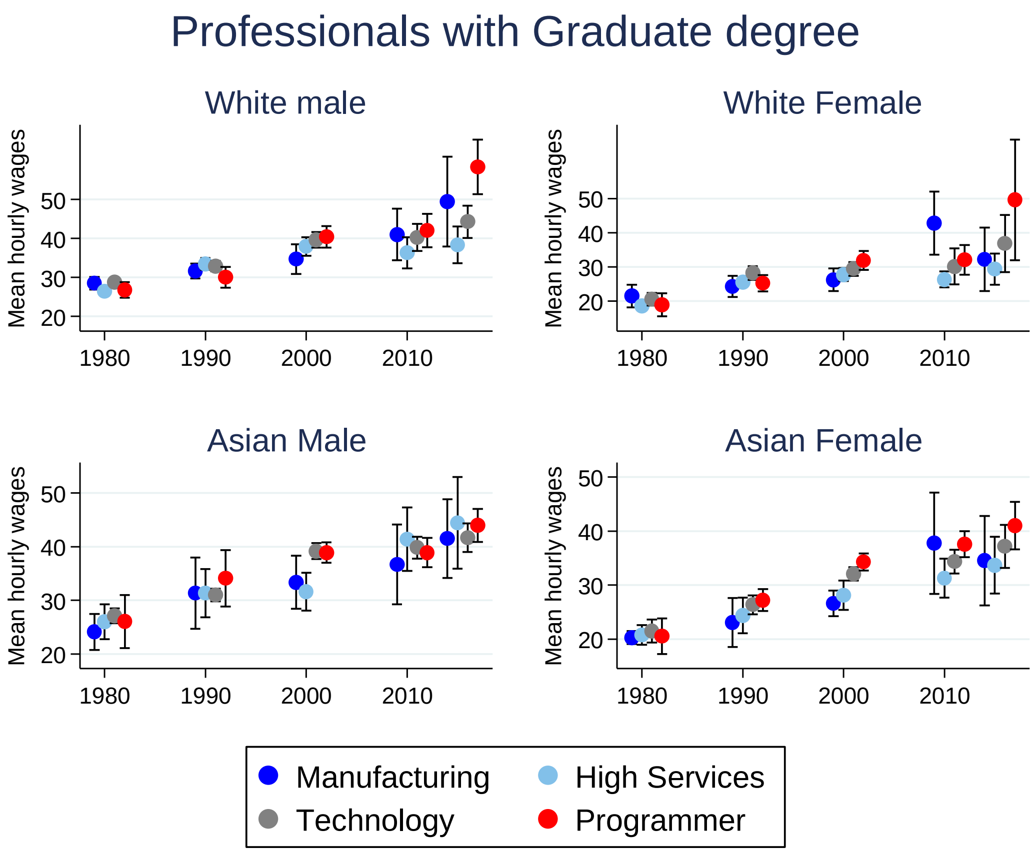


Note: Sample for wage analyses are full-time (at least 35 hours of work a week) full-year (at least 50 weeks in previous year) workers in labor force with positive wages using 1999 dollars. The sample is further limited to workers between 25-44 years old.

Professionals with Undergraduate degree



Note: Sample for wage analyses are full-time (at least 35 hours of work a week) full-year (at least 50 weeks in previous year) workers in labor force with positive wages using 1999 dollars. The sample is further limited to workers between 25-44 years old.



Note: Sample for wage analyses are full-time (at least 35 hours of work a week) full-year (at least 50 weeks in previous year) workers in labor force with positive wages using 1999 dollars. The sample is further limited to workers between 25-44 years old.

Paper 2: Gender differences and the effect of facing harder competition

This paper was published in Journal of Economic Behavior & Organization.
Full Citation: John, J.P. (2017). Gender differences and the effect of facing harder competition. *Journal of Economic Behavior & Organization*, Vol. 143, Pages 201-222. doi: https://doi.org/10.1016/j.jebo.2017.08.012.

1. Introduction

Many studies have shown that females are less competitive than males in stereotypically male tasks (see Niederle and Vesterlund, 2011 for review), which explains some of the gender differences in later education and career outcomes (Almås et al., 2016; Buser et al., 2014; Buser et al., 2017; Ors et al., 2013; Zhang, 2013). One important aspect of competition is the perceived difficulty of the competitors: people may react differently in competition when facing easier or harder opponents. Gender differences in these reactions can help explain dynamics of competition and inform policy decisions about the characteristics of competitions in schools or the workplace. Existing research on the perceived difficulty of the competition primarily relies on information provided in a laboratory context which may have limited applicability in the field. In the current study, I exploit natural sorting within grade levels to randomly assign competitors of different perceived difficulty levels to examine the effect of facing harder competitors by gender in addition to replicating the standard gender gap on a math task in Malaysian public schools.

Gender gaps in competition have been categorized by both choice and performance. Females are shown to be less likely than males to choose into competition, a well-established finding in the literature (Niederle and Vesterlund, 2007). Recent research explores how factors such as task or information affect this gender gap (see Niederle, 2016 for review). There is less consistent evidence, however, of gender differences in performance in competitive environments. A seminal paper finds that females perform worse than males when solving puzzles under a competitive incentive scheme, although there is no difference in performance under a non-competitive incentive scheme (Gneezy et al., 2003). Other studies use similar designs and puzzle tasks with similar results (Datta Gupta et al., 2013; Günther et al., 2010). Niederle et al. (2013) finds that males outperform females in math tasks under competition. However, other studies show no gender differences in performance under either non-competitive or competitive incentives in math tasks (Ertac and Szentes, 2011; Niederle and Vesterlund, 2007; Wozniak et al., 2014).

The literature indicates that gender differences in competitive performance cannot be simply explained by differential ability, which has shifted some recent literature to study how features of competition may differentially affect males' and females' performance. One aspect of competition is how people respond to harder or easier competitors and whether there are gender differences in these responses, the focus of the current study.

Prior research has examined reactions to different levels of competition by providing information or relative feedback during competition[1] in a laboratory environment (Buser, 2016; Cason et al., 2010; Eriksson et al., 2009; Ertac and Szentes, 2011; Gill and Prowse, 2014; Kuhnen and Tymula, 2011; Wozniak et al., 2014), with one recent study conducted in a field setting (Wozniak et al., 2016). In these studies, information about either random competitors or deliberately lower- or higher-performing competitors is given to subjects prior to subsequent competition decisions and performance.

Rational behavior predicts that people would be more reluctant to enter into competition against more difficult competition. Cason et al. (2010) created groups of

[1] The following discussion of existing literature focuses on studies that involve competition in a math-related task and explore gender differences, although Gill and Prowse use a slider task specifically designed to measure effort (Gill and Prowse, 2014). Other studies examine how information affects performance without any differences in incentives and will not be discussed (e.g. Azmat and Iriberri, 2010).

relatively weaker, stronger, or superstar competition and the study finds that, as expected, the fraction of entry into a tournament is highest against the weaker group and lowest against the superstar group. No breakdown by gender is provided, although there is some indication of gender differences-- females under-enter a proportional pay tournament given their expected payout, with no gender difference in under- or over-entry for the winner-take-all tournament. A clear gender difference in choice of competition is demonstrated in an unpublished study by Niederle and Yestrumskas (2008), which shows that females choose a less difficult and less lucrative task than males; however, both genders receive lower payout than if they had optimally chosen their task difficulty.

There is consistent evidence that information about target or relative score provided to subjects decreases or even eliminates the gender gap in entry into competition (Ertac and Szentes, 2011; Wozniak et al., 2014), although Wozniak et al. (2016) finds a persistent gender gap in competition entry among low-ability participants even after information is provided. However, the effect of information on gender differences in performance is less clear.

When subjects must compete, there are mixed results in reactions to information about competitors. Eriksson et al. (2009) finds that feedback on relative performance does not significantly change performance. The study reports positive peer effects in tournaments; frontrunners do not slack off and underdogs rarely quit, although continuous feedback reduces the quality but not quantity of effort for underdogs. However, Gill and Prowse (2014) finds that subjects reduce effort after a loss, although males reduce effort only after failing to win large prizes. Buser (2016) shows somewhat different results depending on gender. Buser created three groups based on random pairing in a first round winner-take-all tournament: winners, losers, and those who receive scores, which he refers to as the no information group. Losers from the first round seek harder challenges, are less successful in the challenges and overall make less money in the second round compared to the winners. While there are no gender differences in average outcomes, such as the challenge level selected or performance in the challenge, males react to losing by becoming more challenge-seeking than winners and females react by lowering their performance.

The findings in these previous studies are contingent on random or contrived information about competitors to elicit a reaction from subjects. Although there is a range in the type of information provided, from relative scores to more direct messages of winning or losing, the explicit information acts as a treatment. The use of explicit information may contribute to results in the previous studies-- a study shows that the possibility of receiving feedback induces subjects to work harder even when they are not compensated for the extra effort, which demonstrates how responsive subjects can be to explicit information (Kuhnen and Tymula, 2011).

I focus on the effect of competitor level on competition performance, a relatively less understood aspect of gender differences in competition. I explore reactions to a subtler but realistic scenario of the perception of competitor difficulty, since people often compete with incomplete information about their competitors. For example, students may not know their rankings in class prior to taking a test; even if these rankings are known from a prior test, they do not perfectly transfer to another subject or even another test in the same subject. Despite this uncertainty, students must perform on assignments or tests. Thus, it is important to explore how a noisier yet realistic signal of competitor difficulty affects performance in competition. Although the context is essentially a lab-in-field environment rather than an actual school competition, the school setting allows students to compete against meaningful categories of competitors instead of relying on artificial competitors created by researchers.

By closely following Buser et al.'s (2014) protocol used in secondary schools in the Netherlands, the current study also provides evidence for replicability of findings in a different context. Cultural context is demonstrated to play a role in gender differences in competition (Gneezy et al., 2009), although not necessarily in expected ways (Cárdenas et al., 2012); thus, it is important to acknowledge potential cultural influences on these differences. Nearly all of the studies use university subject pools in Western countries. To the author's knowledge, this is the first such experiment performed in a Muslim country and one of few performed in Asia. While this paper highlights several differences in the Science, Technology, Engineering & Math (STEM) and gender context particular to Malaysia, the findings are suggestive of gender stereotypes and differences in competition in STEM generally found in the literature.

The results of this study demonstrate that in a context where the standard gender difference in competition entry exists, males appear to be affected by the level of competition while females are not. When students face harder competitors, males respond by lowering performance while the performance of females does not vary significantly by level of competition. These somewhat surprising findings suggest that policies that require females to enter into more difficult competitive situations may not be detrimental to their performance in these situations.

The rest of the article proceeds as follows. Section 2 provides an overview of the study details, including context, data collection procedures and study design. The results from the study are detailed in Section 3. First, I provide descriptive analyses of the behavioral characteristics and other control variables used in later analyses. Then, I provide the analysis of the standard gender differences in competition (same-class competitions). Lastly, I provide an analysis of the response to different levels of competition (other-class competitions). Section 4 discusses potential mechanisms of these findings. Section 5 concludes.

2. Study overview

2.1 Context

Gender differences in competition appear to exist at a young age (Eccles et al., 1993; Gneezy and Rustichini, 2004; Harbaugh et al., 2002; Sutter and Glätzle-Rützler, 2014). These early differences may affect the trajectories of individuals' future decisions and outcomes. To understand competition phenomena in a relevant setting, this study uses a sample of high school students prior to any academic tracking.

This study takes place in public schools in Malaysia, a multicultural developing country in Southeast Asia with a majority Muslim population. Malaysia is a useful context for this study for several reasons. First, the informal but widespread ranking system within grades in public schools provides a unique opportunity to exogenously vary the level of competitor within classrooms, which will be discussed further in Section 2.2. Second, the STEM context in Malaysia appears to favor females compared to the populations used in prior studies, although standard male stereotypes of STEM seem to persist. Several studies view stereotypes associated with tasks as potential explanations for gender differences in math task competitions (Dreber et al., 2014; Grosse and Riener, 2010; Günther et al., 2010; Kamas and Preston, 2010; Shurchkov, 2012), thus any competitive differences found in the Malaysian context could help bring insight into whether gender differences in competition are similar in an environment with greater female STEM participation.

The Malaysian education system consists of six years of primary school and five years of secondary school; during the last two years of secondary school, or upper secondary school[2], students are placed into academic tracks with different associated prestige: the arts track (less prestigious) and the science track (more prestigious). Although there is no official tracking policy prior to the last two years of secondary school, many secondary schools use unofficial methods[3] of ranking and sorting students into classrooms within grade levels. Enrollment in preschool, primary school and secondary school is gender-balanced (49%-50% of enrollment is female). However, there are differences in gender proportions in the upper secondary school academic tracks. In upper secondary school, females constitute about half (47-49%) of the arts stream and the majority (about 58-59%) of students in the science streams[4]. Thus, there are more females than males in the most prestigious science track at the upper secondary level (Ministry of Education Malaysia, 2014). A similar gender distribution is found in the lower secondary Form 3[5] classes in this study, prior to the official academic tracking (see Section 2.2 for details).

[2] Form 4 & 5 are known as upper secondary and are equivalent to grades 10 & 11.
[3] For example, sorting students into classrooms based solely on overall test scores.
[4] Science and arts streams are the two most common streams; some schools offer "sub-science" or "sub-arts" as well.
[5] Equivalent to grade 9.

The female advantage continues in tertiary education. Malaysia has a slightly lower ratio than the U.S. of females to males in tertiary education, although in both countries, females make up the majority of tertiary students (Malaysia: 1.21 to US: 1.36). However, nearly half of Malaysian female students (46%) versus less than a third of U.S. female students (30%) major in STEM fields (World Economic Forum, 2014). In fact, Malaysian females make up the majority of entrants, enrollments and graduates in most fields of study in the public universities including about two-thirds of graduates in Science, Mathematics and Computer; the only field in which females are a minority is Engineering, Manufacturing and Construction (Ministry of Education Malaysia, 2015). A qualitative study of the University of Malaya's[6] Computer Science and Information Technology department reveals that the majority of faculty, heads of department and dean were women in 2001 (Mellström, 2009). Mellström hypothesizes that computer science professions may be considered more suitable for females because of the office rather than field nature of the work; however, labor market data is limited such that it is not possible to identify the percentages of women in these fields.

Thus, females in Malaysia appear to face a more positive STEM climate in education than in many other countries. Nevertheless, gendered stereotypes for STEM and reading exist (see Section 3.1). Furthermore, prevailing gender norms may discourage females from being too "aggressive", which could influence gender responses to competition (Curriculum Development Division, 2016). These features demonstrate that multiple components of culture create a complex atmosphere that may affect gender dynamics in competition.

[6] Malaysia's oldest and most prestigious public university.

2.2 Data collection

This experiment was conducted in public secondary schools in one school district in Selangor, the largest and most urban state in Malaysia. I invited co-educational secondary schools in this district to participate in this study, asking for one classroom period of time; five schools agreed to participate. All schools in this study sort students into classes within grades by prior achievement, a widespread practice in Malaysia, and have a minimum of five classes in Form 3[7] to ensure sufficient variation in competition levels. Three to five classes from Form 3 were selected from each school to participate. The data collection was conducted over the span of one month, from July-August 2015. For a given school, the experiments in different classrooms[8] were conducted during the same day and often at the same time. Not every classroom in Form 3 in a school participated, experiments were often conducted at the same time within a school, and the bulk of the classroom experiments in the entire sample was conducted within one week, so there is little reason to worry that students knew about the experiment and strategized prior to participating. Students were paid two weeks after the experiment through sealed envelopes; there was no fixed participation fee and the average payout was RM10.26[9], with a minimum of RM0 and maximum of RM71.

Four of the five schools provided administrative information including student gender and midterm grades (the most recent official grades). The study was conducted during regular classroom instruction time in eighteen classrooms[10]. Each school engaged

[7] 9th grade equivalent; last year of lower secondary school and prior to academic track specializations.
[8] The experiment for one classroom at one school was conducted about three weeks after the rest of the classrooms at that school because of scheduling problems.
[9] Currency was given in Malaysian Ringgit (MYR), which has a similar purchasing power to USD although the exchange rate was roughly 4 MYR:1 USD in summer 2015.
[10] One additional classroom was dropped due to technical problems.

in some form of classroom rankings such that the classrooms were ordered according to student achievement, prior to official academic tracking practices at the end of Form 3. Students are well aware of this ranking, similar to how students in other countries such as the U.S. are aware of being in advanced or remedial classes. For example, in three of the five schools, classes are named in alphabetical order from top to bottom class. The top class, bottom class, and one to three middle-ranked classes in Form 3 of each school participated in this study. There were 562 secondary school students in Form 3 who participated in this study, but one student was dropped because there was no gender information available, leaving a sample of 561 students (290 males and 271 females). In the sample, females make up 40% of the bottom classes, 48% of the middle classes and 54% of the top classes[11]. The analyses of the effect of facing a different level of competition (i.e., easier or harder competition) are limited to the sample of middle classes (266 students), which were oversampled for this purpose.

The schools in this study represent over a fifth of the 24 public co-educational secondary schools[12] in the district. Although they may not be representative of the country as a whole, the schools appear to be similar on average to Malaysian public secondary schools. The average classroom size in the schools in the sample is 35.28, similar to the national average lower secondary classroom size of 34 (Ministry of Education Malaysia, 2014). Females make up 48% of the sample, similar to the national percentage of 50% (2015 data) in Form 3 (Ministry of Education Malaysia, 2015).

[11] Post hoc ANOVA comparisons using the Sidak (p=0.036), Bonferroni (p=0.036), Scheffe (p=0.043) and Tukey (p=0.032) methods indicate that only the bottom and top classes have a statistically significant different proportion of females at the p<0.10 significance level.
[12] Most students in Malaysia attend co-educational schools. Wiseman (2008) finds that 14.67% of schools (indexed by 8th grade math classrooms) were sex-segregated, which was not statistically different from the international mean of 18.94%.

2.3 Study design

The objective of this experiment is to measure the rates of entering a competition when competing against classroom peers, and, in a subsequent round, to measure differences in performance when forced to compete against students from another higher- or lower-ranked class in the same grade and school.

The experiment has four rounds of tests with varying incentive structures followed by a survey, similar to the design first used in Niederle and Vesterlund (2007). The test instrument for each round was a five-minute math test with 40 double digit multiplication questions, which is a slightly longer and more difficult task than the one used by Niederle and Vesterlund, in order to enable more variance in scores due to an additional incentivized round in this study. This task was designed to measure the level of effort, not mathematical knowledge or attitudes. None of the questions repeat in the study and all numbers with zeroes were removed in order to keep the level of difficulty comparable across each test. There were no penalties for incorrect answers. Students were not allowed to use calculators but were given pieces of scratch paper to solve problems on. Directions about the specific incentive system of the round's test were read out loud in Malay, the language of instruction, prior to each test. All documents were given in both English and Malay. Students were told not to speak during the duration of the study, and had to place their pens down and stand up when the end of each test was announced. Furthermore, students were informed that only 1 out of the 4 rounds of tests would be compensated, randomly chosen at the end of the session, in order to avoid hedging and to encourage each student to try his/her best during each round. Thus, at the end of each session, a representative from the class picked a ball numbered from 1 to 4 out of an opaque bag to choose which round was paid out for that entire class.

Test 1 was scored according to a piece-rate incentive; for this test, students were paid RM0.50 per each correct answer. Test 2 was scored according to a winner-take-all tournament incentive (i.e. a competitive incentive). For this test, students were told they would be competing against 3 other randomly selected students (4 students per group) from their class. If they obtained the highest score (i.e. first place), they received a payment of RM2 per each correct answer, but if they did not obtain the highest score, they received nothing[13].

Prior to Test 3, students were given the choice of how they wanted to be compensated for the third test. Each student chose between one of the prior two incentive schemes, marked the choice on a form, then inserted the form into an envelope. Students were informed prior to decision-making that if they chose the winner-take-all tournament incentive, they would compete against a new set of three randomly selected competitors' scores from Test 2 so they could be competing against any of their classmates, not just those who chose the tournament incentive (see Niederle and Vesterlund, 2007). Test 3 proceeded after every student selected a choice and put away the form in an envelope.

Prior to Test 4, students were given slips of paper that informed them which class they would be competing against in the fourth test. Thus, in the fourth round of the study, students were told they would be competing in a winner-take-all tournament, competing against three randomly selected students from the other class, under the same incentive structure as Test 2 but using only Test 4 scores. In each class, students were randomly assigned to one of two other classes in their grade (e.g. bottom or top class if the student were in a middle class); classes were referred to by their official school names with no

[13] Ties were awarded the same rank, and then skipped the next number of ranking (Stata's egen rank, field option).

explicit reference to positioning within the grade level. However, as described earlier, students are well aware of the implicit differences between classes.

After Test 4, students completed a survey which included incentivized questions on levels of confidence and risk aversion, in addition to non-incentivized questions about their attitudes, opinions and family background. Students never received information about their scores during the experiment. Students could estimate how they had performed only after they were given their payments, a couple weeks after the experiment had been completed.

3. Results

3.1 Same-class competition analysis

The following section presents results from the first three rounds of the study, which replicates the design from Buser et al. (2014). First, I provide the descriptive results of the performance, competition choice, behavioral and other individual characteristics of students. I then present the regression results that confirm the gender gap in competition.

There is no gender gap in performance for this multiplication task, whether students are under piece-rate or tournament incentive against their classroom peers. A table of descriptive characteristics shows the performance and competition choice prior to Test 3, when all students are under the same incentive structures (Table 1). Although it is not a focus of this paper, there is evidence that the sorting mechanism into classrooms by student prior achievement resulted in classes with overall differences in student performance, which is an important component of the analyses of performance against other classes. The average number of questions correct for the first test, under the piece-rate incentive, is 10.141 although this varies between 5.937 in the bottom classes to

12.432 in the top classes. The average number of questions correct for the second test, under the winner-takes-all tournament incentive, is significantly higher at 12.041, ranging from 7.746 in the bottom classes to 14.444 in the top classes[14]. Overall, females appear to outperform males on these first two tests, though these gender differences disappear when taking into account the class level and corresponding differences in gender distribution across class levels. Thus, it is established that there are no gender differences in performance under either of the incentives for this task.

[Table 1 about here]

Furthermore, both genders increase performance under the competition incentive. The different incentive structures between Test 1 and Test 2 affects both genders; the average number of answers correct between Test 1 and Test 2 statistically significantly increases for both males and females (Appendix A-1). This increase could indicate learning with successive tests (discussed further in Section 3.2); however, a recent study finds that the order of piece-rate and tournament rounds does not significantly affect the difference in performance under the two incentives in a similar experiment (Wozniak et al., 2016). Therefore, we can interpret the positive increase as the response to competition.

Unlike performance on the tests, there is a clear difference in the rates at which males and females choose competition, both overall and at each class level. Overall, less than a third of students (29.6%) choose competition for the incentive structure of Test 3. Females choose into competition at almost half the rate of males, with an average of

[14] The numbers of correct answers for both Test 1 and Test 2 are different between all three class levels according to the analysis of variance comparisons, which indicates that student ability in these tasks has been appropriately sorted by class levels (ANOVA analyses available upon request).

20.7% of females versus 37.9% of males choosing competition, with the greatest difference in the top classes (18.5% of females versus 46.8% of males).

The choice into competition for Test 3 does not appear to incentivize students to perform better than those who did not choose into competition for Test 3. There is no difference in the increase in number of correct answers from Test 2 to Test 3 for those who chose competition and those who chose piece-rate (Table 2). This can indicate either insensitivity to the choice, or poor measurement of effort (e.g. ceiling effects) on performance. Subsequent increased performance on Test 4 discussed in Section 3.2 implies that students did not respond to choice, rather than the task failing to measure changes in effort.

[Table 2 about here]

Other factors such as confidence, risk-aversion, academic performance, attitudes and expectations towards math/science, and socio-economic status may be influential in students' choice of competition. A summary of student behavioral and personal characteristics is shown in Table 3 (Appendix A-2 for detail). There are several characteristics that differ by gender.

Males are more confident and over-confident than females in competitions against their own class. Confidence is measured by two questions on the survey, similar to what is used in Niederle and Vesterlund (2007). These questions ask what rank (1-first place to 4-last place) students think they had achieved for the two forced competition rounds, Test 2 (against own class) and Test 4 (against other class). Students received RM1 per correct answer for these questions. Overconfidence is defined as the difference between actual rank[15] and guessed rank, with a range of -3 to 3. This measure provides the student's level of confidence for the particular task rather than a more generalized measure (e.g. soliciting student perceptions about class rank). The average guessed rank of males against their own class is 2.441 versus 2.715 for females (p=0.001); thus, males guessed that they obtained a better rank than females guessed. After accounting for actual ranks, females are under-confident while males' guessed ranks are closer to their actual ranks (slightly under-confident against their own class and slightly over-confident against another class).

It appears that males are more accurate in their rankings, although both males and females appear less confident about winning than other studies have found (e.g. Niederle and Vesterlund, 2007). However, the male percentage is roughly in line with what was found in a sample of similarly-aged students (Buser et al., 2014). About 21% of males and 9% of females believe that they won the tournament in Test 2 (p<0.001), while 30% of males and 24% of females actually win the tournament, with no significant difference.

Males are more risk-seeking than females according to both risk measures in this study. Risk preference is measured in two ways on the survey. First, students answered an incentivized question based on a modified question used by Eckel and Grossman (2002) that asked them to choose between an option with 100% certainty (RM2) or one of four 50/50 lottery options based on a flip of a coin at the end of the study: RM3 or RM1.50, RM4 or RM1, RM5 or RM0.50 or RM6 or RM0. The coin was flipped in front of the classroom at the end of the study and the individual's choice was paid out with the

[15] Actual rank is constructed from 1,000 simulations of random draws of 3 other students from the appropriate class against a given student's score; the modal value was selected as actual rank.

rest of his/her earnings. Second, students answered a non-incentivized risk preference question taken from the 2004 wave of the German Socio-Economic Panel Study following Dohmen and Falk (2011), which finds that this question predicts incentivized lottery choices. The question is: How do you see yourself: Are you generally a person who is fully prepared to take risks or do you try to avoid taking risks? Check ONE box on the scale, where the value 0 means: 'unwilling to take risks' and the value 10 means: 'fully prepared to take risks'. Males choose a more risky lottery option and also choose a higher level of risk to describe themselves. In this sample, the correlation between these two measures is 0.243 overall, 0.208 for males and 0.230 for females ($p<0.001$ in both cases).

Females and males perform similarly on their school math midterm grades[16]; there is no gender difference (Appendix A-3 for detail). However, there is a significant female advantage for overall midterm grades: females have a 5 percentage point higher overall midterm grade than males (57.436 versus 52.414, $p=0.005$). Despite this academic context, the student survey responses show that male-favoring stereotypes exist for math and science and female-favoring stereotypes exist for reading, similar to Western stereotypes (Appendix A-4 for detail).

Females and males have similar levels of enjoyment of math; 74.3% of males and 69.7% of females agree or strongly agree that they like math (no significant difference) although a higher percentage of males than females like science while a higher percentage of females than males like reading ($p=0.015$, $p<0.001$ respectively). In addition, a higher percentage of males believe they are good at math; almost half of males

[16] Administrative grade data was obtained from four out of the five schools.

(47.2%) versus a little over a third of females (36.8%) agree or strongly agree that they are good at math (p=0.014). A similar pattern follows for science although it is reversed for reading; over three-quarters of females (77.2%) versus two-thirds of males (67.5%) think they are good at reading (p=0.010).

The science and math fields are most prestigious; 71.4% of all students rate the Science track as the best academic track in upper secondary school, with no statistically significant gender differences. A marginally higher percentage of males than females think that they will end up in the Science track in the next academic year, 47.6% versus 40.6% (p=0.097). On average, students believe that boys are better at math and science while girls are better at reading; males tend to rate boys as better in each of these subjects (Appendix A-4 for detail).

There do not appear to be gender differences in socioeconomic status (SES), using parental education as a proxy. On average, 45.1% of students' fathers and 36.7% of students' mothers hold at least bachelor's degrees (Appendix A-5).

[Table 3 about here]

Given that these variables may contribute to an individual's decision to enter into competition, it is important to control for these variables when determining whether there is a gender difference in competitiveness; that is, choosing competition for Test 3. The measure of competitiveness in this paper is similar to the measure first used in Niederle and Vesterlund (2007). Student choice of whether to enter into competition or piece-rate compensation prior to Test 3, controlling for other variables, is used as the measure of competitiveness (choosing competition is used interchangeably with choosing the tournament incentive for Test 3).

When controlling for only the score on the piece-rate test (Test 1) and the difference between the tournament and piece-rate scores (Test 2-Test 1), females are 17.3 percentage points less likely than males are to choose competition (Table 4, Model 1). When adding in the level of overconfidence, the difference decreases to 14.9 percentage points, which is different from the coefficient in Model 1 at the p=0.005 level[17] (Model 2). This difference remains largely stable when adding in both measures of risk preferences (Model 3), and is not significantly different from Model 2. When student attitudes and SES are added, the gender gap is 13.9 percentage points, although none of the coefficients for these characteristics appear to influence competition entry (Model 4). Lastly, although one school did not provide midterm scores, the gender gap remains when including math and overall midterm grades in addition to all the other covariates (Model 5).

[Table 4 about here]

Similar results hold for the previous models when this school is excluded from the analyses (Appendix A-6) or when session fixed effects are used instead of class fixed effects to account for simultaneous experimental sessions (Appendix A-7). Thus, the gender gap is still significant although the power from the reduced sample size is lower, and is very similar to the gap found in a similar age sample of ninth-grade students in the Netherlands (Buser et al., 2014).

Secondary students in Malaysia show the standard gender gap in choosing competition that has been demonstrated in many different contexts. When only

[17] Comparisons of the coefficient for female use seemingly unrelated estimations with clustered standard errors (not exact standard errors from main analyses, since Stata's suest command does not accept xtreg models). The coefficient for "Female" in Model 1 is significantly different from the coefficients in Models 2-4 at the $p<0.10$ level.

controlling for previous performance, the gender gap is 17.3 percentage points. The gender gap is reduced a total of about 20% when controlling for confidence, risk preferences, student attitudes about math and socioeconomic status, but females are still 13.9 percentage points less likely than males to choose competition ($p<0.05$).

3.2 Other-class competition analysis

The previous analysis confirms that the standard gender gap in choosing into math competition exists for this sample of secondary school students. This section focuses on the novel contribution of this paper: how students react to different levels of competition. I present several descriptive findings of the difference in performance when facing different competitors. I then present the experimental results in addition to exploring heterogeneity in these results and whether changes in questions answered or accuracy led to these results.

The sample for the following analyses is restricted to the middle-ranked (middle) classes so that there are both easier (bottom class) and harder (top class) competitors. There are 266 students in 8 middle classes (137 male and 129 female), which represents a little less than half the number of students in the original sample. As described in Section 2.3, students in the middle classes were randomized to compete against either the top ranked class or the bottom ranked class in the same grade and school, although classes were only named by their official titles as to not directly prime students to the level of their competitors. Students received a slip of paper informing them which class their competitors would come from, and were told to put the slip of paper in an envelope and not talk so that treatment assignments remained concealed.

As in the overall sample, there is a general upward trend in the number of correct answers in successive tests, which suggests that learning[18] could play a role in the observed scores (Table 5). This brings up concerns about whether the observed scores reflect learning or ability rather than the effort put into the task. The randomization should alleviate these concerns for this last round, unless learning or ability is not balanced within genders across treatment groups. The randomization produced balanced groups competing against higher and lower competitors across all observable baseline characteristics (gender, math midterm score and overall midterm score). In addition, most student characteristics measured prior to treatment are balanced across groups, including scores on Test 2, Test 3, the difference between Test 2 and Test 1, and the competition choice. Treatment assignment predicts the score on Test 1 at the 10% significance level, although there is no significant correlation between treatment and Test 1 score within gender (Appendix A-8). The following analyses control for Test 1 score, difference between Test 1 and Test 2 score, and competition choice as robustness checks.

[Table 5 about here]

Although the upward trend in scores on successive tests is clear in the treatment against the bottom class, it is less apparent for those who competed against the top class. However, the incentives between the third and fourth test vary by student choice thus it is most relevant to compare results from Test 4 against Test 2.

---

[18] A limitation of this study is the difficulty in separating out learning effects and response to incentives, given that the order of the rounds remained constant in order to replicate the Niederle and Vesterlund (2007) experiment to determine gender differences in competition. Cotton and colleagues show that repeated competition eliminates the gender gap in performance in their study (Cotton et al., 2013). The results from the current study show some indication that genders may perform differently in successive competitions. The average scores increase from Test 1 to Test 2 for both genders, for only females from Test 2 to Test 3, and then do not increase from Test 3 to Test 4 for either gender. However, there is no indication that males lower their performance during successive rounds, unlike what Cotton and colleagues find.

In the following analysis, the primary variable of interest is the difference between performance in Test 2 and Test 4. Similar variables are constructed for the difference between total number of questions answered and the difference in accuracy of answers, which are used to explore the main results. Thus, a student's performance against another class (Test 4) is compared against performance against a student's own class (Test 2). This within-subject design allows us to see the effect of a different level of competitor using each subject's baseline value (i.e. performance on Test 2). The average value of the difference in the number of correct answers from Test 2 to Test 4 is 1.34 with a standard deviation of 2.90 and a range of -7 to 10. As Figure 1 shows, there is no gender difference in the change in performance when the competitors are from the bottom class. Both genders perform about 1.5 questions better. However, when matched against competitors from the top class, females increase the number of correct answers by significantly more than males, 1.806 correct answers compared to 0.507 correct answers (Appendix A-9, p=0.017).

[Figure 1 about here]

Since treatment is randomized within class, the following equation can be used to determine the effect of the treatment.

$$y_{ij} = \Gamma_j + \beta_1 Treatment_{ij} + \beta_2 Female_{ij} + \beta_3 (Treatment * Female)_{ij} + \theta \mathrm{X_{ij}} + \epsilon_{ij}$$

where:

y_{ij} is the difference in number of correct answers between other and own class (Test 4 - Test 2) for student i in class j

Γ_j is the class fixed effects

Treatment is 1 if assigned to the top class and 0 if assigned to the bottom class for student i in class j

Female is 1 if female and 0 if male for student i in class j

Treatment * Female is 1 if student i in class j is assigned to the top class and is female; 0 otherwise. This represents the gender difference in the effect of treatment on the difference of performance between other and own class

X_{ij} is a vector of student attributes

The regressions in Table 6 show the effects of competing against the top class (competition against bottom class as reference group), relative to competing against the student's own class. Since the treatments were randomly assigned, the estimates of the effect of the treatment can be directly interpreted. Baseline covariates are included in subsequent models, which lowers the precision of the estimates (Columns 2-3). The regressions are also performed separately for males (Columns 4-6) and females (Columns 7-9).

The effect of facing the top class versus the bottom class is about one question less, -1.029 ($p<0.05$) (Table 6, Column 1). However, the interaction effect of being female and facing the top class is positive and similar in magnitude to this negative effect, 1.184 ($p<0.10$). When adding in baseline variables including Test 1 performance, response to competition incentive (difference in Test 1 and Test 2 performance), and competition choice, the pattern remains similar; there is a stable negative main effect although precision decreases so that the female interaction effect is not statistically significantly different from zero (Table 6, Column 3).

[Table 6 about here]

The gender difference in response to harder competition is clearer when examining the regression results separately by gender (Table 6, Columns 4-9). The effect of facing the top class instead of the bottom class is consistently negative and close to 1 question for males after controlling for behavior and performance from prior rounds[19],

[19] When the baseline variables are added in models 3, 6 and 9, the difference between Test 1 and Test 2 performance (T-PR) shows a consistently large negative coefficient, which could possibly be due to ceiling effects.

ranging from -0.996 to -0.893 (Table 6, Columns 4-6). On the other hand, females do not seem affected by facing the top class as opposed to the bottom class; the effect is not statistically different from zero (Table 6, Columns 7-9). These findings indicate that males are negatively affected by facing a difficult competitor while females are not. Qualitatively similar results hold when the whole sample of students is included and treatment is defined as competing against any higher class (Appendix A-10), session fixed effects are used (Appendix A-11) or absolute score on Test 4 is used controlling for Test 2 performance and other variables (Appendix A-12). Males perform worse when competing against the top class rather than the bottom class, even after controlling for prior performance and competitive behavior, while there is no evidence that females perform differently according to the level of their competitors.

To explore these results, I examine heterogeneity in the sample in addition to whether the effects are due to differential numbers of questions answered or a change in the accuracy of answers.

An important characteristic of this sample is the variance in performance both within schools (e.g. average scores in middle classes compared to top classes) and across schools. All previous results include class fixed effects, which help capture this heterogeneity. However, it is also instructive to view these results in a more easily comparable manner such as the chance of winning against the top class. The chance of winning against the top class conditional on the number of correct answers varies by school; for example, with 18 correct answers, a student in a middle class at School 4 has an 83% chance of winning, while a student in a middle class at School 5 has a 9% chance of winning (Table 7). When the chances of winning are used as controls instead of the numbers of answers correct, the effects of facing harder competition remain negative for males and null for females (Appendix A-13).

[Table 7 about here]

These effects of facing more difficult competitors appear to differ along the distribution of baseline performance by gender. For males, the difference between Test 2 and Test 4 score is greatest at the best and worst quintiles of the baseline (Test 1) performance distribution (Figure 2). Males at the best and worst quintiles who face the top class perform about two questions worse than males who face the bottom class. Females in the top two quintiles perform similarly when facing either the top or bottom class, although females in the bottom two quintiles who face the top class appear to perform a little better than those who face the bottom class. Overall, it appears that males from the top and bottom of the performance distributions respond most to the level of competition.

The change in performance from Test 2 to Test 4 could be due to a combination of the quantity and accuracy of answered questions. For example, individuals can obtain a higher score by answering more questions with the same (or lower) level of accuracy or by answering the same number (or fewer) of questions with higher accuracy. It appears that competitor difficulty has no effect on the number of questions answered; there is a negative effect for males that is not significant after controlling for prior number of questions answered and competitive behavior (Table 8).

[Table 8 about here]

However, males but not females are less accurate when facing more difficult competitors; the difference between females and males when facing harder competition is about 5 percentage points and significant at the 5% level (Table 9, column 3). After controlling for prior accuracy and competitive behavior, the accuracy of males who face harder competitors is a little over 3 percentage points (significant at the 10% level) less than the accuracy of males who face easier competitors (Table 9, column 6). Thus, it appears that males change the quality (accuracy) of performance rather than the quantity of effort against more difficult competition.\

[Table 9 about here]

4. Discussion

This study shows the robustness of the gender gap in competition. Overall, females choose into competition at about half the rate of males—20.7% versus 37.9%. After controlling for student performance, confidence, risk preferences, and other student characteristics, females still have a 13.9 percentage point lower probability of choosing into competition less than males. This gender gap is very similar to what is found in the Netherlands with a similar age group and experiment protocol, although the overall rates of competition are lower in Malaysia.

There is another gender gap that emerges when facing different levels of competitors. The performance of females is not affected by facing harder competitors. However, males perform almost one question worse when facing competitors from the top class (about one-third of a standard deviation) than when competing against the bottom class. It appears that accuracy decreases for males when facing the top class compared to the bottom class. There may be several explanations for the gender difference in performance against harder competitors, such as the gender composition of groups, differential expectations when facing different classes or changes in the chance of winning or expected earnings.

One possible explanation for these results may be the gender composition of the competitor groups. Existing research indicates that the gender composition of competitors can affect performance in competitions (Booth and Nolen, 2012; De Paola et al., 2015; Gneezy et al., 2003; Kuhnen and Tymula, 2011). Thus, the perceived gender composition of the competitors could also play a role in these results. As noted in Section 2.2, there is a higher proportion of females in the top classes than in the middle or bottom classes, although the difference is not statistically significant between the top and middle classes, which is the relevant comparison in these analyses. The range in female composition of the top class across the five schools in the study is reasonably small, from 48.48% to 60.71%. These factors make it unlikely that the female composition of the top classes affected results.

These results could also be explained by different expectations between genders when competing against harder or easier competition, and a corresponding differential change in effort. For example, Kuhnen and Tymula (2011) use gender composition of the group as a proxy for perceived difficulty of competitor and find that females have lower output, worse expected rank and worse actual rank with more males in their group while males are not affected by the gender composition of the group. However, gender composition of the group may be an inappropriate proxy for perceived difficulty of competitors. It is worth noting that they observe that males expect better rankings than females (similar to this study) yet males also outperform females (different from this study).

I use a similar task but more clearly designated groups of easier or harder competitors and find that expectations of males rather than that of females appear to be affected. There are no gender differences in the actual rankings in either treatment condition, although both genders guess a better rank when competing against the bottom class (Table 10). These rankings also confirm that the difficulty levels of competitors are appropriately categorized; students in the sample have a 55% chance of winning the tournament against the bottom class and a 16% chance of winning against the top class, with no gender difference. However, males guess they are a better rank than females do and are more overconfident when facing the bottom class (p-values 0.019 and 0.061, respectively). There are no gender differences in guessed rank or overconfidence when facing the top class, although males are slightly overconfident and females are under-confident. Since baseline measures of confidence against different classes were not elicited in this study in order to prevent priming, it is not possible to distinguish whether the treatment of facing more difficult competition changed male and female priors about their performance differentially. Nevertheless, these ex-post elicited measures of confidence could indicate a possible mechanism difference between genders; that is, males may lower performance because they expect to do worse against harder competition (on par with females' confidence), relative to their confidence against easier competition (more confident than females).

[Table 10 about here]

Finally, there is a negative effect on the chance of winning (Table 11) and expected earnings (Table 12) when facing harder competition for both females and males. The relatively lower performance of males when facing harder competition does not appear to result in a lower chance of winning or decreased expected earnings for males.

Thus, the lower performance of males could reflect greater efficiency (e.g. lower performance for the same financial outcomes).

[Table 11 about here]

[Table 12 about here]

The gender difference in performance under more difficult competition is somewhat surprising, given findings from previous literature which generally show an equal response if not female disadvantage when encountering difficult competition. For example, Eriksson et al. (2009) finds that relative information does not affect performance, Gill and Prowse (2014) finds that both genders lower performance after a loss and Buser (2016) finds that females lower their performance after a loss but males do not.

However, this study design does not depend on explicit information, as previous studies have used, but a more realistic yet less certain competitive situation. The experiment exploited pre-existing differences in levels of competitors without an explicit message about relative position, which could affect the dynamics in competition. There is suggestive evidence that males may have lowered expectations when facing harder competition, although the gender gap in the effect of facing harder competition on performance does not appear to extend to a gender difference in the chance of winning or expected earnings.

5. Conclusion

This paper presents experimental evidence that females and males have different reactions to more difficult competitors—males lower their performance while females' performance does not change. In addition, it appears that standard gender differences in competitive behavior apply even within a STEM context with more female participation.

Given the similar gender gaps in competition choice, it is reasonable to believe that these findings about reacting to harder competition apply in broader contexts.

The results from this study confirm the gender gap in choosing into competition in a math task similar to those that have been linked to future educational choices. Although several previous studies have found that females perform worse than males in competition, the current study adds to the body of literature that finds no gender difference in competitive performance. Furthermore, the within-subject study design shows a gender difference in the response to harder or easier competition.

These findings have implications for policies designed to attract females into more competitive environments. Existing research clearly indicates that, when given a choice, females choose into competition less than males do. There are many situations in which people face competition choices, such as which courses to take in school or which jobs to apply for. Early decisions could have lasting consequences; for example, there may be prerequisite courses for certain majors which are required to pursue certain occupations (e.g. advanced math/science courses required for engineering degrees to become an engineer). If females differentially decline to enter into competition early, gender gaps may widen over time as fewer opportunities remain open.

However, it appears that females may not be negatively affected by the level of competition once they are in a more competitive situation. Thus, if females do not perform worse in more competitive environments even when they do not choose into these environments, perhaps policies can be designed to compel people into more difficult competitive environments. For example, schools could require more advanced STEM courses or companies could provide mandatory leadership programs, which would require females who may not otherwise choose those programs to participate in them. Then, they may thrive in the more competitive environment. On the other hand, it is important to ensure that males do not perform worse in these more demanding situations where there could be negative outcomes from lowered performance. The results of this study are found in a sample of students in middle-ranked classes with no gender differences in performance, thus these proposed policies may not apply among high or low performance individuals or when gender differences in performance exist. These policies also do not address other barriers such as chilly climates that females face in competitive environments.

Future research could look at the generalizability of and possible mechanisms underlying the results. This study was conducted among secondary students in middle-ranked classes in an Asian country; it would be illuminating to see whether the results hold among different ages, performance levels or cultural contexts. In addition to addressing generalizability, future studies can examine more deeply the potential mechanisms for these results, such as a differential change in expectations when facing different levels of competition. Other possibilities from the psychology literature could be differences in persistence or grit; for example, females may be grittier than males in learning environments. Thus, even if females would not choose more competitive environments, they could persist and succeed in them. Understanding these mechanisms could help design policies that could result in greater participation and performance in environments with more difficult competition.

References

Almås, I., Cappelen, A. W., Salvanes, K. G., Sørensen, E. Ø., & Tungodden, B. (2016). What Explains the Gender Gap in College Track Dropout? Experimental and Administrative Evidence. *American Economic Review*, 106(5), 296–302.

Azmat, G., & Iriberri, N. (2010). The importance of relative performance feedback information: Evidence from a natural experiment using high school students. *Journal of Public Economics*, *94*(7–8), 435–452.

Booth, A., & Nolen, P. (2012). Choosing to compete: How different are girls and boys? *Journal of Economic Behavior & Organization*, 81(2), 542–555.

Buser, T. (2016). The Impact of Losing in a Competition on the Willingness to Seek Further Challenges. *Management Science, 62*(12), 3439 - 3449.

Buser, T., Niederle, M., & Oosterbeek, H. (2014). Gender, Competitiveness, and Career Choices. *The Quarterly Journal of Economics*, *129*(3), 1409–1447.

Buser, T., Peter, N., & Wolter, S. C. (2017). Gender, Competitiveness, and Study Choices in High School: Evidence from Switzerland. *American Economic Review*, 107(5), 125–130.

Cárdenas, J.-C., Dreber, A., von Essen, E., & Ranehill, E. (2012). Gender differences in competitiveness and risk taking: Comparing children in Colombia and Sweden. *Journal of Economic Behavior & Organization*, 83(1), 11–23.

Cason, T. N., Masters, W. A., & Sheremeta, R. M. (2010). Entry into winner-take-all and proportional-prize contests: An experimental study. *Journal of Public Economics*, *94*(9–10), 604–611.

Cotton, C., McIntyre, F., & Price, J. (2013). Gender differences in repeated competition: Evidence from school math contests. *Journal of Economic Behavior & Organization*, *86*(C), 52–66.

Curriculum Development Division, Ministry of Education Malaysia. 2016. Sharing Malaysian Experience in Participation of Girls in STEM Education. Geneva, Switzerland, UNESCO International Bureau of Education (IBE).

Datta Gupta, N., Poulsen, A., & Villeval, M. C. (2013). Gender Matching and Competitiveness: Experimental Evidence. *Economic Inquiry,* 51(1), 816–835.

De Paola, M., Gioia, F., & Scoppa, V. (2015). Are females scared of competing with males? Results from a field experiment. *Economics of Education Review*, *48*(C), 117–128.

Dohmen, T., & Falk, A. (2011). Performance Pay and Multidimensional Sorting: Productivity, Preferences, and Gender. *American Economic Review*, *101*(2), 556–90.

Dreber, A., Essen, E., & Ranehill, E. (2014). Gender and competition in adolescence: task matters. *Experimental Economics*, *17*(1), 154–172.

Eccles, J., Wigfield, A., Harold, R. D., & Blumenfeld, P. (1993). Age and gender differences in children's self-and task perceptions during elementary school. *Child Development*, 64(3), 830–847.

Eckel, C. C., & Grossman, P. J. (2002). Sex Differences and Statistical Stereotyping in Attitudes Toward Financial Risk. *Evolution and Human Behavior*, *23*(4), 281–295.

Eriksson, T., Poulsen, A., & Villeval, M. C. (2009). Feedback and incentives: Experimental evidence. *Labour Economics*, *16*(6), 679–688.

Ertac, S., & Szentes, B. (2011). The Effect of Information on Gender Differences in Competitiveness: Experimental Evidence (Koç University-TUSIAD Economic Research Forum Working Paper No. 1104). Koc University-TUSIAD Economic Research Forum.

Gill, D., & Prowse, V. (2014). Gender differences and dynamics in competition: The role of luck. *Quantitative Economics*, *5*(2), 351–376.

Gneezy, U., Leonard, K. L., & List, J. A. (2009). Gender Differences in Competition: Evidence From a Matrilineal and a Patriarchal Society. *Econometrica*, 77(5), 1637–1664.

Gneezy, U., Niederle, M., & Rustichini, A. (2003). Performance in Competitive Environments: Gender Differences. *The Quarterly Journal of Economics*, *118*(3), 1049–1074.

Gneezy, U., & Rustichini, A. (2004). Gender and Competition at a Young Age. *American Economic Review*, 94(2), 377–381.

Grosse, N. D., & Riener, G. (2010). Explaining Gender Differences in Competitiveness: Gender-Task Stereotypes (Jena Economic Research Paper No. 2010-017). Friedrich-Schiller-University Jena, Max-Planck-Institute of Economics.

Günther, C., Ekinci, N. A., Schwieren, C., & Strobel, M. (2010). Women can't jump?—An experiment on competitive attitudes and stereotype threat. *Journal of Economic Behavior & Organization*, 75(3), 395–401.

Harbaugh, W., Krause, K., & Vesterlund, L. (2002). Risk Attitudes of Children and Adults: Choices Over Small and Large Probability Gains and Losses. *Experimental Economics*, 5(1), 53–84.

Kamas, L., & Preston, A. (2010). Are Women Really Less Competitive Than Men? Working Paper, Santa Clara University.

Kuhnen, C. M., & Tymula, A. (2011). Feedback, Self-Esteem, and Performance in Organizations. *Management Science*, 58(1), 94–113.

Mellström, U. (2009). The Intersection of Gender, Race and Cultural Boundaries, or Why is Computer Science in Malaysia Dominated by Women? *Social Studies of Science*, *39*(6), 885–907.

Ministry of Education Malaysia. (2014). Quick Facts 2014. Retrieved 19 February 2016 from http://www.moe.gov.my/cms/upload_files/publicationfile/2014/pubfile_file_002100.pdf

Ministry of Education Malaysia. (2015). Quick Facts 2015. Retrieved 19 February 2016 from http://www.moe.gov.my/cms/upload_files/publicationfile/2015/pubfile_file_002101.pdf

Niederle, M. (2016). Gender. In J. Kagel & A. E. Roth (Eds.), *Handbook of Experimental Economics* (2nd ed., pp. 481–553). Princeton University Press.

Niederle, M., Segal, C., & Vesterlund, L. (2013). How costly is diversity? Affirmative action in light of gender differences in competitiveness. *Management Science*, *59*(1), 1–16.

Niederle, M., & Vesterlund, L. (2007). Do Women Shy Away from Competition? Do Men Compete Too Much? *The Quarterly Journal of Economics*, 122(3), 1067–1101.

Niederle, M., & Vesterlund, L. (2011). Gender and Competition. *Annual Review in Economics*, *3*, 601–630.

Niederle, M., & Yestrumskas, A. H. (2008). Gender Differences in Seeking Challenges: The Role of Institutions (Working Paper No. 13922). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w13922

Ors, E., Palomino, F. & Peyrache, E. (2013). Performance Gender Gap: Does Competition Matter? *Journal of Labor Economics*, 31(3), 443–499.

Shurchkov, O. (2012). Under Pressure: Gender Differences in Output Quality and Quantity Under Competition and Time Constraints. *Journal of the European Economic Association*, *10*(5), 1189–1213.

Sutter, M., & Glätzle-Rützler, D. (2014). Gender Differences in the Willingness to Compete Emerge Early in Life and Persist. *Management Science*, *61*(10), 2339–2354.

Wiseman, A. W. (2008). A culture of (in) equality?: A cross-national study of gender parity and gender segregation in national school systems. *Research in Comparative and International Education*, *3*(2), 179–201.

World Economic Forum. (2014). The Global Gender Gap Report 2014. Retrieved from http://reports.weforum.org/global-gender-gap-report-2014/

Wozniak, D., Harbaugh, W. T., & Mayr, U. (2014). The Menstrual Cycle and Performance Feedback Alter Gender Differences in Competitive Choices. *Journal of Labor Economics*, *32*(1), 161 – 198.

Wozniak, D., Harbaugh, W. T., & Mayr, U. (2016). The Effect of Feedback on Gender Differences in Competitive Choices. (SSRN Scholarly Paper No. ID 1976073).

Zhang, Y. J. (2013). Can Experimental Economics Explain Competitive Behavior Outside the Lab? (SSRN Scholarly Paper No. ID 2292929). Rochester, NY: Social Science Research Network. Retrieved from http://papers.ssrn.com/abstract=2292929

Tables

Table 1. Descriptive statistics of number of correct answers and competition choice, by class level.

Variable	Class level	Overall	Male	Female	Diff	p-value
Test 1 (Piece-Rate)	Overall	10.141	9.693	10.620	-0.927	0.040
	Bottom	5.937	5.908	5.980	-0.072	0.948
	Middle	10.677	10.307	11.070	-0.763	0.173
	Top	12.432	12.338	12.511	-0.173	0.847
Test 2 (Tournament)	Overall	12.041	11.710	12.395	-0.684	0.082
	Bottom	7.746	7.789	7.680	0.109	0.785
	Middle	12.549	12.482	12.620	-0.138	0.650
	Top	14.444	14.208	14.641	-0.434	0.354
T-PR	Overall	1.900	2.017	1.775	0.242	0.578
	Bottom	1.810	1.882	1.700	0.182	0.469
	Middle	1.872	2.175	1.550	0.625	0.205
	Top	2.012	1.870	2.130	-0.260	0.378
Competition choice	Overall	0.296	0.379	0.207	0.173	<0.001
	Bottom	0.325	0.395	0.220	0.175	0.041
	Middle	0.271	0.321	0.217	0.104	0.057
	Top	0.314	0.468	0.185	0.283	<0.001

Number of observations are from the whole sample: 561 overall, with 290 males and 271 females overall. The gender breakdown is: 76 males and 50 females in the bottom classes; 137 males and 129 females in the middle classes; 77 males and 92 females in the top classes. T-PR is the difference between number correct on the tournament (Test 2) versus piece-rate test (Test 1). Competition choice is the proportion that chose the tournament rather than the piece-rate incentive. P-values are from Mann-Whitney *U* tests.

Table 2. Change in number of correct answers between Test 2 and Test 3.

	Overall		Chose Piece-rate		Chose Competition			
	Mean	N	Mean	N	Mean	N	Diff	p-value
Overall	0.720	561	0.681	395	0.813	166	-0.132	0.518
Male	0.638	290	0.600	180	0.700	110	-0.100	0.771
Female	0.808	271	0.749	215	1.036	56	-0.287	0.385

Differences are calculated by student (Test 3-Test 2). P-values are from Mann-Whitney *U* tests.

Table 3. Descriptive statistics of student characteristics.

Variable	Overall Mean	Overall N	Male Mean	Male N	Female Mean	Female N	Diff	p-value
Confidence								
Guessed Rank Test 2	2.573	560	2.441	290	2.715	270	-0.273	0.001
Guessed Rank Test 4	2.455	560	2.360	289	2.557	271	-0.197	0.028
Overconfidence Test 2	-0.221	560	-0.097	290	-0.356	270	0.259	0.016
Overconfidence Test 4	-0.136	560	0.042	289	-0.325	271	0.366	<0.001
Risk								
Incentivized risk scale (1-5; 5 most risky)	2.588	561	3.103	290	2.037	271	1.067	<0.001
Non-incentivized risk scale (0-10; 10 most risky)	6.161	559	6.410	288	5.897	271	0.513	0.001
Midterm scores								
Math	49.842	463	49.457	230	50.223	233	-0.767	0.852
Overall GPA	54.936	432	52.414	215	57.436	217	-5.022	0.005
Attitudes and Beliefs								
Like Math	0.721	555	0.743	284	0.697	271	0.046	0.232
Like Science	0.770	556	0.812	287	0.725	269	0.087	0.015
Like Reading	0.752	537	0.647	275	0.863	262	-0.215	<0.001
Good at Math	0.422	552	0.472	286	0.368	266	0.104	0.014
Good at Science	0.410	554	0.455	286	0.362	268	0.093	0.027
Good at Reading	0.722	554	0.675	286	0.772	268	-0.098	0.010
Rank Science 1	0.714	532	0.722	270	0.706	262	0.016	0.681
Guess Science Stream	0.442	559	0.476	288	0.406	271	0.070	0.097
Stereotype views								
Gender better at math (-1 to 1)	-0.220	549	-0.270	282	-0.169	267	-0.101	0.053
Gender better at science (-1 to 1)	-0.160	550	-0.236	284	-0.079	266	-0.157	0.003
Gender better at reading (-1 to 1)	0.376	553	0.320	284	0.435	269	-0.115	0.048
Socioeconomic status								
Father is college grad	0.451	552	0.483	286	0.417	266	0.065	0.124
Mother is college grad	0.367	551	0.384	284	0.348	267	0.035	0.388

Guess Rank ranges from 1 to 4 (1 is the best rank and 4 is the worst rank). Overconfidence is calculated as Actual-Guessed rank (actual rank based on modal rank in 1,000 simulations). Midterm scores are available for 4 schools, and are on a scale of 0-100. Attitudes and beliefs are based on dichotomized variables where 1=yes/agree and 0=no/disagree, except for "Gender better at" questions which are coded -1 (Boys are better) 0 (Both are equally as good) and 1 (Girls are better). Socioeconomic status are dichotomized variables for each parent holding at least a bachelor's degree. P-values are from Mann-Whitney *U* tests.

Table 4. Models for tournament entry (Competitiveness).

	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Female	-0.173**	-0.149**	-0.145**	-0.139**	-0.150*
	(0.048)	(0.047)	(0.044)	(0.047)	(0.056)
Num. Correct-Test 1	0.018**	0.026***	0.024***	0.026**	0.028**
	(0.005)	(0.006)	(0.006)	(0.007)	(0.008)
T-PR	0.009	0.018*	0.015*	0.019**	0.019*
	(0.008)	(0.007)	(0.006)	(0.006)	(0.007)
Overconfidence Test 2		0.060**	0.051*	0.053**	0.067*
		(0.017)	(0.018)	(0.018)	(0.022)
Nonincentivized risk			0.025+	0.024*	0.037**
			(0.012)	(0.011)	(0.011)
Incentivized risk			-0.004	-0.004	-0.020*
			(0.012)	(0.013)	(0.009)
Math stereotype				-0.013	-0.036
				(0.038)	(0.045)
Likes math				0.006	-0.010
				(0.037)	(0.050)
Thinks is good at math				-0.010	-0.018
				(0.044)	(0.058)
Expects science stream				0.035	0.074
				(0.042)	(0.042)
Father is college grad				-0.033	-0.056
				(0.060)	(0.055)
Mother is college grad				0.009	-0.017
				(0.059)	(0.070)
Midterm math score					-0.000
					(0.002)
Midterm overall score					-0.005
					(0.004)
Observations	561	560	558	524	409

All models provide OLS linear probability results that include class fixed effects. T-PR is the difference between number correct on the tournament (Test 2) versus piece-rate test (Test 1). Overconfidence Test 2 is measured as the difference between Actual and Guessed rank on Test 2. Nonincentivized risk is a scale from 0 to 10 (10 is most risky). Incentivized risk is the choice between a certain option or set of lotteries, ranging from 1 to 5 (5 is most risky). Robust standard errors are provided in parentheses. Significance levels are set at + p<0.10 * p<0.05 ** p<0.01 *** p<0.001.

Table 5. Number of correct answers, by treatment condition.

				Treatment Condition					
		Overall		Bottom Class		Top Class			
		Mean	N	Mean	N	Mean	N	Diff	p-value
Overall	Test 1	10.677	266	10.977	133	10.376	133	0.602	0.228
	Test 2	12.549	266	12.714	133	12.383	133	0.331	0.567
	Test 3	13.429	266	13.669	133	13.188	133	0.481	0.387
	Test 4	13.891	266	14.286	133	13.496	133	0.789	0.253
Males	Test 1	10.307	137	10.848	66	9.803	71	1.046	0.134
	Test 2	12.482	137	12.636	66	12.338	71	0.298	0.725
	Test 3	13.263	137	13.652	66	12.901	71	0.750	0.436
	Test 4	13.467	137	14.136	66	12.845	71	1.291	0.270
Females	Test 1	11.070	129	11.104	67	11.032	62	0.072	0.870
	Test 2	12.620	129	12.791	67	12.435	62	0.356	0.623
	Test 3	13.605	129	13.687	67	13.516	62	0.170	0.631
	Test 4	14.341	129	14.433	67	14.242	62	0.191	0.627

Analyses are limited to the sample of students in the middle classes. P-values are from Mann-Whitney *U* tests.

Table 6. Change in number of correct answers between Test 2 and Test 4 due to level of competition.

	All			**Male**			**Female**		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Versus top class	-1.029*	-1.043*	-0.815+	-0.996*	-1.008*	-0.893+	0.181	0.179	0.059
	(0.342)	(0.348)	(0.421)	(0.326)	(0.325)	(0.392)	(0.603)	(0.622)	(0.626)
Female	-0.002	-0.053	-0.112						
	(0.564)	(0.576)	(0.615)						
Female * Vs top class	1.184+	1.197+	0.864						
	(0.596)	(0.618)	(0.756)						
Competition		-0.448	-0.305		-0.380	-0.237		-0.722	-0.598
		(0.257)	(0.244)		(0.417)	(0.362)		(0.563)	(0.583)
Test 1			-0.047			-0.080+			0.016
			(0.035)			(0.037)			(0.049)
T-PR			-0.403***			-0.284**			-0.485***
			(0.045)			(0.066)			(0.079)
Observations	266	266	266	137	137	137	129	129	129

Analyses are limited to the sample of students in the middle classes. All models provide OLS linear probability results that include class fixed effects. Competition is the competition choice prior to Round 3. Test 1 is the number of correct answers on Test 1 (Piece-Rate). T-PR is the difference between number correct on the tournament (Test 2) versus piece-rate test (Test 1). Robust standard errors are provided in parentheses. Significance levels are set at: + p<0.10 * p<0.05 ** p<0.01 *** p<0.001.

Table 7. Chance of winning in Test 4 against top class, by school.

Questions	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**25**	**29**
School 1	0.2	0.6	1.2	1.4	5.7	10.8	12.4	17.5	30.5	50.1	71.3	91.5	93.2	-	-	-
School 2	0.5	0.5	1	1.1	3.1	-	20.3	23.4	39.2	55.6	-	-	81.5	-	-	-
School 3	0.3	0.3	-	-	-	3.4	-	-	24.5	-	53.8	-	-	-	-	-
School 4	1.7	-	21.5	-	-	53.1	66.7	-	83.1	-	-	-	-	-	-	-
School 5	0	0	0	0.8	1	2.4	3	7	9.4	13.7	23.7	39.4	-	52.6	61.8	100

Analyses only include the sample of students in the middle classes who face the top class. The chance of winning in Test 4 is the chance of getting 1st place in a group of 4 total competitors: the individual and 3 competitors from the top class at the same school (percentages are obtained by simulating 1,000 random draws of groups of 3 competitors from the top class for each individual).

Table 8. Change in number of answered questions between Test 2 and Test 4 due to level of competition.

	All			Male			Female		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Versus top class	-0.645*	-0.637*	-0.465	-0.616*	-0.610*	-0.448	-0.146	-0.146	-0.234
	(0.238)	(0.238)	(0.254)	(0.250)	(0.252)	(0.256)	(0.423)	(0.421)	(0.455)
Female	0.062	0.088	0.085						
	(0.369)	(0.373)	(0.400)						
Female * Vs top class	0.474	0.468	0.212						
	(0.404)	(0.393)	(0.539)						
Competition		0.223	0.174		0.182	0.165		0.268	0.060
		(0.224)	(0.210)		(0.173)	(0.226)		(0.480)	(0.461)
Test 1			0.029			-0.012			0.096*
			(0.032)			(0.036)			(0.038)
T-PR			-0.235***			-0.296***			-0.144+
			(0.036)			(0.042)			(0.070)
Observations	266	266	266	137	137	137	129	129	129

Analyses are limited to the sample of students in the middle classes. All models provide OLS linear probability results that include class fixed effects. Competition is the competition choice prior to Round 3. Test 1 is the number of total (incorrect + correct) answers on Test 1 (Piece-Rate). T-PR is the difference between number of total answers on the tournament (Test 2) versus piece-rate test (Test 1). Robust standard errors are provided in parentheses. Significance levels are set at: + p<0.10 * p<0.05 ** p<0.01 *** p<0.001.

Table 9. Change in accuracy between Test 2 and Test 4 due to level of competition.

	All			Male			Female		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Versus top class	-0.067+	-0.067+	-0.042*	-0.066+	-0.066+	-0.034+	0.020	0.019	0.008
	(0.035)	(0.035)	(0.015)	(0.034)	(0.034)	(0.015)	(0.021)	(0.022)	(0.016)
Female	-0.014	-0.017	-0.013						
	(0.033)	(0.033)	(0.022)						
Female * Vs top class	0.086+	0.086+	0.051*						
	(0.042)	(0.042)	(0.019)						
Competition		-0.023	-0.020		-0.004	-0.013		-0.062*	-0.037+
		(0.024)	(0.015)		(0.039)	(0.027)		(0.022)	(0.019)
Test 1			-0.519***			-0.592***			-0.566***
			(0.050)			(0.100)			(0.070)
T-PR			-0.875***			-1.021***			-0.785***
			(0.095)			(0.164)			(0.067)
Observations	266	266	266	137	137	137	129	129	129

Analyses are limited to the sample of students in the middle classes. All models provide OLS linear probability results that include class fixed effects. Competition is the competition choice prior to Round 3. Test 1 is the percentage of correct answers on Test 1 (Piece-Rate). T-PR is the difference between percentages of correct answers on the tournament (Test 2) versus piece-rate test (Test 1). Robust standard errors are provided in parentheses. Significance levels are set at: + p<0.10 * p<0.05 ** p<0.01 *** p<0.001.

Table 10. Confidence on Test 4 by treatment and gender.

Variable	Treatment	Overall Mean	Overall N	Male Mean	Male N	Female Mean	Female N	Diff	p-value
Actual Rank	Bottom class	1.541	133	1.500	66	1.582	67	-0.082	0.830
	Top class	2.932	133	2.944	71	2.919	62	0.024	0.761
Guessed Rank	Bottom class	1.962	133	1.758	66	2.164	67	-0.407	0.019
	Top class	3.000	133	2.930	71	3.081	62	-0.151	0.685
Overconfidence	Bottom class	-0.421	133	-0.258	66	-0.582	67	0.325	0.061
	Top class	-0.068	133	0.014	71	-0.161	62	0.175	0.579
Probability of win	Bottom class	55.024	133	56.114	66	53.951	67	2.162	0.601
	Top class	15.794	133	15.437	71	16.203	62	-0.767	0.474

Analyses are limited to the sample of students in the middle classes. Actual rank is based on modal rank on Test 4 based on 1,000 simulations. Guessed rank is from the survey question asking students to guess their rank. Overconfidence is the difference between Actual and Guessed rank. Probability of win is calculated as the percentage of wins (i.e. rank 1) based on the 1,000 simulations. P-values are from Mann-Whitney *U* tests.

Table 11. Change in chance of winning Test 4 due to level of competition.

	All			Male			Female		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Versus top class	-40.047***	-39.864***	-36.768**	-39.707***	-39.569***	-37.446***	-38.539***	-38.522***	-37.635***
	(5.087)	(5.168)	(6.887)	(5.105)	(5.153)	(6.726)	(5.871)	(5.721)	(4.387)
Female	-1.425	-0.776	-0.689						
	(4.320)	(4.377)	(4.060)						
Female * Vs top class	1.247	1.088	-1.011						
	(4.680)	(4.698)	(5.095)						
Competition		5.652*	-3.551		4.093	-2.671		8.357	-5.595
		(2.108)	(2.080)		(4.977)	(2.155)		(4.794)	(7.064)
Test 1			4.479***			4.078***			5.306***
			(0.263)			(0.341)			(0.441)
T-PR			2.185***			2.865**			1.816*
			(0.309)			(0.651)			(0.625)
Observations	266	266	266	137	137	137	129	129	129

Analyses are limited to the sample of students in the middle classes. All models provide OLS linear probability results that include class fixed effects. The chance of winning in Test 4 is the chance of getting 1st place in a group of 4 total competitors: the individual and 3 competitors from the other class (percentages are obtained by simulating 1,000 random draws of groups of 3 competitors for each individual, by class). Competition is the competition choice prior to Round 3. Test 1 is the number of correct answers on Test 1 (Piece-Rate). T-PR is the difference between number correct on the tournament (Test 2) versus piece-rate test (Test 1). Robust standard errors are provided in parentheses. Significance levels are set at: + p<0.10 * p<0.05 ** p<0.01 *** p<0.001.

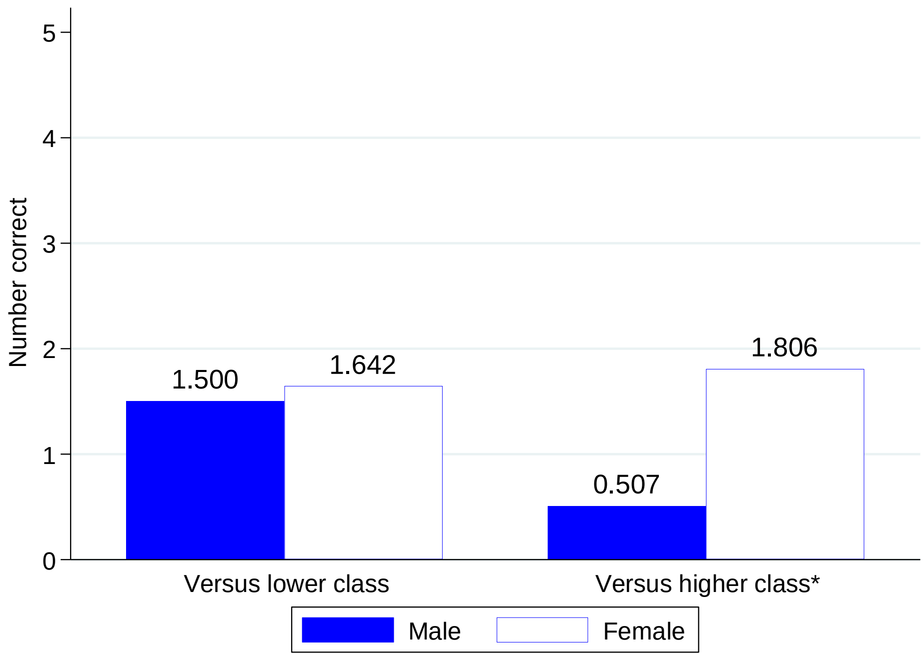
Table 12. Change in expected earnings in Test 4 due to level of competition.

	All			Male			Female		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Versus top class	-12.006***	-11.907***	-10.512**	-11.836***	-11.755***	-10.772**	-12.041***	-12.032***	-11.570***
	(1.573)	(1.632)	(2.225)	(1.560)	(1.584)	(2.115)	(2.153)	(2.118)	(1.393)
Female	-0.181	0.169	0.229						
	(1.913)	(1.921)	(1.785)						
Female * Vs top class	-0.151	-0.236	-1.125						
	(2.175)	(2.161)	(2.185)						
Comp.		3.054*	-1.354		2.424	-0.828		4.260*	-2.409
		(1.080)	(0.965)		(2.362)	(1.165)		(1.357)	(2.857)
Test 1			2.138***			1.959***			2.515***
			(0.215)			(0.241)			(0.320)
T-PR			1.164**			1.430**			1.029*
			(0.270)			(0.407)			(0.340)
Obs.	266	266	266	137	137	137	129	129	129

Analyses are limited to the sample of students in the middle classes. All models provide OLS linear probability results that include class fixed effects. The expected earnings in Test 4 is the chance of getting 1st place in the group of 4 multiplied by 2 (percentages are obtained by simulating 1,000 random draws of groups of 3 competitors for each individual, by class). Competition is the competition choice prior to Round 3. Test 1 is the number of correct answers on Test 1 (Piece-Rate). T-PR is the difference between number correct on the tournament (Test 2) versus piece-rate test (Test 1). Robust standard errors are provided in parentheses. Significance levels are set at: + p<0.10 * p<0.05 ** p<0.01 *** p<0.001.

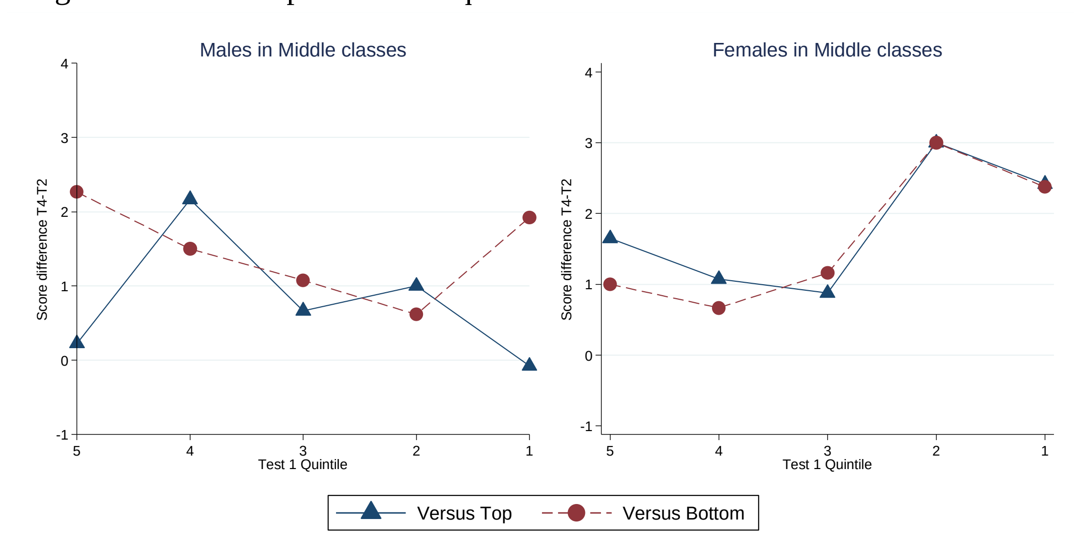
Figures

Figure 1: Change in number of correct answers between Test 2 and Test 4, by treatment and gender



Note: Significance levels are set at: + p<0.10 * p<0.05 ** p<0.01 *** p<0.001.

Figure 2. Change in number of correct answers between Test 2 and Test 4, by treatment and gender and initial performance quintile



Note: Quintiles calculated within each class. 1 is best and 5 is worst.

Appendix

A-1. Average difference in number of correct answers between tests.

	Class level	Overall Diff	Overall p-value	Male Diff	Male p-value	Female Diff	Female p-value
Test 1 to Test 2	All	1.900	<0.001	2.017	<0.001	1.775	<0.001
	Bottom	1.810	0.012	1.882	0.043	1.700	0.142
	Middle	1.872	<0.001	2.175	0.001	1.550	0.003
	Top	2.012	<0.001	1.87	0.037	2.130	0.002
Test 2 to Test 3	All	0.720	0.010	0.638	0.153	0.808	0.026
	Bottom	0.214	0.747	0.105	0.945	0.380	0.734
	Middle	0.880	0.017	0.781	0.204	0.984	0.031
	Top	0.846	0.055	0.909	0.217	0.793	0.148
Test 3 to Test 4	All	0.451	0.203	0.345	0.569	0.565	0.163
	Bottom	0.492	0.657	0.539	0.637	0.420	0.895
	Middle	0.462	0.308	0.204	0.783	0.736	0.192
	Top	0.402	0.483	0.403	0.751	0.402	0.409
Test 2 to Test 4	All	1.171	<0.001	0.983	0.045	1.373	0.001
	Bottom	0.706	0.451	0.645	0.565	0.800	0.599
	Middle	1.342	0.001	0.985	0.114	1.721	0.002
	Top	1.249	0.012	1.312	0.147	1.196	0.029

This table reports the differences in number correct. Number of observations are from the whole sample: 561 overall, with 290 males and 271 females overall. The gender breakdown is: 76 males and 50 females in the bottom classes; 137 males and 129 females in the middle classes; 77 males and 92 females in the top classes. P-values are from Mann-Whitney *U* tests for the difference in number correct between tests.

A-2. Student behavioral characteristics, by gender and class level.

		Overall			Male		Female			
Variable	**Class level**	**Mean**	**N**	**SD**	**Mean**	**N**	**Mean**	**N**	**Diff**	**p-value**
Guess Rank Test 2	Overall	2.573	560	0.941	2.441	290	2.715	270	-0.273	0.001
	Bottom	2.698	126	0.998	2.632	76	2.800	50	-0.168	0.502
	Middle	2.515	266	0.933	2.358	137	2.682	129	-0.325	0.005
	Top	2.571	168	0.906	2.403	77	2.714	91	-0.312	0.030
Guess Rank Test 4	Overall	2.455	560	1.072	2.36	289	2.557	271	-0.197	0.028
	Bottom	3.048	126	0.954	2.934	76	3.220	50	-0.286	0.088
	Middle	2.481	266	1.068	2.365	137	2.605	129	-0.240	0.072
	Top	1.970	168	0.925	1.776	76	2.130	92	-0.354	0.011
Overconfidence Test 2 (Actual-Guessed rank)	Overall	-0.221	560	1.159	-0.097	290	-0.356	270	0.259	0.016
	Bottom	-0.484	126	1.129	-0.539	76	-0.400	50	-0.139	0.627
	Middle	-0.139	266	1.185	0.051	137	-0.341	129	0.392	0.011
	Top	-0.155	168	1.116	0.078	77	-0.352	91	0.430	0.036
Overconfidence Test 4 (Actual-Guessed rank)	Overall	-0.136	560	1.136	0.042	289	-0.325	271	0.366	<0.001
	Bottom	0.310	126	1.196	0.408	76	0.160	50	0.248	0.180
	Middle	-0.244	266	1.128	-0.117	137	-0.380	129	0.263	0.078
	Top	-0.298	168	1.018	-0.039	76	-0.511	92	0.471	0.001
Incentivized risk scale (1-5; 5 most risky)	Overall	2.588	561	1.573	3.103	290	2.037	271	1.067	<0.001
	Bottom	2.317	126	1.505	2.763	76	1.640	50	1.123	0.001
	Middle	2.613	266	1.555	3.066	137	2.132	129	0.934	<0.001
	Top	2.751	169	1.632	3.506	77	2.120	92	1.387	<0.001
Non-incentivized risk scale (0-10; 10 most risky)	Overall	6.161	559	2.150	6.410	288	5.897	271	0.513	0.001
	Bottom	5.427	124	2.292	5.824	74	4.840	50	0.984	0.006
	Middle	6.308	266	2.049	6.489	137	6.116	129	0.373	0.048
	Top	6.467	169	2.087	6.831	77	6.163	92	0.668	0.028

Guess Rank ranges from 1 to 4 (1 is the best rank and 4 is the worst rank). Overconfidence is calculated as Actual-Guessed rank (actual rank based on modal rank in 1,000 simulations). P-values are from Mann-Whitney *U* tests.

A-3. Student midterm scores, by gender and class level.

		Overall			Male		Female			
Variable	**Class level**	**Mean**	**N**	**SD**	**Mean**	**N**	**Mean**	**N**	**Diff**	**p-value**
Math	Overall	49.842	463	23.379	49.457	230	50.223	233	-0.767	0.852
	Bottom	27.064	94	19.407	25.204	54	29.575	40	-4.371	0.363
	Middle	48.043	234	18.009	50.534	116	45.593	118	4.941	0.045
	Top	68.822	135	18.073	69.200	60	68.520	75	0.680	0.629
Malay	Overall	53.620	463	19.600	49.700	230	57.489	233	-7.789	<0.001
	Bottom	31.638	94	18.686	28.833	54	35.425	40	-6.592	0.138
	Middle	53.667	234	14.681	51.009	116	56.280	118	-5.271	0.004
	Top	68.844	135	11.615	65.950	60	71.160	75	-5.210	0.014
English	Overall	63.641	462	17.896	62.000	230	65.267	232	-3.267	0.125
	Bottom	41.462	93	16.731	39.259	54	44.513	39	-5.254	0.188
	Middle	63.889	234	12.823	64.483	116	63.305	118	1.178	0.264
	Top	78.489	135	7.753	77.667	60	79.147	75	-1.480	0.167
Overall	Overall	54.936	432	17.403	52.414	215	57.436	217	-5.022	0.005
	Bottom	32.739	94	12.150	30.973	54	35.123	40	-4.150	0.200
	Middle	54.637	203	11.165	54.089	101	55.179	102	-1.091	0.526
	Top	70.843	135	9.114	68.892	60	72.404	75	-3.512	0.060

Midterm scores are available for 4 schools, and are on a scale of 0 to 100. P-values are from Mann-Whitney *U* tests.

A-4. Student opinions and stereotypes, by gender and class level.

		Overall			Male		Female			
Variable	**Class level**	**Mean**	**N**	**SD**	**Mean**	**N**	**Mean**	**N**	**Diff**	**p-value**
Like Math	Overall	0.721	555	0.449	0.743	284	0.697	271	0.046	0.232
	Bottom	0.585	123	0.495	0.548	73	0.640	50	-0.092	0.311
	Middle	0.736	265	0.442	0.787	136	0.682	129	0.105	0.054
	Top	0.796	167	0.404	0.853	75	0.750	92	0.103	0.100
Like Science	Overall	0.770	556	0.421	0.812	287	0.725	269	0.087	0.015
	Bottom	0.637	124	0.483	0.635	74	0.640	50	-0.005	0.956
	Middle	0.795	264	0.404	0.853	136	0.734	128	0.119	0.017
	Top	0.827	168	0.379	0.909	77	0.758	91	0.151	0.010
Like Reading	Overall	0.752	537	0.432	0.647	275	0.863	262	-0.215	<0.001
	Bottom	0.648	122	0.480	0.514	72	0.840	50	-0.326	<0.001
	Middle	0.789	251	0.409	0.714	126	0.864	125	-0.150	0.004
	Top	0.774	164	0.419	0.662	77	0.874	87	-0.211	0.001
Good at Math	Overall	0.422	552	0.494	0.472	286	0.368	266	0.104	0.014
	Bottom	0.240	121	0.429	0.233	73	0.250	48	-0.017	0.830
	Middle	0.392	263	0.489	0.485	136	0.291	127	0.194	0.001
	Top	0.601	168	0.491	0.675	77	0.538	91	0.137	0.072
Good at Science	Overall	0.410	554	0.492	0.455	286	0.362	268	0.093	0.027
	Bottom	0.295	122	0.458	0.274	73	0.327	49	-0.053	0.534
	Middle	0.407	263	0.492	0.500	136	0.307	127	0.193	0.001
	Top	0.497	169	0.501	0.545	77	0.457	92	0.089	0.251
Good at Reading	Overall	0.722	554	0.448	0.675	286	0.772	268	-0.098	0.010
	Bottom	0.677	124	0.469	0.622	74	0.760	50	-0.138	0.107
	Middle	0.695	262	0.461	0.667	135	0.724	127	-0.058	0.311
	Top	0.798	168	0.403	0.740	77	0.846	91	-0.106	0.090
Rank Science 1	Overall	0.714	532	0.452	0.722	270	0.706	262	0.016	0.681
	Bottom	0.567	104	0.498	0.567	60	0.568	44	-0.002	0.988
	Middle	0.695	262	0.461	0.716	134	0.672	128	0.045	0.435
	Top	0.837	166	0.370	0.855	76	0.822	90	0.033	0.567
Guess Science Stream	Overall	0.442	559	0.497	0.476	288	0.406	271	0.070	0.097
	Bottom	0.208	125	0.408	0.267	75	0.120	50	0.147	0.049
	Middle	0.406	266	0.492	0.453	137	0.357	129	0.096	0.112
	Top	0.673	168	0.471	0.724	76	0.630	92	0.093	0.201

Gender better at math									
	Overall	-0.220	549	-0.270	282	-0.169	267	-0.101	0.053
	Bottom	-0.169	118	-0.157	70	-0.188	48	0.030	0.811
	Middle	-0.229	262	-0.259	135	-0.197	127	-0.062	0.415
	Top	-0.243	169	-0.390	77	-0.120	92	-0.270	0.003
Gender better at reading									
	Overall	0.376	553	0.320	284	0.435	269	-0.115	0.048
	Bottom	0.248	121	0.236	72	0.265	49	-0.029	0.974
	Middle	0.420	264	0.378	135	0.465	129	-0.087	0.289
	Top	0.399	168	0.299	77	0.484	91	-0.185	0.040
Gender better at science									
	Overall	-0.160	550	-0.236	284	-0.079	266	-0.157	0.003
	Bottom	-0.092	120	-0.194	72	0.063	48	-0.257	0.041
	Middle	-0.206	262	-0.237	135	-0.173	127	-0.064	0.433
	Top	-0.137	168	-0.273	77	-0.022	91	-0.251	0.007

Attitudes and beliefs are based on dichotomized variables where 1=yes/agree and 0=no/disagree, except for "Gender better at" questions which are coded -1 (Boys are better) 0 (Both are equally as good) and 1 (Girls are better). P-values are from Mann-Whitney *U* tests.

A-5. Descriptive statistics of student characteristics, by gender and class level.

		Overall			**Male**		**Female**			
Variable	**Class level**	**Mean**	**N**	**SD**	**Mean**	**N**	**Mean**	**N**	**Diff**	**p-value**
Female	Overall	0.483	561	0.500						
	Bottom	0.397	126	0.491						
	Middle	0.485	266	0.501						
	Top	0.544	169	0.500						
Father is college grad	Overall	0.451	552	0.498	0.483	286	0.417	266	0.065	0.124
	Bottom	0.276	123	0.449	0.297	74	0.245	49	0.052	0.526
	Middle	0.462	262	0.499	0.518	137	0.400	125	0.118	0.056
	Top	0.563	167	0.498	0.600	75	0.533	92	0.067	0.384
Mother is college grad	Overall	0.367	551	0.482	0.384	284	0.348	267	0.035	0.388
	Bottom	0.281	121	0.451	0.292	72	0.265	49	0.026	0.752
	Middle	0.341	264	0.475	0.372	137	0.307	127	0.065	0.265
	Top	0.470	166	0.501	0.493	75	0.451	91	0.043	0.584

P-values are based on Mann-Whitney *U* tests.

A-6. Models for tournament entry (Competitiveness), excluding school without administrative records.

	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Female	-0.170**	-0.144*	-0.136*	-0.142*	-0.150*
	(0.048)	(0.048)	(0.047)	(0.048)	(0.056)
Num. Correct Test 1	0.018*	0.026**	0.023**	0.027**	0.028**
	(0.006)	(0.007)	(0.007)	(0.008)	(0.008)
T-PR	0.007	0.018*	0.014+	0.019*	0.019*
	(0.008)	(0.007)	(0.007)	(0.007)	(0.007)
Overconfidence Test 2		0.064**	0.053*	0.058*	0.067*
		(0.020)	(0.021)	(0.020)	(0.022)
Nonincentivized risk			0.038**	0.037**	0.037**
			(0.010)	(0.010)	(0.011)
Incentivized risk			-0.005	-0.010	-0.020*
			(0.012)	(0.013)	(0.009)
Math stereotype				-0.038	-0.036
				(0.043)	(0.045)
Likes math				-0.009	-0.010
				(0.046)	(0.050)
Thinks is good at math				-0.036	-0.018
				(0.051)	(0.058)
Expects science stream				0.069	0.074
				(0.042)	(0.042)
Father is college grad				-0.084	-0.056
				(0.058)	(0.055)
Mother is college grad				0.003	-0.017
				(0.070)	(0.070)
Midterm math score					-0.000
					(0.002)
Midterm overall score					-0.005
					(0.004)
Observations	464	463	462	439	409

All models provide OLS linear probability results that include session fixed effects (13 session vs 18 classes). T-PR is the difference between number correct on the tournament (Test 2) versus piece-rate test (Test 1). Overconfidence Test 2 is measured as the difference between Actual and Guessed rank on Test 2. Nonincentivized risk is a scale from 0 to 10 (10 is most risky). Incentivized risk is the choice between a certain option or set of lotteries, ranging from 1 to 5 (5 is most risky). Robust standard errors are provided in parentheses. Significance levels are set at + p<0.10 * p<0.05 ** p<0.01 *** p<0.001.

A-7. Models for tournament entry (Competitiveness), clustered by session.

	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Female	-0.174**	-0.154**	-0.153**	-0.152**	-0.151*
	(0.047)	(0.047)	(0.042)	(0.041)	(0.058)
Num. Correct Test 1	0.017***	0.022***	0.020***	0.022***	0.028***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)
T-PR	0.008	0.015+	0.012	0.016+	0.018+
	(0.010)	(0.008)	(0.007)	(0.009)	(0.009)
Overconfidence Test 2		0.051**	0.042*	0.046**	0.062**
		(0.013)	(0.017)	(0.014)	(0.018)
Nonincentivized risk			0.025+	0.024+	0.037*
			(0.014)	(0.013)	(0.012)
Incentivized risk			-0.007	-0.007	-0.023*
			(0.012)	(0.014)	(0.008)
Math stereotype				-0.000	-0.026
				(0.025)	(0.029)
Likes math				-0.004	-0.021
				(0.024)	(0.033)
Thinks is good at math				-0.005	-0.003
				(0.057)	(0.058)
Expects science stream				-0.002	0.058
				(0.032)	(0.034)
Father is college grad				-0.041	-0.061
				(0.084)	(0.064)
Mother is college grad				0.015	-0.004
				(0.050)	(0.059)
Midterm math score					0.000
					(0.001)
Midterm overall score					-0.006*
					(0.002)
Observations	561	560	558	524	409

All models provide OLS linear probability results that include session fixed effects (13 session vs 18 classes). T-PR is the difference between number correct on the tournament (Test 2) versus piece-rate test (Test 1). Overconfidence Test 2 is measured as the difference between Actual and Guessed rank on Test 2. Nonincentivized risk is a scale from 0 to 10 (10 is most risky). Incentivized risk is the choice between a certain option or set of lotteries, ranging from 1 to 5 (5 is most risky). Robust standard errors are provided in parentheses. Significance levels are set at + p<0.10 * p<0.05 ** p<0.01 *** p<0.001.

A-8. Balance check of covariates for middle classes.

	All			Males			Females		
Variable	**Coeff**	**SE**	**Obs**	**Coeff**	**SE**	**Obs**	**Coeff**	**SE**	**Obs**
Female	-0.038	0.033	266						
Math midterm score	0.282	2.399	234	-3.071	2.100	116	3.563	2.991	118
Overall midterm score	0.610	1.320	203	-0.103	1.762	101	1.530	1.545	102
Test 1 (Piece-Rate)	-0.626+	0.290	266	-1.058	0.651	137	-0.086	0.570	129
Test 2 (Tournament)	-0.373	0.359	266	-0.372	0.532	137	-0.338	0.394	129
Test 3	-0.513	0.539	266	-0.793	0.518	137	-0.026	0.797	129
Tournament-Piece Rate	0.252	0.337	266	0.686	0.409	137	-0.251	0.477	129
Competition choice	-0.015	0.031	266	-0.034	0.050	137	-0.002	0.054	129

Analyses are limited to the sample of students in the middle classes. This table presents results of regressions of the covariates on treatment, for the overall sample and then by gender. Each row represents a regression. All regressions use class fixed effects. Robust standard errors are provided in parentheses. Significance levels are set at + p<0.10 * p<0.05 ** p<0.01 *** p<0.001

A-9. Change in number of correct answers between Test 2 and Test 4 by treatment in middle classes.

	Overall			Males		Females			
Class Level	**Mean**	**N**	**SD**	**Mean**	**N**	**Mean**	**N**	**Diff**	**p-value**
Versus lower	1.571	133	2.827	1.500	66	1.642	67	-0.142	0.849
Versus higher	1.113	133	2.972	0.507	71	1.806	62	-1.299	0.017

Analyses are limited to the sample of students in the middle classes. Differences are calculated by individual (Test 4-Test 2). P-values are from Mann-Whitney *U* tests.

A-10. Change in number of correct answers between Test 2 and Test 4 due to level of competition, using whole school sample.

$$y_{ij} = \Gamma_j + \beta_1 Vshigher_{ij} + \beta_2 Female_{ij} + \beta_3 (Vshigher * Female)_{ij} + \theta X_{ij} + \epsilon_{ij}$$

y_{ij} = Difference in Number of Correct Answers between Other and Own class (Test 4 - Test 2) for student i in class j.
Γ_j is the class fixed effects.
$Vshigher$ is 1 if assigned higher class and 0 if assigned lower class for student i in class j. This means that for all the bottom classes & half of middle classes, Treatment=1.
$Female$ is 1 if female and 0 if male for student i in class j.
$Female * Vshigher$ is 1 if subject is assigned to higher class & is female; 0 otherwise.
X_{ij} is vector of student attributes.

	All			**Male**			**Female**		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Vs higher	-0.793*	-0.794*	-0.571	-0.996**	-0.991**	-0.814*	0.181	0.180	0.060
	(0.373)	(0.376)	(0.373)	(0.313)	(0.311)	(0.381)	(0.579)	(0.590)	(0.584)
Female	-0.083	-0.088	-0.054						
	(0.328)	(0.337)	(0.363)						
Female * Vs higher	0.672	0.674	0.356						
	(0.438)	(0.441)	(0.481)						
Competition		-0.021	-0.014		0.143	0.120		-0.473	-0.227
		(0.280)	(0.237)		(0.377)	(0.328)		(0.437)	(0.400)
Test 1			-0.032			-0.030			-0.064
			(0.021)			(0.031)			(0.042)
T-PR			-0.374***			-0.305***			-0.456***
			(0.041)			(0.047)			(0.066)
Observations	561	561	561	290	290	290	271	271	271

All models provide OLS linear probability results that include class fixed effects. The whole school sample is used; thus those who received the treatment "Vshigher" are half the students in the middle classes and all the students in the bottom classes, which is not random. Competition is the competition choice prior to Round 3. Test 1 is the number of correct answers in Test 1 (Piece-Rate). T-PR is the difference between number correct on the tournament (Test 2) versus piece-rate test (Test 1). Robust standard errors are provided in parentheses. Significance levels are set as: + p<0.10 * p<0.05 ** p<0.01 *** p<0.001.

A-11. Change in number of correct answers between Test 2 and Test 4 due to level of competition, clustered by session.

	All			Male			Female		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Versus top class	-1.032*	-1.046*	-0.816	-1.013*	-1.025*	-0.896+	0.177	0.175	0.051
	(0.417)	(0.419)	(0.464)	(0.404)	(0.402)	(0.444)	(0.674)	(0.693)	(0.698)
Female	-0.005	-0.056	-0.116						
	(0.544)	(0.550)	(0.612)						
Female * Vs top class	1.189+	1.201	0.866						
	(0.602)	(0.625)	(0.760)						
Competition		-0.452	-0.317		-0.341	-0.223		-0.645	-0.488
		(0.240)	(0.227)		(0.419)	(0.366)		(0.549)	(0.572)
Test 1			-0.043			-0.064			0.000
			(0.030)			(0.035)			(0.040)
T-PR			-0.401***			-0.282**			-0.496***
			(0.044)			(0.053)			(0.082)
Observations	266	266	266	137	137	137	129	129	129

Analyses are limited to the sample of students in the middle classes. All models provide OLS linear probability results that include session fixed effects (7 sessions vs 8 classes). Competition is the competition choice prior to Round 3. Test 1 is the number of correct answers on Test 1 (Piece-Rate). T-PR is the difference between number correct on the tournament (Test 2) versus piece-rate test (Test 1). Robust standard errors are provided in parentheses. Significance levels are set at: + p<0.10 * p<0.05 ** p<0.01 *** p<0.001.

A-12. Number of correct answers on Test 4 due to level of competition.

	All			Male			Female		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Versus top class	-1.438*	-1.384+	-0.815+	-1.368+	-1.325+	-0.893+	-0.157	-0.153	0.059
	(0.596)	(0.616)	(0.421)	(0.590)	(0.605)	(0.392)	(0.715)	(0.694)	(0.626)
Female	-0.343	-0.151	-0.112						
	(0.738)	(0.763)	(0.615)						
Female * Vs top class	1.234	1.187	0.864						
	(0.826)	(0.750)	(0.756)						
Competition		1.671*	-0.305		1.293	-0.237		2.132**	-0.598
		(0.580)	(0.244)		(0.975)	(0.362)		(0.459)	(0.583)
Test 2			0.953***			0.920***			1.016***
			(0.035)			(0.037)			(0.049)
T-PR			-0.356***			-0.205*			-0.502***
			(0.043)			(0.065)			(0.091)
Observations	266	266	266	137	137	137	129	129	129

Analyses are limited to the sample of students in the middle classes. All models provide OLS linear probability results that include class fixed effects. Competition is the competition choice prior to Round 3. Test 2 is the number of correct answers in Test 2 (Tournament). T-PR is the difference between number correct on the tournament (Test 2) versus piece-rate test (Test 1). Robust standard errors are provided in parentheses. Significance levels are set at: + p<0.10 * p<0.05 ** p<0.01 *** p<0.001.

A-13. Change in number of correct answers between Test 2 and Test 4 due to level of competition, controlling for chance of winning.

	All			**Male**			**Female**		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Versus top class	-1.029*	-1.043*	-0.971*	-0.996*	-1.008*	-0.993*	0.181	0.179	0.100
	(0.342)	(0.348)	(0.393)	(0.326)	(0.325)	(0.364)	(0.603)	(0.622)	(0.680)
Female	-0.002	-0.053	-0.049						
	(0.564)	(0.576)	(0.595)						
Female * Vs top class	1.184+	1.197+	1.055						
	(0.596)	(0.618)	(0.780)						
Competition		-0.448	-0.279		-0.380	-0.168		-0.722	-0.808
		(0.257)	(0.280)		(0.417)	(0.426)		(0.563)	(0.516)
Chance win T1			0.000			-0.008			0.010
			(0.007)			(0.007)			(0.011)
Chance T2-T1			-0.032**			-0.019			-0.037*
			(0.009)			(0.012)			(0.013)
Observations	266	266	266	137	137	137	129	129	129

Analyses are limited to the sample of students in the middle classes. All models provide OLS linear probability results that include class fixed effects. Competition is the competition choice prior to Round 3. Chance winning T1 is the chance of getting 1st place if Test 1 were a tournament with groups of 4 competitors (percentages obtained by simulating 1,000 draws of groups of 3 competitors for each individual, by class). Chance T2-T1 is the difference in the chances of winning in Test 1 and Test 2. Robust standard errors are provided in parentheses. Significance levels are set at: + p<0.10 * p<0.05 ** p<0.01 *** p<0.001.

Paper 3: The (unintended) effects of allowing Computer Science to count as a mathematics graduation requirement in Texas

1. Introduction

There have been recent efforts to expand Computer Science (CS) K-12 education, although data on the quantity and quality of CS offerings in U.S. K-12 schools is thin. Nevertheless, many states have adopted policies in support of expanding CS education such as allowing CS to count towards an academic graduation requirement (Code.org). As CS becomes an increasingly important part of K-12 education, it is crucial to examine policies pertinent to CS education; however, current research on such policies is thin. To address this gap in the literature, this paper analyzes the effects of allowing CS to count as a core academic graduation credit by race and gender, providing one of the first causal analyses of any K-12 CS education policy. This paper uses data from Texas, one of the earliest adopters of K-12 CS policies.

With the growing prominence of CS, it is also important to address equity issues in CS. There are well-known gender and race gaps in Science, Technology, Engineering & Mathematics (STEM), and some indication that these gaps may be particularly severe in CS. For example, CS has one of the lowest proportion of women among STEM degrees (National Science Foundation, 2015), and Blacks and Hispanics have relatively low representation in computer occupations compared to in other STEM occupations (Landivar, 2013). Thus, this paper analyzes student race-gender subgroups separately to examine whether the policy has differential effects by race and gender.

This paper describes the effects of a recent and growing policy intended to expand CS education in high school by allowing CS to count towards a mathematics graduation requirement. The background literature is provided in Section 2. The Texas policy context is described in Section 3. Section 4 describes the data and empirical methods.

Section 5 provides the main and spillover results. Section 6 includes a discussion of the results and concludes.

2. Literature

This paper contributes to two main bodies of literature, K-12 CS education and high school graduation requirements. These two areas will increasingly overlap as CS education gains more traction within K-12 education.

K-12 CS education background

Until recently, CS has not been considered a core academic subject and has been inconsistently categorized across states and districts; for example, it is often part of Careers and Technology Education (Wilson et al., 2010). The K-12 Computer Science Framework was recently created to provide guidance on what K-12 CS education should look like, distinguishing between CS and other computer-related skills, such as computer literacy or information technology. This paper focuses on AP Computer Science, which is the only course that counts towards the mathematics graduation requirement in Texas (see Section 3).

Along with the emerging standardization of the definition of K-12 CS education, the role of CS in K-12 education appears to be changing in significant ways. Although there have been calls to expand CS education over the past several decades, the growing ubiquity of computing supports "broad acknowledgment" of the importance of computing (Grover & Pea, 2012; 40). In a seminal article, Wing (2006) argues that the skills associated with CS are a "fundamental skill for everyone" (33). This growing consensus has now reached a national level of recognition, as CS is included under the definition of a well-rounded education in the Every Student Succeeds Act (2015).

Evidence suggests that gender and racial disparities in CS attitudes begin early and endure. A recent study finds that 6-year old children believed gender stereotypes about programming (i.e. boys are better at programming than girls), yet not about math or science, which is in line with other research that suggests that math and science gender stereotypes develop later in childhood (Master et al., 2017). Thus, it appears that gender differences in CS stereotypes may develop even earlier than other STEM stereotypes. In higher education, there is suggestive evidence that females were less confident than males in an advanced CS class at Stanford (Irani, 2004). There is also evidence of racial differences in math and science attitudes (see Else-Quest et al., 2013 for review).

In addition to differences in attitudes about CS, CS attainment varies by gender and race. For example, 24% of male versus 14% of female high school graduates had a CS credit in 2009 (U.S. Department of Education, 2011). Similarly, only 19% of AP CS test-takers were female in 2013 (Cheryan et al., 2017). In several states, there were no female, or no Black/Hispanic students who took the AP CS test in 2013[1] (Ericson, 2014). In higher education, gender and racial differences exist among students in Computer/Information Science majors (Chen & Weko, 2009; NSF, 2017). Gender differences in CS degrees hold in all racial groups, with the lowest proportion of females among Whites (Cheryan et al., 2017).

While there is evidence of gaps in CS attainment, it is difficult to calculate the extent of these gaps because of the lack of state and national data on CS course-taking. Thus, it is unclear what the enrollment in CS is overall, moreover by student subgroups. However, there is some evidence of “a marked decline in the number of introductory and

[1] I conduct original analyses that corroborate this and show historical trends in more detail.

AP Computer Science courses being taught in secondary schools" as of 2010, even though state and national policies have sought to increase STEM course offerings (Wilson et al., 2010). Nationally, the percentage of high school graduates who took a CS course declined from 25% in 1990 to 19% in 2009 (U.S. Department of Education, 2011).

One major barrier to CS course-taking appears to be access to these classes. Only 2,100 of the 42,000 high schools in the US were certified to teach the AP computer science course in 2011 (Exploring computer science, 2017). However, there appears to be a more positive recent trend: 40% of K-12 principals stated that a CS course is offered in their schools in 2015-16, up from 25% in 2014-15. The percentage was higher in secondary school; almost four-fifths of high school principals (78%) reported that their schools offered at least one CS class in 2016 (Google Inc. & Gallup Inc., 2016b).

Access to CS courses can vary by factors such as race and socio-economic status. Less than half (47%) of Black students compared to 58% of White 7th-12th grade students state their school has a dedicated CS class (Google Inc. & Gallup Inc., 2016). In addition, there is qualitative evidence that AP CS classes are unequally distributed by race/SES in Los Angeles Unified School District (Margolis et al., 2003). These data indicate that CS is typically offered only in certain schools or regions, which could be one of the main barriers for these underrepresented groups. Thus, increasing access CS courses could potentially help mitigate these gaps.

Existing research suggests that females and other minority students may face differential barriers to taking CS courses, beyond physical access to CS courses. A meta-study finds that males display greater sex-role stereotyping of computers, higher computer self-efficacy, and more positive affect about computers than females do; these gender differences are largest among high school students (Whitley Jr., 1997). Furthermore, stereotypes about CS may discourage females or minority students from participating in CS (Goode, 2008). However, it appears that Black students are more confident than White or Latino students about CS despite lower rates of access (Google Inc. & Gallup Inc., 2016a). Thus, the direction of racial stereotypes about CS is unclear. Nevertheless, research suggests that access to CS courses may not necessarily remove all barriers to taking CS courses.

K-12 CS policies and programs

To my knowledge, there is no extant research on the effects of CS policies in the U.S.. Studies have focused on the creation of CS curriculum rather than the effects of implementation, which makes sense given the relatively recent initiatives towards providing CS at scale in public school classrooms. Internationally, there are several studies that have looked at school CS programs in Israel, New Zealand and Germany (Hazzan, Gal-Ezer, & Blum, 2008; Bell, Andreae, & Lambert, 2010; Hubwieser, 2012). However, these studies provide only descriptions of either the implementation or planned implementation of the CS programs, with no empirical evidence provided. The relatively sparse and descriptive literature about these CS programs demonstrates the nascent nature of these types of programs, which underscores the importance of providing rigorous analyses to help guide the development of effective CS programs.

An increasing body of literature explores how particular components of CS K-12 education, such as cultural relevance or mode of instruction, may be beneficial (or at least, not harmful) for underrepresented groups (Goode, 2008). For example, digital game-based learning of CS was more effective for students' computer memory concepts and student motivation, with no differences by gender (Papastergiou, 2010). There is suggestive evidence that a high-intensity CS program targeting females may increase their understanding of what computer scientists do and the significance of CS, yet there were no significant differences in personal beliefs about CS (Gannod et al., 2014). These types of contextual concerns have contributed to the development of other types of K-12 CS courses (e.g. AP CS Principles) which may be a promising path towards expanding K-12 CS access.

The existing literature demonstrates that CS programs are of increasing interest, both domestically and globally. Furthermore, these programs may affect student learning and attitudes surrounding CS, and programs may incorporate these factors to help create a better learning environment for underrepresented students. However, the existing literature is still relatively sparse.

High school graduation requirements

Another body of relevant literature for understanding the effects of allowing CS to count as a graduation requirement is the existing evidence on high school graduation requirements. Although there is a robust body of literature on the effects of changing various aspects of graduation requirements, the role of electives in high school graduation is less-explored in the literature. Thus, this paper contributes to the high school graduation literature by highlighting the effect of a change in the elective status of AP CS.

Underlying policies to increase graduation requirements is the assumption that greater course-taking is associated with improved student outcomes. Several studies show that there are positive associations between more advanced course-taking and high school achievement and graduation (Long, Conger & Iatarola, 2012) in addition to college outcomes (Attewell & Domina, 2008; Aughinbaugh, 2012). There is also evidence of inequalities of access to advanced courses (Attewell & Domina, 2008) and differences in the effects of taking more advanced courses by student group (Long et al., 2012).

Teitelbaum directly examines a widespread policy requiring three mathematics and science credits in the late 1980s and 1990s, which was intended to increase student fluency in these subjects (2003). Teitelbaum finds that the policy was associated with increased course-taking but not student achievement; only a small percentage of students took an advanced course to fulfill the third requirement. Thus, this policy also did not meet all of its intended goals. There were no analyses by student subpopulation reported.

Other studies that have examined high school graduation policy show consonant results. For example, Schiller & Muller (2003) look at overall graduation unit requirements and find that higher requirements are associated with higher level math coursework as freshmen, but fewer advanced mathematics credits overall, which is consistent with Teitelbaum's findings. There is some evidence of differences by race and socioeconomic status (e.g. differential expected amounts of coursework by race in states with more or less extensive testing). Another study shows that increased graduation requirements in math and science were associated with increased course-taking, but only among students who were taking minimum courses (Chaney, Burgdorf & Atash, 1997). The majority of students took above the minimum course requirement and the policy was not associated with increased course-taking. A low percentage of students actually took an advanced level of math/science in relationship to the increased graduation requirements, and there was increased student achievement among those who took the relatively advanced level courses.

The literature suggests that there is not a simple relationship between graduation policies and student outcomes. Only certain groups of students may increase course-taking, or students may shift towards increasing units but not in advanced courses. These considerations could also be applied to the policy studied in the current paper.

3. Texas policy theory and context

Texas policy theory

There have been recent policy attempts to expand access to K-12 CS education, including integrating CS into core academic requirements. The assumption underlying the policy is that the elective status of CS acts as a barrier to students taking the course. Allowing CS to count as an academic elective would thus remove a potential barrier (K–12 Computer Science Framework). The number of states that adopted this policy increased from 12 in 2013 to 35 in 2018 (Code.org). Despite the rapid adoption of this policy, the effect on student participation in CS courses has not been studied.

According to the CS Framework, the elective status of CS acts as a barrier to students taking CS by perpetuating existing “misconceptions” about CS and limits exposure to CS, since students may self-select out more when CS is only an elective credit. A policy that grants mathematics credit for taking CS class ostensibly lowers the barrier for taking CS. There could be several reasons for this, such as allowing students to take an equally attractive elective in addition to CS, signaling the importance of CS, or simply highlighting the option of taking CS. Although this policy may increase enrollment in CS classes, it is not obvious whether this policy improves or aggravates existing racial and gender disparities in CS course-taking. If this policy lowers the barrier to taking CS, does it lower it for all students or more for certain groups? Furthermore, does the policy result in lowered enrollment for mathematics courses? There is some concern that this policy may detract from student participation in advanced math courses (Loewus, 2016).

Data suggests that females who have exposure to CS classes in high school are much more likely to take CS classes in college; thus, providing access could be a critical component to increasing participation rates of under-represented groups (Google, 2014). A blog post from the Association for Computing Machinery (ACM) explains, “Our first step is to provide access. If computer science counts towards high school graduation, then schools have a reason to offer it” (Guzdial, 2014).

Although access may be provided, the demographics of those who choose into CS may not necessarily change. A challenge is to incentivize students, especially underrepresented students, to choose to take CS courses. Allowing CS to count towards a core academic requirement for graduation may increase enrollment in CS classes, yet may differentially increase enrollment for different races and genders. It is important to understand how these policies affect underrepresented student populations, so that there are no negative unintended consequences (e.g. decline in enrollment, or greater disparities) as a result of the policies.

Texas policy context

The State Board of Education sets graduation requirements in Texas, thus the policy affects all districts. Starting in 2007-08, one CS course, AP Computer Science A (AP CS), was allowed to count as a 4th graduation requirement under the Recommended High School Program. The policy was rolled out for incremental graduation cohorts (i.e.

for entering 9th grade cohorts in 2007-08, then entering 9th grade cohorts in 2008-09, etc.)[2].

There were several other concurrent policy changes around STEM graduation requirements during this period. Most importantly, the overall math requirement increased from 3 to 4 units of math, and the science requirement also increased from 3 to 4 units of science in 2007-08. Graduation requirements changed more dramatically starting in 2014-15 with the transition to the Foundation High School Program, which completely replaced the old graduation plans with one program, including endorsements and distinguished levels of achievement. To isolate the effects of the CS policy, I limit the years of analysis to prior to 2014-15 and even more specifically between 2005-06 to 2009-2010, such that there are affected and unaffected cohorts present in the analyses.

4. Data & estimation strategy

The data for this study come from two main sources: a custom dataset from the Texas Education Agency (TEA) and publicly available data from the Common Core Dataset (CCD). The TEA dataset consists of enrollment data by entering cohort (defined as year entered 9th grade), grade, race/gender, course and school (e.g. cohort of 2005 9th grader Latina females in AP Computer Science in School A). These enrollment numbers were used as the numerator for the percentage of each corresponding group of students. Data spanned from 2002-03 to 2015-16 data, yet complete data began in 2005-06.

The data from CCD consists of both grade-level and school-level characteristics. The enrollment by race/gender by grade by school (parallel to the TEA data described earlier) was used as the denominator to construct the percentage of each race/gender

[2] See http://tea.texas.gov/graduation.aspx for implementation details.

group of students. Although these grade-level enrollment numbers did not always match the cohort-level data provided by TEA[3], I used enrollment numbers from 9th grade (which should correspond to entering 9th grade cohort, as defined by TEA) as the denominator for the percentage of each corresponding group of students (i.e. total number of cohort of 2005 9th grader Latina females in School A). Thus, the percentage of students in each course by cohort, grade, race/gender, subject and school were created using both the TEA and CCD datasets.

TEA data included censoring when enrollment for a particular group was between 1 and 4 students. Due to high levels of censoring in the TEA data, I constructed lower (1) and upper (4) bounds in addition to imputing random numbers from 1 to 4 when data is censored. I bounded the percentage data so that the highest percentage possible is 100%. For robustness checks, I also included data when CCD enrollment data was missing and TEA data was positive (forcing the percentage to be 100%) as additional outcome variables in the analyses.

I use a triple difference estimation strategy in my main analyses (see Appendix A-1 for visual explanation). I exploit the cascading nature of the policy implementation to create groups of treated and untreated grades within the same school as the first difference. Although the policy was implemented in 2007-08, it was binding only for the entering 9th grade cohort in 2007-08, thus higher grades were not subject to the policy. Thus, in 2007-08, 10th-12th graders were not subject to the policy; in 2008-09, 11th–12th graders were not subject to the policy; and in 2009-10, 12th graders were not subject to the policy. These grades serve as the control group to treated grades in their treated years.

[3] TEA provided similar data that I constructed from the CCD, but with missing data and suppressed data. Thus, I relied on CCD data to construct these percentages.

Thus, the difference-in-difference model consists of treated grades in treated years, which depends on entering cohort. The triple difference consists of constructing and comparing the difference-in-difference for the treated course, AP CS, in addition to an untreated course, AP Psychology. AP Psychology is an ideal control course since it is a social science, and not subject to any of the changing requirements for both math and science courses during this time period.

In a robustness exercise, I use the triple difference method on a national AP dataset. I use publicly available data from the College Board. There are several limitations of this dataset. The dataset only consists of AP test-takers, not those who enrolled in the AP course, which includes an additionally self-selected sample of students. Although official percentages do not exist, I calculate that about two-thirds of White/Asian students who take the course take the AP CS:A test while about two-fifths (less than half) of Black/Latinx students who take the course take the AP CS:A test, using the randomly-chosen censored values[4]. While these percentages may not be completely accurate due to discrepancies between datasets, the low percentages demonstrate the challenge of using AP test data in lieu of AP course-taking data. Another difference is that the unit of observation (state), does not correspond directly with student enrollment data from the CCD. Instead, the data at the state level are aggregated by public school AP test takers and only by race, or by all AP test takers by race and gender (which include non-public school students); both of these aggregations include all grades. When using CCD data to create percentages by race or race-gender group, the CCD data uses just 9-12th grades for public schools.

[4] Author calculated percentages, which are available upon request.

For the spillover analyses, the difference-in-difference is constructed the same as in the main analyses (i.e. treated grades in treated years, or using differences pre and post policy between different student cohorts). However, for the triple difference, I use the difference between treated and untreated schools, defining treated schools as those that offered AP Computer Science in addition to the specified math class and untreated schools as those that did not offer AP Computer Science while offering the specified math class. Offering a class is defined as having at least one student take the class during the year.

5. Results

This section provides descriptive analyses of the availability of and enrollment in AP CS courses. It then provides causal analyses of the effects of the graduation policy on AP CS course enrollment and select math course enrollment (spillover effects), including falsification and robustness exercises.

To understand the enrollment trends in AP CS, it is important to understand the Texas high school context, focusing on the period between 2005-06 to 2009-10 (Table 1). Texas high schools are increasingly rural, making up almost half of high schools in 2009-10. About 16% of high schools were located in large cities during this time. There was also a large and increasing percentage of schools with a majority of students who qualified for free/reduced priced lunch, from about a third in 2005-06 to 41% by 2009-10. Racial demographics have shifted, with Whites making up half of the high school population in 2005-06 to about 47% in 2009-10, and Latinx students increasing from 35% to 39% (Table 2).

The variation in the types of schools makes salient the question of whether access to AP CS is uniformly distributed across high schools in the state. Overall, less than a fifth of schools offer AP CS, hovering around 17% from 2005 to 2010 (Table 3). In comparison, about 17% of schools offered AP Psychology in 2005 but that percentage climbed to almost a quarter (24%) in 2009-10. Advanced math courses studied in this paper were offered at higher rates, from about a quarter of schools (AP Calculus BC), a third of schools (AP Statistics), three-quarters of schools (AP Calculus BC) and nearly all schools (Pre-calculus). Besides Pre-calculus, there is significant variation by school location, with rural schools (Figure 1), Title I eligible schools and schools with a majority of students with free/reduced lunch status less likely to offer these courses, although there are less pronounced differences by the racial composition of schools[5].

As mentioned in the Section 4, the participation rates in each course were calculated in several ways to deal with the suppressed data, including using lower-bound, upper-bound and random number imputations. The overall participation rates (i.e. throughout the high school career) varied by race-gender group; Asian males had the highest participation rates, with lower bound estimates above 10% and upper bound estimates hovering around 30%, while Black and Latina females have the lowest participation rates at about 5% even at the upper bounds (Table 4, Appendix A-2). Asian female participation rates ranged from about 6-8% at the lower bound to 15-20% at the upper bound, Latino male participation rates ranged from about 4% (lower bound) to 10% (upper bound) and Black male participation rates ranged from about 2% (lower bound) to 8% (upper bound). White males' participation rates were slightly less than 10% at the lower bound to 15-17% at the upper bound, while White females' participation rates ranged from about 4% at the lower bound to 7-10%. The rest of the analyses focus

[5] Figures available upon request.

primarily on White, Latinx and Black students since Asian student enrollment is low, which results in comparatively noisy estimated percentage enrollments.

Female participation rates are higher than male participation rates within race for AP Psychology. There are generally higher rates in the math subjects (AP Calculus BC has similar rates) than in AP CS.

Difference-in-difference results

The results of the difference-in-difference model are shown below. All analyses are conducted separately conducted by race-gender group. In the difference-in-difference model, the comparison of interest is the enrollment rates in AP CS of treated grades, which change over time, compared to enrollment rates in AP CS of untreated grades.

$$y_{igy} = \alpha + \beta(\mathrm{Treatment})_{\mathrm{igy}} + \gamma_g + \delta_y + \rho_i + \epsilon_{igy}$$

y_{igy} Percent enrollment for grade g in year y in school i

$(\mathrm{Treatment})_{\mathrm{igy}}$ is 1 in treated "grade" g (1/2/3) in post-period y (depends on "grade" and year; cohort year >=2007) for school i; 0 otherwise

γ_g grade fixed effect

δ_y year fixed effect

ρ_i school fixed effect

The difference-in-difference results that compare the difference between the treated grades compared to the control grades in treated/untreated years are shown in Table 5. The grade fixed effects account for fixed differences in course enrollment rates across grades and the year fixed effects account for fixed differences in course enrollment rates across years (i.e. common shocks by year). School fixed effects are included to account for fixed differences in course enrollment rates across schools. The results include negative estimates for White males and some negative estimates for White females, which could indicate a negative effect of the policy on AP CS course enrollment for White students. The difference-in-difference model includes fixed effects for year and for grade in addition to school fixed effects, so there are still threats to internal validity if there are time-varying grade trends.

I run a falsification exercise to examine the internal validity of this model, where I look at "effects" by year; a credible result would be if there are no pre-treatment effects. However, this does not appear to be the case for Whites (Appendix A-3), since there appear to be non-positive effects prior to the centered treatment year. To address these possible validity threats, I use a triple difference model that incorporates an untreated subject (AP Psychology).

Triple Difference results

A difference-in-difference is susceptible to several threats to internal validity, such as time-varying grade trends. As described earlier, there is some evidence that there may be trends that precede the policy period, and thus this paper uses a control subject, AP Psychology, that should not be affected by any policy changes during this time. AP Psychology, as a social science, was not subject to either the math or science graduation requirement changes and thus enrollment in this course should not have been affected by these policies.

The triple difference consists of estimating the differences between the treated and untreated grades before and after the policy as defined in the difference-in-difference, in the treated subject (AP Computer Science) compared to an untreated subject (AP

Psychology). The triple difference model uses a fully saturated model and includes school fixed effects.

$$y_{igys} = \alpha + \beta(\text{Treatment})_{igys} + \gamma_{gy} + \delta_{gs} + \mu_{sy} + \rho_i + \epsilon_{ist}$$

y_{igys} Percent enrollment for school i for year in school y in year t in subject s

$(\text{Treatment})_{\text{igys}}$ is 1 for treated "grade" g (1/2/3) in post-period t (cohort year >=2007) in treated subject s in school i; 0 otherwise

ρ_i are the school fixed effects

γ_{gy} grade g by year y fixed effect

δ_{gs} grade g by subject s fixed effect

μ_{sy} subject s by year y fixed effect

As in the difference-in-difference model, school fixed effects are used to account for fixed differences in course enrollment rates across schools. I then include all two-way fixed effects: γ_{gy} refers to grade by year fixed effects, to account for time-varying fixed differences in enrollment across grades; δ_{gs} refers to grade by subject fixed effects, to account for subject-varying fixed differences in enrollment across grades; and μ_{sy} refers to subject by year fixed effects, to account for time-varying fixed differences in enrollment across subjects. These two-way fixed effects preclude the need to include the individual fixed effects. The coefficient β is the effect of the policy, identifying the effect of the CS policy on CS enrollment. The assumptions underlying a causal interpretation of this model is that there are no other plausible explanations for the differences between course enrollments in AP CS and AP Psychology after the policy is binding for the respective grades. For example, the estimate may be susceptible to validity concerns if there are differential trends in the perception of AP CS and AP Psychology by grade over time.

The triple difference results are shown in Table 6. The results indicate a consistent negative effect of about 3 percentage points for White males, and similar (yet not always significant) results for White females. For other race-gender groups, results are not significantly different from zero yet are somewhat negative for Latinx and Black students and slightly positive for Asian students[6].

The analyses are then performed by school characteristics, including rural/non-rural status, majority White, majority Latinx, and majority/non-majority free/reduced price lunch schools[7]. The negative effects persist for White males and females in rural or non-rural schools, White-majority, Latinx-majority schools, FRL-majority or FRL-minority schools, although the effects are not always significant. On the other hand, there are nonsignificant but consistently positive effects for Black and Latina females in Latinx-majority schools and for Latina females in White-majority schools and Black females in FRL majority schools. Thus, it appears that institutional factors, such as racial/socioeconomic composition of schools, could play a role in the effects of this policy.

I conduct several robustness exercises for the triple difference. First, I use the year prior to implementation for each grade and find generally null results (Appendix A-4). Next, I use AP Macroeconomics instead of AP CS as the treated subject and find null results as well (Appendix A-5).

---

[6] The results are qualitatively similar for White males (varies between negative and positive for White females; and although not statistically significant but positive for Latinx and Black males, Asian females) when balanced by outcome variable, although the number of observations is reduced by over 75% and thus the whole sample analyses are reported in the rest of the paper.
[7] Analyses available upon request.

In another robustness exercise, I use a national dataset of AP test-taking to compare AP CS to AP Psychology test enrollment. I leverage the policy used across multiple states using the triple difference method and find null results for all race-gender groups (Appendix A-6). The data used in these analyses are much coarser than the Texas analyses, yet the null results help corroborate the main results. In conjunction with the null results using a prior year or another social science subject, these robustness checks lend credence to the significant effects found in the triple difference analyses.

Spillover results

In addition to the effects of the policy on AP CS enrollment, it is important to explore whether there are impacts on the math courses that AP CS would in essence replace. I focus on four math courses that count as the fourth math graduation requirement: Pre-calculus, AP Calculus AB, AP Calculus BC and AP Statistics. Other eligible math courses did not have sufficient observations to conduct the spillover analyses.

In the following analyses, I test whether there are any spillover effects on these math courses. The triple difference for the math spillover analyses were constructed by using the same difference-in-difference, treated vs untreated grades before and after the policy. However, the triple difference is constructed using treated and untreated schools; that is, schools that offered both the math subject and AP CS and those that offered the math subject but not AP CS. I use a fully saturated model, including grade by year, grade by school and school by year fixed effects.

$$y_{gts} = \alpha + \beta(\text{Treatment})_{gts} + \gamma_{gy} + \delta_{gs} + \mu_{sy} + \epsilon_{gts}$$

y_{yts} Percent enrollment for grade g in year y in school s

$(\text{Treatment})_{\text{igys}}$ is 1 for treated "grade" g (1/2/3) in post-period t (cohort year >=2007) in treated school s in school i; 0 otherwise

γ_{gy} grade g by year y fixed effect

δ_{gs} grade g by school s fixed effect

μ_{sy} school s by year y fixed effect

Due to the model, the number of observations is dramatically reduced from the main analyses. Thus, the following analyses are suggestive, but are sensitive to model specification.

Surprisingly, there were positive effects for Latina and Black females in Pre-calculus, ranging from about 2 percentage point for Latina females to 3 percentage points for Black females (Table 7A). There were no significant spillover effects for other race-gender groups. The results were qualitatively similar by rural/non-rural school status, White or Latinx majority school status, or FRL-majority/minority school status, although the effects not always statistically significant[8]. This suggests that the policy which was intended to increase CS enrollment actually attracted Latina and Black female students to take Pre-calculus, which will be further discussed in Section 6.

However, there were negative significant effects for Latino males on AP Calculus AB, with no other significant effects for other race-gender groups (Table 7B). These results are somewhat consistent across the different types of schools (e.g. rural, non-rural, etc.). There were no significant results for any race gender group for AP Calculus BC

[8] Analyses available upon request.

(Table 7C). Lastly, there were negative effects for White males and the upper bound for Latino students and lower bound/random imputation for Latina students in AP Statistics, from about 3 to 4 percentage points lower (Table 7D). Thus, it appears that this policy may have had negative spillover effects on AP Statistics and AP Calculus AB for certain race-gender groups.

As a robustness exercise, this analysis was also conducted using AP Psychology as a potential spillover subject instead of AP CS (Appendix A-7). There were no significant results except for the upper-bound estimates for Latino males and for Asian males. When prior year implementation (2006) was used, there were no significant effects for any math subjects except for Asians and one estimate for White females in pre-calculus (Appendix A-8). These generally null robustness checks help support the significant findings of the main spillover results.

6. Discussion & Conclusion

The main analysis indicates that the policy of allowing CS to count towards a math graduation requirement does not have the intended consequence of increasing enrollment in AP CS. In fact, the enrollment for White males and females decreases after the policy, although there is some fluctuation depending on the type of school. These results pose a puzzling question: why doesn't enrollment increase in a subject after a policy that grants AP CS more than elective status?

One potential explanation is that the policy elevated the perceived difficulty of AP CS as a potential math course. This explanation is supported by evidence from the spillover analyses that Latina and Black females take more Pre-calculus after the policy. These students appear to have chosen an "easier" course, pre-calculus, when given the option that includes AP CS.

Another possible explanation could be that the policy was not widely announced to or understood by students, which could explain null results. This policy was in the context of a set of many significant graduation policy changes around STEM, and may not have been made as public as these broader policies. However, the existence of several negative effects on CS enrollment and corresponding positive and negative effects on math enrollment suggest that there was some level of awareness of AP CS as an option.

These findings highlight that it is unclear whom the intended targets of this policy were. Which types of students would be induced by this policy to take AP CS? Such students are those who did not take AP CS when it only counted as an elective but would take AP CS when it could count towards a fourth math requirement, perhaps because of limited space in their schedules for electives. This suggests a high-achieving student who is taking an already-high course load, which probably describes only a minority of students and which may not be the intended targeted demographic for inducing more students to take CS. Additionally, it appears that students who have been traditionally underrepresented in CS, Latina and Black females, may react by taking the easier course that fulfills the requirement, which may further contribute to existing inequalities.

The results of this study suggest that this widespread policy to encourage CS course-taking is not effective in increasing student participation. There are generally null results for most race-gender groups, and negative results for White students. These results could be explained by the narrow set of students who might be motivated by this policy in addition to potentially highlighting the relative difficulty of AP CS in comparison to other math courses.

Although Texas has played a pioneering role in CS K-12 education policies, future research could examine whether these results hold across other states that implemented this policy. These additional analyses would help determine whether these null and negative results are specific to Texas. Furthermore, since the majority of states have adopted this policy in the recent years, future analyses could also examine heterogeneity by region or other characteristics.

In addition to this policy, there are several promising alternatives to expand CS participation, including offering CS in all high schools and creating CS courses that are attractive to underrepresented groups. Texas, for example, was the first state to enact a policy that requires all districts to offer CS courses (policy created in 2014). Arkansas followed with a similar policy in 2015, and included significant funding to ensure that students had access to CS courses in all public high schools (Code.org). Alongside the new CS policies, a new AP CS course (AP CS Principles) was developed to attract broader participation in CS and was first administered in 2017. These efforts appear to be targeted to increase participation from a diverse group of students, but it is important to analyze explicitly the impact on subpopulations of students in order to assess the effectiveness of these policies and courses.

References

Attewell, P. & Domina, T. (2008). Raising the bar: Curricular intensity and academic performance, *Educational Evaluation and Policy Analysis*, 30(1), 51-71.

Bell, T., Andreae, P., & Lambert, L. (2010). Computer science in New Zealand high schools. In *Proceedings of the Twelfth Australasian Conference on Computing Education - Volume 103* (pp. 15–22). Darlinghurst, Australia, Australia: Australian Computer Society, Inc. Retrieved from http://dl.acm.org/citation.cfm?id=1862219.1862223

Chaney, B., Burgdorf, K. & Atash, N. (1997). Influencing achievement through high school graduation requirements, *Educational Evaluation and Policy Analysis*, 19(3), 229-244.

Chen, X., & Weko, T. (2009). Students who study science, technology, engineering, and mathematics (STEM) in postsecondary education.

Cheryan, S., Ziegler, S. A., Montoya, A. K., & Jiang, L. (2017). Why are some STEM fields more gender balanced than others? Psychological Bulletin, 143(1), 1–35. https://doi.org/10.1037/bul0000052

Exploring computer science. (2017). CS Education Statistics. Retrieved May 10, 2017, From http://www.exploringcs.org/resources/cs-statistics

Code.org. (n.d.). Anybody can learn | Code.org. Retrieved May 26, 2016, from https://code.org/

Every Student Succeeds Act of 2015, Pub. L. No. 114-95. 20 U.S.C.A. 6301 (2016).

Else-Quest, N.. Mineo, C & Higgins, A. (2013). Math and science attitudes and achievement at the intersection of gender and ethnicity, *Psychology of Women Quarterly*, 37(3), 293-309.

Ericson, B. (2014). Detailed data on pass rates, race, and gender for 2013. Retrieved May 10, 2017, from http://home.cc.gatech.edu/ice-gt/556

Gannod, G., Burge, J., McIe, V., Doyle, M., & Davis, K. (2014). Increasing awareness of computer science in high school girls.

Goode, J. (2008). Increasing diversity in K-12 computer science: Strategies from the field. In *Proceedings of the 39th SIGCSE Technical Symposium on Computer Science Education* (pp. 362–366). New York, NY, USA: ACM. https://doi.org/10.1145/1352135.1352259

Google Inc. (2014). Women who choose computer science—What really matters. Retrieved from https://static.googleusercontent.com/media/edu.google.com/en//pdfs/women-who-choose-what-really.pdf

Google Inc. & Gallup Inc. (2016a). Diversity gaps in computer science: Exploring the underrepresentation of girls, blacks and hispanics. Retrieved from http://goo.gl/PG34aH.

Google Inc. & Gallup Inc. (2016b). Trends in the state of computer science in U.S. K-12 schools. Retrieved from http://goo.gl/j291E0

Grover & Pea (2013). Computational thinking in K-12: A review of the state of the field. *Educational Researcher*, 42(1), 38-43.

Guzdial, M. (2014, January 17). We may be 100 years behind in making computing education accessible to all | blog@CACM | Communications of the ACM. Retrieved May 10, 2017, from https://cacm.acm.org/blogs/blog-cacm/171475-we-may-be-100-years-behind-in-making-computing-education-accessible-to-all/fulltext

Hazzan, O., Gal-Ezer, J., & Blum, L. (2008). A model for high school computer science education: The four key elements that make it! In *Proceedings of the 39th SIGCSE Technical Symposium on Computer Science Education* (pp. 281–285). New York, NY, USA: ACM. https://doi.org/10.1145/1352135.1352233

Hubwieser, P. (2012). Computer science education in secondary schools – the introduction of a new compulsory subject. *Trans. Comput. Educ.*, *12*(4), 16:1–16:41. https://doi.org/10.1145/2382564.2382568

Irani, L. (2004). Understanding gender and confidence in CS course culture. In Proceedings of the 35th SIGCSE Technical Symposium on Computer Science Education (pp. 195–199). New York, NY, USA: ACM. https://doi.org/10.1145/971300.971371

K–12 computer science framework. (2016). Retrieved from http://www.k12cs.org.

Kulik, J. A. (1994). Meta-analytic studies of findings on computer-based instruction. In E. Baker, H. F. O'Neil Jr, & H. F. O'Neil (Eds.), *Technology Assessment in Education and Training*. Routledge.

Landivar, L. (2013). Disparities in STEM employment by sex, race, and hispanic origin. American Community Survey Reports. Retrieved from https://www.census.gov/prod/2013pubs/acs-24.pdf

Loewus, L. (2016). Math teachers group questions allowing computer science to count as math credit, Education Week. Retrieved from http://blogs.edweek.org/edweek/curriculum/2016/03/math_group_questions_allowing_computer_science_count_as_math_credit.html

Long, M., Conger, D. & Iatarola, P. (2012). Effects of high school course-taking on secondary and postsecondary success, *American Educational Research Journal*, 49(2) 285-322.

Margolis, J., Holme, J. J., Estrella, R., Goode, J., Nao, K., & Stumme, S. (2003). The computer science pipeline in urban schools: Access to what? For whom. *IEEE Technology and Society Magazine*.

Master, A., Cheryan, S., Moscatelli, A., & Meltzoff, A. N. (2017). Programming experience promotes higher STEM motivation among first-grade girls. Journal of Experimental Child Psychology, 160, 92–106. https://doi.org/10.1016/j.jecp.2017.03.013

National Science Foundation, National Center for Science and Engineering Statistics. (2017). Women, minorities, and persons with disabilities in science and engineering: 2017. Retrieved from www.nsf.gov/statistics/wmpd/

Papastergiou, M. (2009). Digital game-based learning in high school computer science education: Impact on educational effectiveness and student motivation. *Computers & Education*, *52*(1), 1–12. https://doi.org/10.1016/j.compedu.2008.06.004

Schiller, K. & Muller, C. (2003). Raising the bar and equity? Effects of state high school graduation requirements and accountability policies on students' mathematics course taking, *Educational Evaluation and Policy Analysis*, 25(3), 299–318.

Sheehy, K. High school graduation requirements get limber. 24 April 2013. US News. https://www.usnews.com/education/blogs/high-school-notes/2013/04/24/high-school-graduation-requirements-get-limber

Teitelbaum, P. (2003). The influence of high school graduation requirement policies in mathematics and science on student course-taking patterns and achievement, *Educational Evaluation and Policy Analysis*, 25(1), 31–57.

U.S. Department of Education, National Center for Education Statistics. (2011). America's high school graduates: Results of the 2009 NAEP high school transcript study. Retrieved from https://nces.ed.gov/nationsreportcard/pubs/studies/2011462.aspx

Whitley Jr., B. E. (1997). Gender differences in computer-related attitudes and behavior: A meta-analysis. Computers in Human Behavior, 13(1), 1–22. https://doi.org/10.1016/S0747-5632(96)00026-X

Wilson, C., Sudol, L. A., Stephenson, C., & Stehlik, M. (2010). Running on empty: The failure to teach K–12 computer science in the digital age. The Association for Computing Machinery & The Computer Science Teachers Association.

Wing, J. (2006). Computational thinking. Communications of the ACM, 49(3), 33–36.

Tables

Table 1. Characteristics of regular public high schools in Texas

	2005	2006	2007	2008	2009
Large City	15.55	15.64	15.83	15.78	15.92
Small/Medium City	7.86	7.64	8.09	7.66	7.65
Town/Suburb	32.35	30.22	29.68	27.90	26.82
Rural	44.24	46.49	46.40	48.66	49.60
Total	*1119*	*1125*	*1112*	*1122*	*1137*
Title I eligible	48.61	49.28	48.73	70.06	72.87
Total	*1119*	*1110*	*1102*	*1119*	*1128*
Charter school	0.71	0.89	0.72	1.34	1.85
Total	*1119*	*1125*	*1112*	*1122*	*1137*
Total Free/Reduced Price Lunch					
Low	20.05	19.95	20.60	20.38	17.13
Medium	72.18	73.56	73.23	72.21	74.71
High	7.77	6.50	6.17	7.42	8.16
Majority	33.33	32.40	32.67	34.14	40.55
Total	*1107*	*1108*	*1102*	*1119*	*1127*
Asian					
Low	99.55	99.46	99.46	99.37	99.29
Medium	0.45	0.54	0.54	0.63	0.71
High	0	0	0	0	0
Majority	0.00	0.00	0.00	0.09	0.09
Latinx					
Low	50.23	48.38	47.01	45.22	43.48
Medium	35.44	36.67	37.48	39.32	40.11
High	14.34	14.95	15.52	15.46	16.42
Majority	26.15	26.76	27.59	28.15	29.46
Black					
Low	83.50	84.77	85.21	85.34	85.36
1Medium	14.52	13.42	12.98	12.87	13.04
High	1.98	1.80	1.81	1.79	1.60
Majority	4.96	4.59	4.45	4.29	3.90
White					
Low	24.80	26.04	26.77	28.42	29.02
Medium	47.70	46.31	47.19	46.38	46.94
High	27.50	27.66	26.04	25.20	24.05
Majority	54.37	54.05	52.81	51.03	49.42
Total	*1109*	*1110*	*1102*	*1119*	*1127*

Percentages shown.

Table 2. Characteristics of schools that offer AP Computer Science, AP Psych

	2005	2006	2007	2008	2009
All schools					
Total enrollment	1,012	1,031	1,045	1,054	1,052
Percentage					
White	50.65	49.87	48.99	47.97	46.66
Latinx	34.52	35.58	36.44	37.34	38.65
Black	12.71	12.36	12.24	12.18	12.09
Asian	1.73	1.81	1.91	2.10	2.16
Total FRL	42.61	41.96	41.82	42.69	45.16
Schools that offer: AP Computer Science					
Total enrollment	2,043	2,010	2,032	2,019	2,001
Percentage					
White	44.27	44.45	45.02	44.22	44.57
Latinx	34.73	35.23	34.33	35.04	35.32
Black	15.65	14.70	14.73	14.23	13.66
Asian	4.99	5.23	5.52	6.13	6.05
Total FRL	35.79	34.10	32.23	33.72	34.76
Schools that offer: AP Psychology					
Total enrollment	2,116	2,116	2,119	2,042	2,049
Percentage					
White	46.11	43.68	43.39	41.59	39.84
Latinx	32.74	35.90	35.32	36.04	38.91
Black	16.05	15.42	15.62	16.29	15.26
Asian	4.69	4.60	5.28	5.68	5.61
Total FRL	33.41	34.29	33.05	35.34	37.70

Averages shown.

Table 3. Percentage of regular high schools that offer the following classes (at least 1 student enrolled)

	AP CS	**AP Psych**	**AP Stat**	**Pre-calculus**	**AP Calculus AB**	**AP Calculus BC**
2005	16.89	17.25	25.74	93.74	78.37	22.79
2006	16.18	17.96	27.91	94.31	78.13	24.00
2007	16.64	20.32	31.29	95.59	77.97	23.92
2008	17.65	23.26	32.62	95.99	76.56	24.15
2009	17.85	24.45	34.30	96.48	75.64	24.89

Table 4. Percentage of students in regular high schools who take AP Computer Science, by race-gender

AP Computer Science	2005	2006	2007	2008	2009
White Male					
Lower bound	9.71	9.70	9.48	10.34	7.91
Random (1-4)	13.75	13.24	13.55	14.19	11.06
Upper bound	17.18	17.23	16.67	17.75	15.73
Obs.	204	215	233	217	218
White Female					
Lower bound	4.23	4.49	4.55	3.71	4.03
Random (1-4)	6.37	5.89	7.10	6.26	6.79
Upper bound	8.45	7.25	9.76	8.47	8.72
Obs.	201	210	228	216	219
Latino Male					
Lower bound	4.19	3.41	4.25	4.59	4.65
Random (1-4)	7.33	6.44	7.09	7.48	8.06
Upper bound	9.54	9.65	9.67	10.56	10.60
Obs.	208	217	232	219	221
Latino Female					
Lower bound	2.53	3.28	3.09	3.19	2.72
Random (1-4)	4.29	4.18	4.80	4.83	5.08
Upper bound	5.84	5.81	5.88	6.30	6.94
Obs.	203	216	231	219	222
Black Male					
Lower bound	3.30	2.65	1.77	1.76	3.03
Random (1-4)	5.88	4.48	3.30	4.74	6.19
Upper bound	7.82	7.38	5.15	6.66	8.14
Obs.	191	199	212	201	204
Black Female					
Lower bound	1.73	1.51	1.74	1.78	1.99
Random (1-4)	2.71	2.34	2.53	3.10	2.83
Upper bound	3.93	3.18	3.97	4.60	3.61
Obs.	188	199	211	196	202
Asian Male					
Lower bound	11.94	11.72	11.17	11.67	11.08
Random (1-4)	21.14	20.70	18.78	18.58	19.95
Upper bound	27.43	28.71	26.11	27.51	27.30
Obs.	159	173	186	173	180
Asian Female					
Lower bound	8.34	7.46	7.14	8.28	5.92
Random (1-4)	14.77	11.67	13.29	14.04	10.77
Upper bound	19.96	14.97	16.43	18.35	15.21
Obs.	155	176	184	173	186

Table 5. Difference-in-difference estimates by grade (starting year)

	AP Computer Science					
	By start year			Force 100%		
	LB	UB	Rand	LB	UB	Rand
White male	-2.100+	-2.173+	-2.167+	-2.464+	-2.397*	-2.447*
SE	(1.216)	(1.218)	(1.220)	(1.261)	(1.206)	(1.221)
Obs.	2342	2342	2342	2384	2384	2384
R-squared	0.570	0.641	0.615	0.626	0.676	0.656
White female	-1.272*	0.349	-0.643	-1.559+	0.0632	-0.999
SE	(0.594)	(0.779)	(0.718)	(0.898)	(0.974)	(0.965)
Obs.	2306	2306	2306	2340	2340	2340
R-squared	0.516	0.564	0.560	0.556	0.587	0.584
Latino male	-0.430	-1.322	-1.101	-1.426*	-2.262*	-2.087*
SE	(0.379)	(0.832)	(0.739)	(0.704)	(0.999)	(0.929)
Obs.	2368	2368	2368	2382	2382	2382
R-squared	0.359	0.408	0.403	0.416	0.431	0.427
Latina female	-0.805	-0.475	-0.370	-1.038	-0.629	-0.522
SE	(0.579)	(0.664)	(0.632)	(0.793)	(0.840)	(0.812)
Obs.	2348	2348	2348	2358	2358	2358
R-squared	0.342	0.319	0.318	0.427	0.381	0.392
Black male	-0.410	-0.991	-0.275	-1.374+	-1.945+	-1.233
SE	(0.351)	(0.810)	(0.558)	(0.707)	(1.014)	(0.830)
Obs.	2194	2194	2194	2208	2208	2208
R-squared	0.238	0.294	0.228	0.285	0.301	0.266
Black female	0.234	0.601	0.242	-0.667	-0.286	-0.658
SE	(0.522)	(0.740)	(0.605)	(0.888)	(1.030)	(0.936)
Obs.	2182	2182	2182	2192	2192	2192
R-squared	0.236	0.336	0.278	0.247	0.313	0.266
Asian male	0.0229	0.0946	0.252	-0.0733	0.0348	0.166
SE	(1.057)	(1.838)	(1.613)	(1.181)	(1.894)	(1.692)
Obs.	1927	1927	1927	1968	1968	1968
R-squared	0.282	0.238	0.224	0.475	0.323	0.358
Asian female	0.383	1.321	1.176	-0.641	0.449	0.305
SE	(0.732)	(1.647)	(1.147)	(1.241)	(1.870)	(1.463)
Obs.	1910	1910	1910	1938	1938	1938
R-squared	0.298	0.252	0.264	0.435	0.319	0.358

AP Psychology						
	By start year			Force 100%		
	LB	UB	Rand	LB	UB	Rand
White male	0.296	0.525	0.354	-0.0420	-0.0189	-0.182
SE	(0.925)	(1.175)	(1.049)	(1.132)	(1.337)	(1.236)
Obs.	2435	2435	2435	2468	2468	2468
R-squared	0.593	0.585	0.597	0.603	0.601	0.612
White female	0.550	0.799	0.467	0.101	0.347	0.0656
SE	(1.170)	(1.355)	(1.302)	(1.365)	(1.488)	(1.453)
Obs.	2415	2415	2415	2463	2463	2463
R-squared	0.669	0.639	0.647	0.660	0.644	0.649
Latino male	0.834+	0.899	0.919	1.115	0.921	1.034
SE	(0.465)	(0.883)	(0.654)	(0.831)	(1.019)	(0.864)
Obs.	2471	2471	2471	2487	2487	2487
R-squared	0.408	0.520	0.462	0.389	0.500	0.447
Latina female	-0.188	0.343	-0.0541	-0.665	-0.244	-0.541
SE	(0.669)	(1.195)	(1.025)	(0.936)	(1.320)	(1.207)
Obs.	2459	2459	2459	2485	2485	2485
R-squared	0.577	0.561	0.537	0.553	0.557	0.535
Black male	-0.139	-0.372	0.134	-0.705	-0.900	-0.428
SE	(0.717)	(1.149)	(0.957)	(1.084)	(1.402)	(1.258)
Obs.	2326	2326	2326	2347	2347	2347
R-squared	0.336	0.342	0.306	0.391	0.389	0.374
Black female	0.224	0.315	0.286	0.550	0.684	0.625
SE	(0.735)	(1.135)	(1.044)	(1.183)	(1.409)	(1.347)
Obs.	2317	2317	2317	2344	2344	2344
R-squared	0.344	0.354	0.342	0.430	0.410	0.415
Asian male	-0.670	-1.212	-0.815	-1.233	-1.657	-1.307
SE	(1.133)	(1.939)	(1.646)	(1.531)	(2.126)	(1.918)
Obs.	2062	2062	2062	2096	2096	2096
R-squared	0.352	0.305	0.309	0.510	0.380	0.416
Asian female	0.699	2.154	1.190	0.278	1.470	0.623
SE	(1.221)	(2.127)	(1.713)	(1.735)	(2.397)	(2.074)
Obs.	2035	2035	2035	2091	2091	2091
R-squared	0.362	0.300	0.304	0.462	0.356	0.381

All regressions use school, year and yearinschool (grade) fixed effects. Clustered (school) standard errors are used. + p<0.10 * p<0.05 ** p<0.01 *** p<0.001. These tables use all available observations (not balanced). CCD enrollment that is 0 is treated as missing. Similar results hold when positive course enrollment with CCD enrollment that is 0 is treated as 100% instead of missing (significantly negative for LM, BM in APCS by start year), or when data is restricted to balanced by outcome variable (positive for LM in AP CS under certain dependent variables) or by schools that offer the respective course (no significant results for AP CS).

Table 6. Triple difference estimates, AP CS and AP Psych, treated/untreated grades

	By start year			Force 100%		
	LB	UB	Rand	LB	UB	Rand
White male	-2.977*	-3.355*	-3.113*	-3.066*	-2.983+	-2.867+
SE	(1.275)	(1.519)	(1.283)	(1.524)	(1.653)	(1.470)
Obs.	4791	4791	4791	4867	4867	4867
R-squared	0.548	0.596	0.583	0.595	0.627	0.619
White female	-3.313*	-2.307	-2.945*	-3.325*	-2.273	-3.024+
SE	(1.297)	(1.576)	(1.497)	(1.455)	(1.634)	(1.591)
Obs.	4734	4734	4734	4817	4817	4817
R-squared	0.509	0.533	0.525	0.540	0.556	0.551
Latino male	-0.967	-1.904	-1.757	-2.770*	-3.408*	-3.393*
SE	(0.663)	(1.371)	(1.103)	(1.121)	(1.516)	(1.322)
Obs.	4852	4852	4852	4883	4883	4883
R-squared	0.336	0.413	0.385	0.377	0.426	0.406
Latina female	-1.015	-1.458	-0.961	-0.970	-1.271	-0.799
SE	(0.880)	(1.444)	(1.252)	(1.198)	(1.603)	(1.467)
Obs.	4820	4820	4820	4858	4858	4858
R-squared	0.386	0.394	0.377	0.427	0.424	0.418
Black male	-0.311	-0.713	-0.598	-0.946	-1.391	-1.254
SE	(0.609)	(1.301)	(1.008)	(1.113)	(1.603)	(1.381)
Obs.	4538	4538	4538	4575	4575	4575
R-squared	0.212	0.258	0.205	0.284	0.289	0.265
Black female	-0.0256	0.0426	-0.148	-1.311	-1.205	-1.419
SE	(0.786)	(1.284)	(1.109)	(1.407)	(1.702)	(1.569)
Obs.	4515	4515	4515	4553	4553	4553
R-squared	0.241	0.299	0.259	0.321	0.340	0.321
Asian male	0.470	0.832	0.226	0.729	1.073	0.470
SE	(1.665)	(2.516)	(2.188)	(1.822)	(2.558)	(2.288)
Obs.	4011	4011	4011	4086	4086	4086
R-squared	0.290	0.240	0.233	0.472	0.325	0.360
Asian female	0.319	0.163	0.990	-1.012	-0.889	-0.157
SE	(1.562)	(2.757)	(2.175)	(2.096)	(3.017)	(2.531)
Obs.	3972	3972	3972	4053	4053	4053
R-squared	0.302	0.258	0.262	0.416	0.317	0.344

All regressions use school by year, school by year in school,and yearinschool by year fixed effects. Clustered (school) standard errors are used. + p<0.10 * p<0.05 ** p<0.01 *** p<0.001

Table 7. Math triple difference estimates
All regressions use all two-way (year-yearinschool, year-school, yearinschool-school) fixed effects. Clustered (school) standard errors are used. + p<0.10 * p<0.05 ** p<0.01 *** p<0.001")

7A. Precalculus

	By start year			Force 100%		
	LB	UB	Rand	LB	UB	Rand
White male	-0.515	-0.540	-0.781	-0.584	-0.382	-0.799
SE	(0.878)	(1.330)	(1.093)	(1.094)	(1.442)	(1.268)
Obs.	11886	11886	11886	12044	12044	12044
R-squared	0.848	0.792	0.792	0.856	0.803	0.807
White female	-0.309	0.393	-0.0862	-0.361	0.555	0.0284
SE	(0.868)	(1.207)	(1.094)	(1.135)	(1.395)	(1.294)
Obs.	11727	11727	11727	11948	11948	11948
R-squared	0.834	0.788	0.785	0.854	0.804	0.809
Latino male	0.685	1.573	1.672	0.559	1.421	1.496
SE	(0.885)	(1.213)	(1.057)	(1.068)	(1.337)	(1.205)
Obs.	11649	11649	11649	11773	11773	11773
R-squared	0.825	0.768	0.769	0.833	0.775	0.780
Latina female	1.522+	2.322*	2.182*	1.490	2.220+	2.043+
SE	(0.864)	(1.095)	(1.050)	(1.049)	(1.216)	(1.188)
Obs.	11529	11529	11529	11705	11705	11705
R-squared	0.813	0.765	0.765	0.829	0.778	0.783
Black male	-0.788	0.693	0.339	-0.0393	1.244	0.797
SE	(0.995)	(1.387)	(1.289)	(1.204)	(1.494)	(1.424)
Obs.	8896	8896	8896	9059	9059	9059
R-squared	0.759	0.711	0.711	0.801	0.740	0.750
Black female	2.558*	3.112+	3.631*	3.058*	3.629+	4.234**
SE	(1.125)	(1.823)	(1.472)	(1.261)	(1.878)	(1.570)
Obs.	8698	8698	8698	8856	8856	8856
R-squared	0.747	0.701	0.704	0.789	0.727	0.738
Asian male	1.238	4.873	2.304	1.710	5.206	2.633
SE	(2.076)	(3.325)	(2.870)	(2.263)	(3.355)	(2.956)
Obs.	5117	5117	5117	5320	5320	5320
R-squared	0.721	0.705	0.701	0.754	0.717	0.720
Asian female	-0.556	2.653	1.269	-0.805	2.413	1.208
SE	(2.170)	(3.400)	(2.949)	(2.517)	(3.467)	(3.103)
Obs.	4990	4990	4990	5234	5234	5234
R-squared	0.724	0.701	0.685	0.765	0.717	0.714

7B. AP Calculus AB

	By start year			Force 100%		
	LB	UB	Rand	LB	UB	Rand
White male	-1.788	-1.198	-0.766	-0.169	1.501	1.567
SE	(1.899)	(2.595)	(2.416)	(2.409)	(2.882)	(2.764)
Obs.	2986	2986	2986	3042	3042	3042
R-squared	0.871	0.783	0.800	0.877	0.810	0.828
White female	-5.598*	-4.231+	-5.607*	-3.276	-0.875	-2.335
SE	(2.176)	(2.536)	(2.351)	(2.641)	(2.856)	(2.718)
Obs.	2948	2948	2948	3005	3005	3005
R-squared	0.840	0.835	0.831	0.835	0.839	0.834
Latino male	-1.781	-3.110	-1.820	-1.985	-3.219	-1.967
SE	(1.444)	(2.036)	(1.841)	(2.004)	(2.424)	(2.282)
Obs.	3104	3104	3104	3132	3132	3132
R-squared	0.798	0.806	0.795	0.804	0.810	0.800
Latina female	-0.630	0.950	0.458	0.645	2.082	1.714
SE	(1.027)	(1.231)	(1.241)	(1.687)	(1.770)	(1.813)
Obs.	3079	3079	3079	3121	3121	3121
R-squared	0.810	0.821	0.800	0.818	0.827	0.812
Black male	-0.263	1.135	0.528	0.196	1.570	0.939
SE	(0.966)	(2.296)	(1.719)	(1.501)	(2.536)	(2.069)
Obs.	2613	2613	2613	2637	2637	2637
R-squared	0.790	0.787	0.759	0.781	0.783	0.756
Black female	-2.146	-2.282	-3.140	-1.580	-1.970	-2.715
SE	(1.811)	(2.645)	(2.291)	(2.027)	(2.757)	(2.440)
Obs.	2599	2599	2599	2615	2615	2615
R-squared	0.750	0.762	0.753	0.765	0.771	0.767
Asian male	2.276	4.541	-1.178	4.544	7.080	1.213
SE	(3.257)	(5.763)	(5.062)	(3.663)	(5.848)	(5.283)
Obs.	2013	2013	2013	2067	2067	2067
R-squared	0.769	0.755	0.747	0.806	0.766	0.766
Asian female	2.050	-0.247	1.520	0.622	-1.607	0.376
SE	(3.111)	(5.584)	(4.979)	(4.131)	(5.955)	(5.392)
Obs.	1991	1991	1991	2060	2060	2060
R-squared	0.766	0.732	0.710	0.797	0.742	0.735

7C. AP Calculus BC

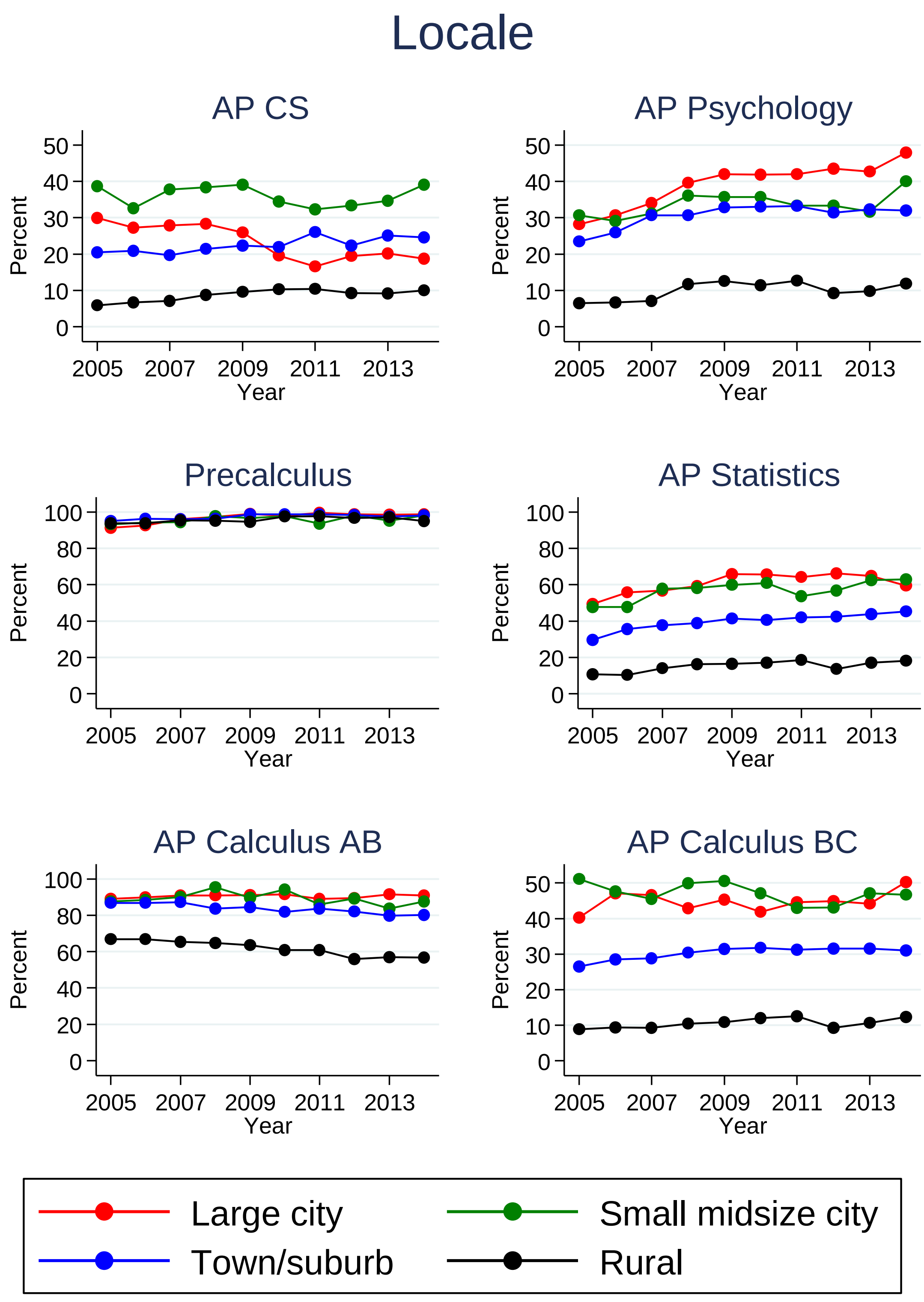
	By start year			Force 100%		
	LB	UB	Rand	LB	UB	Rand
White male	-0.717	-0.143	-1.069	-1.523	-0.645	-1.894
SE	(2.191)	(3.059)	(2.374)	(2.767)	(4.071)	(3.098)
Obs.	1027	1027	1027	1043	1043	1043
R-squared	0.915	0.882	0.899	0.851	0.853	0.853
White female	-2.651	-0.599	-1.773	-4.345	-2.247	-3.380
SE	(3.801)	(4.035)	(3.832)	(4.291)	(4.503)	(4.338)
Obs.	1015	1015	1015	1037	1037	1037
R-squared	0.883	0.852	0.892	0.892	0.873	0.896
Latino male	0.708	0.353	1.292	0.339	-0.0531	0.889
SE	(1.050)	(2.779)	(1.727)	(2.084)	(3.261)	(2.465)
Obs.	1033	1033	1033	1041	1041	1041
R-squared	0.849	0.784	0.774	0.809	0.782	0.774
Latina female	0.719	4.144	1.876	-2.294	1.165	-1.121
SE	(0.833)	(2.678)	(2.181)	(2.268)	(3.346)	(2.972)
Obs.	1039	1039	1039	1047	1047	1047
R-squared	0.831	0.779	0.801	0.809	0.774	0.790
Black male	-0.802	-3.579	-2.103	0.301	-2.456	-1.031
SE	(0.963)	(3.684)	(2.219)	(1.466)	(3.804)	(2.443)
Obs.	967	967	967	973	973	973
R-squared	0.666	0.662	0.697	0.763	0.700	0.749
Black female	-0.860	-1.402	-1.752	-2.491	-2.938	-3.313
SE	(1.494)	(3.444)	(3.239)	(2.207)	(3.737)	(3.561)
Obs.	968	968	968	974	974	974
R-squared	0.858	0.722	0.722	0.897	0.782	0.796
Asian male	-1.491	-1.331	2.266	-7.262	-5.033	-3.823
SE	(4.116)	(7.604)	(6.571)	(5.202)	(7.575)	(7.084)
Obs.	937	937	937	963	963	963
R-squared	0.792	0.788	0.782	0.817	0.804	0.799
Asian female	3.716	2.584	5.332	-1.442	-2.253	-0.226
SE	(3.404)	(6.441)	(6.031)	(6.180)	(7.684)	(7.592)
Obs.	906	906	906	930	930	930
R-squared	0.810	0.784	0.781	0.801	0.783	0.779

7D. AP Statistics

	By start year			Force 100%		
	LB	UB	Rand	LB	UB	Rand
White male	-2.714*	-3.914*	-4.246*	-3.035	-3.003	-3.997+
SE	(1.334)	(1.971)	(1.793)	(1.897)	(2.288)	(2.159)
Obs.	2355	2355	2355	2412	2412	2412
R-squared	0.888	0.862	0.868	0.896	0.878	0.881
White female	-1.169	-1.566	-1.466	-0.0175	-0.292	-0.162
SE	(1.681)	(2.007)	(1.787)	(2.071)	(2.324)	(2.143)
Obs.	2332	2332	2332	2401	2401	2401
R-squared	0.896	0.882	0.891	0.893	0.884	0.890
Latino male	-1.019	-3.272*	-1.849	-2.034	-4.318*	-2.926
SE	(0.711)	(1.412)	(1.320)	(1.604)	(1.973)	(1.933)
Obs.	2366	2366	2366	2396	2396	2396
R-squared	0.821	0.808	0.774	0.787	0.795	0.764
Latina female	-2.939*	-1.806	-2.824+	-2.512	-1.510	-2.466
SE	(1.309)	(1.641)	(1.476)	(1.851)	(2.070)	(1.954)
Obs.	2374	2374	2374	2416	2416	2416
R-squared	0.844	0.833	0.827	0.826	0.826	0.820
Black male	0.729	1.829	2.080	-0.0873	0.983	1.254
SE	(0.916)	(2.077)	(1.481)	(1.575)	(2.391)	(1.926)
Obs.	2207	2207	2207	2245	2245	2245
R-squared	0.697	0.685	0.667	0.706	0.704	0.684
Black female	-0.671	0.809	1.075	-1.098	0.240	0.555
SE	(1.071)	(2.264)	(1.925)	(1.811)	(2.647)	(2.392)
Obs.	2208	2208	2208	2232	2232	2232
R-squared	0.696	0.725	0.707	0.704	0.730	0.714
Asian male	0.135	3.971	5.151	1.877	5.162	6.365
SE	(2.519)	(5.442)	(4.408)	(2.732)	(5.459)	(4.468)
Obs.	1903	1903	1903	1952	1952	1952
R-squared	0.735	0.724	0.702	0.797	0.750	0.746
Asian female	0.615	-4.981	-6.462	2.377	-3.188	-4.484
SE	(3.033)	(6.251)	(5.287)	(3.274)	(6.222)	(5.316)
Obs.	1790	1790	1790	1839	1839	1839
R-squared	0.718	0.716	0.790	0.735	0.742	0.718

Figure

Figure 1. Percentage of regular high schools that offer classes, by school characteristics



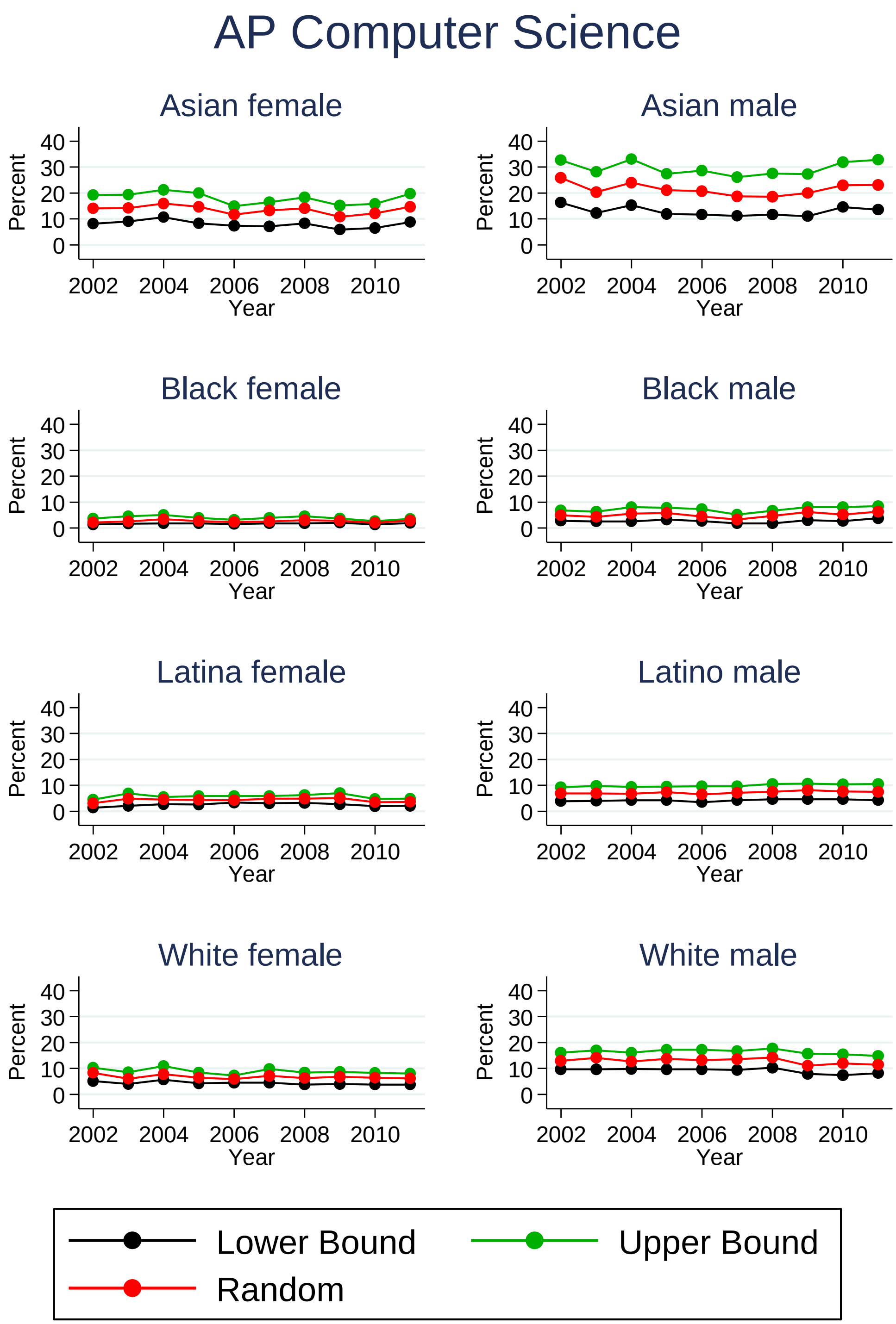
Appendix

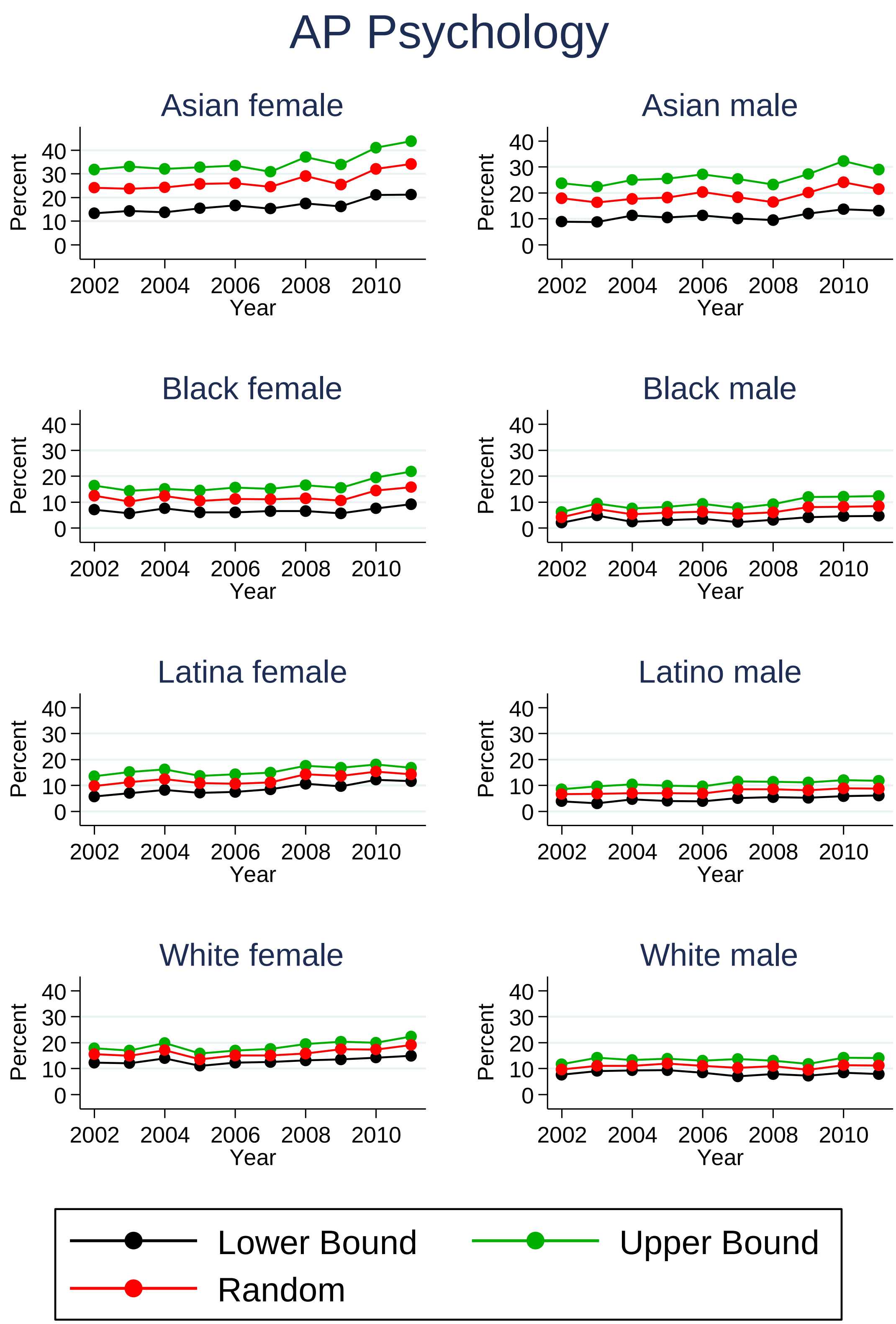
Appendix A-1. Implementation of allowing CS to count towards a math graduation requirement

	2005-06	2006-07	2007-08	2008-09	2009-10
Freshman	2005	2006	2007	2008	2009
Sophomore	2004	2005	2006	2007	2008
Junior	2003	2004	2005	2006	2007
Senior	2002	2003	2004	2005	2006

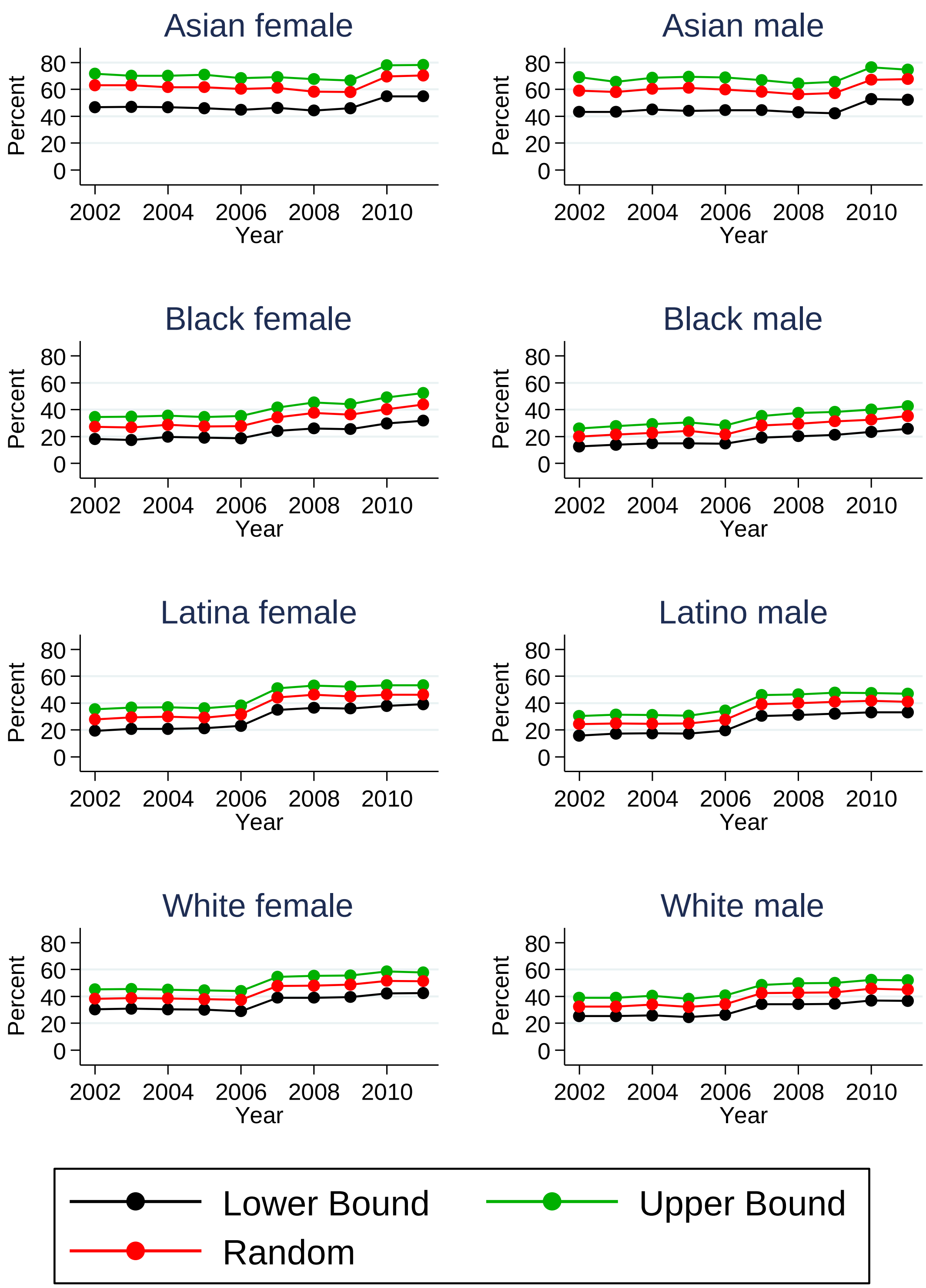
Note: This table describes the cascading nature of the implementation of the policy. The cohort of incoming students could be followed in a diagonal line (i.e. freshmen who entered in 2005 were freshmen in 2005-06, then sophomores in 2006-07, etc.). The green line indicates when the policy is binding for which cohorts of students and the red values indicate when a cohort of students is impacted by the policy. The table illustrates the difference in difference: treated grades in treated years. A similar table can be constructed for the control subject (AP Psychology) for the triple difference.

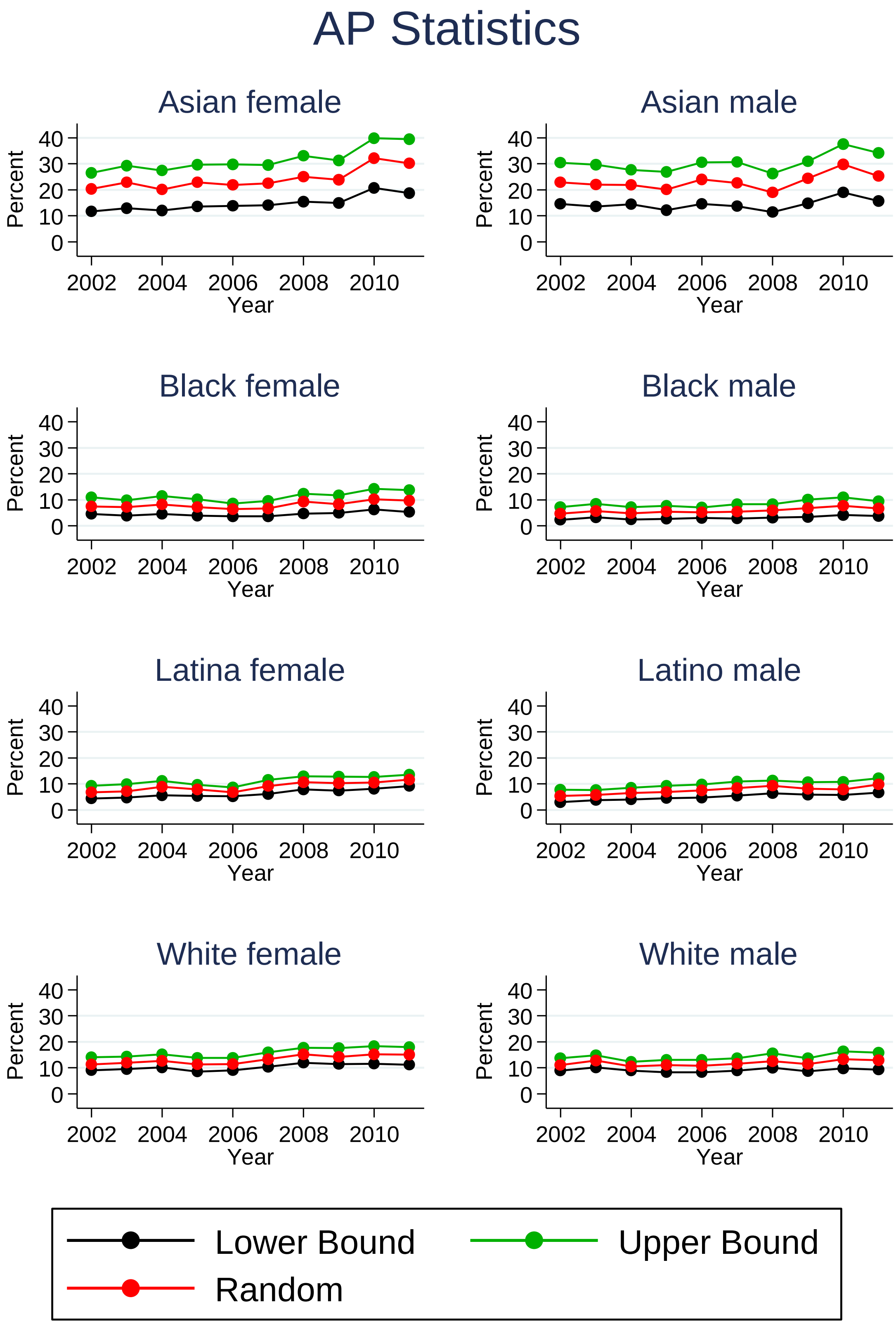
Appendix A-2. Overall course-taking, by race/gender and cohort

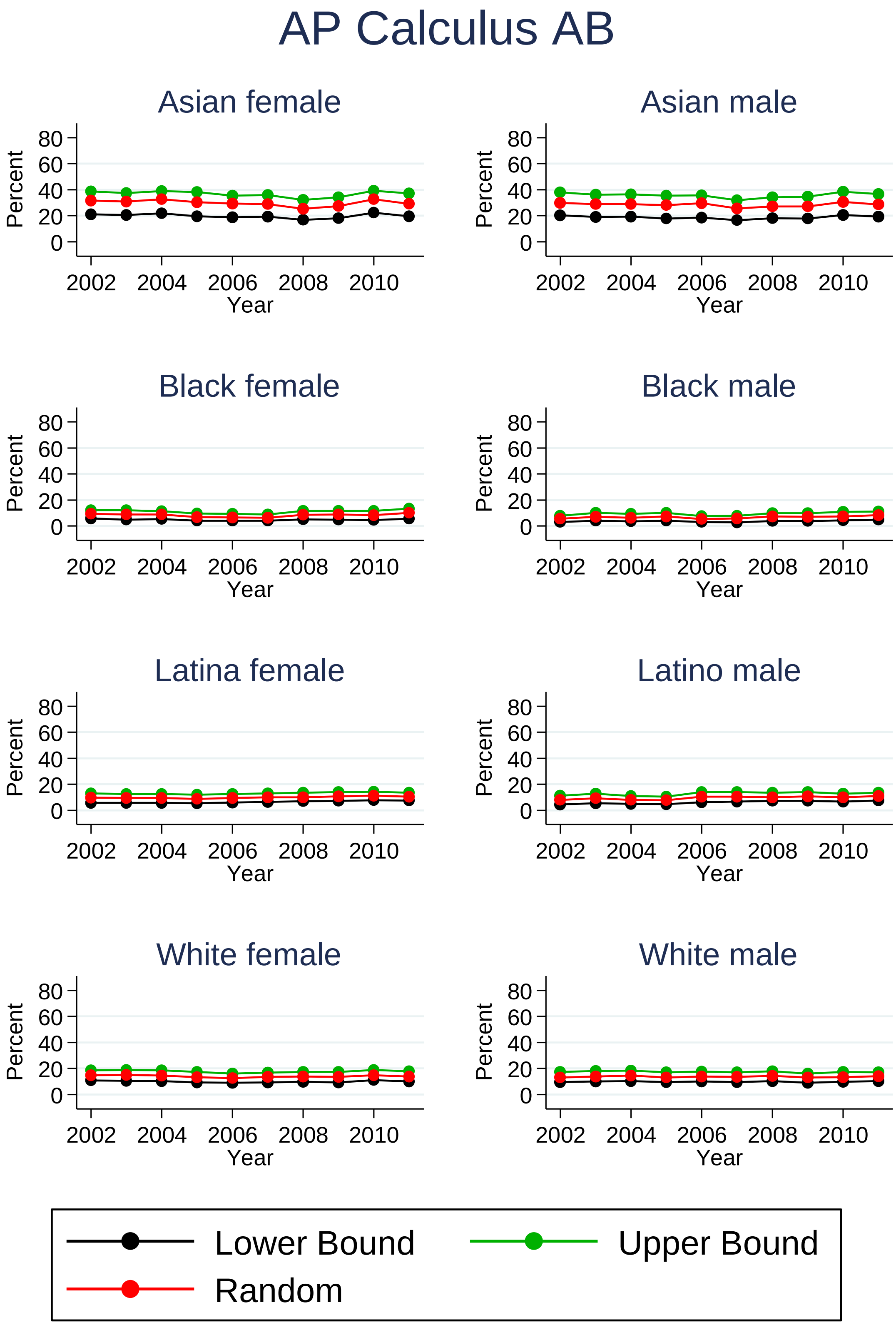


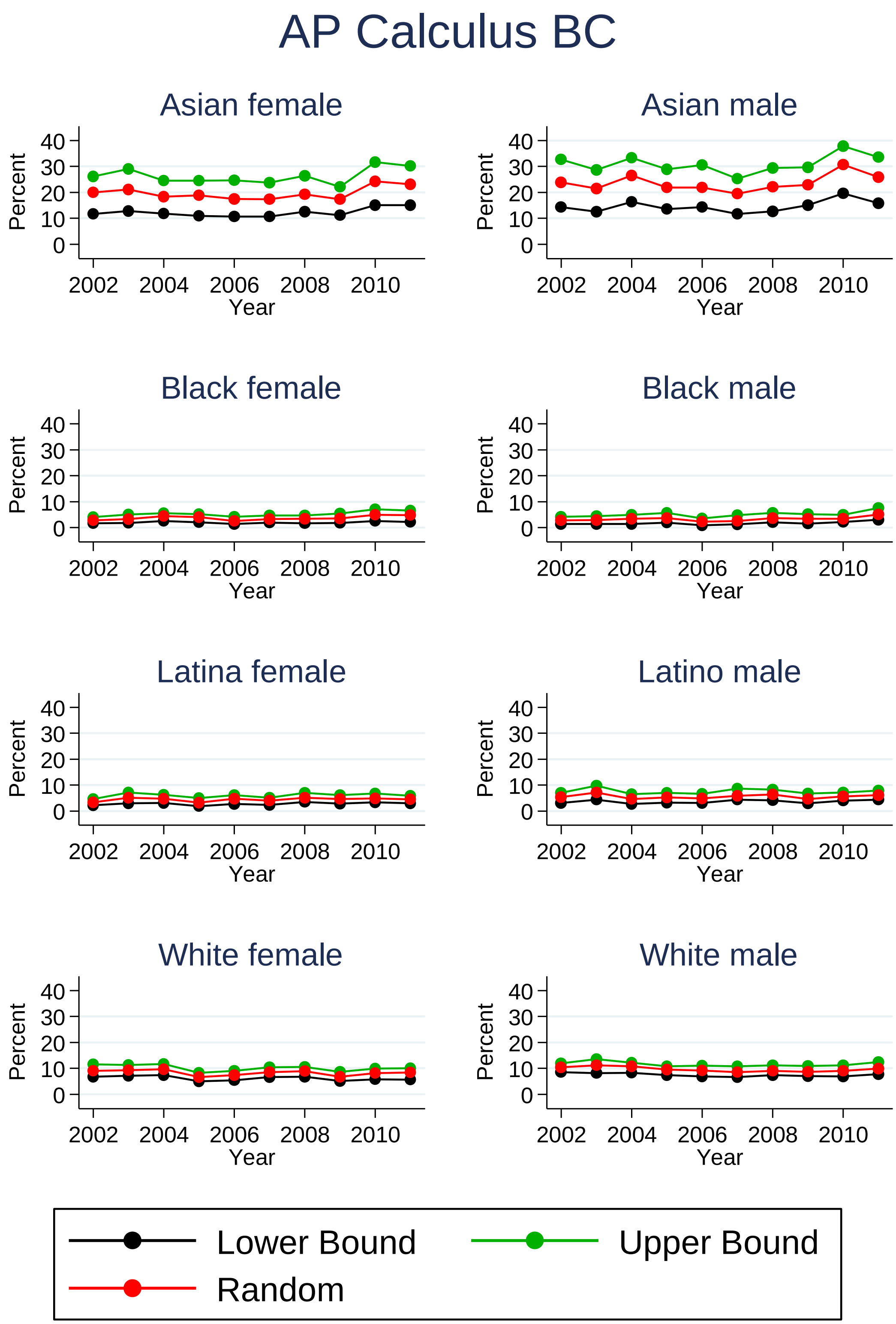

AP Psychology
Asian female
Asian male
Black female
Black male
Latina female
Latino male
White female
White male
Percent
0
10
20
30
40
2002
2004
2006
2008
2010
Year
Lower Bound
Upper Bound
Random

Pre-Calculus

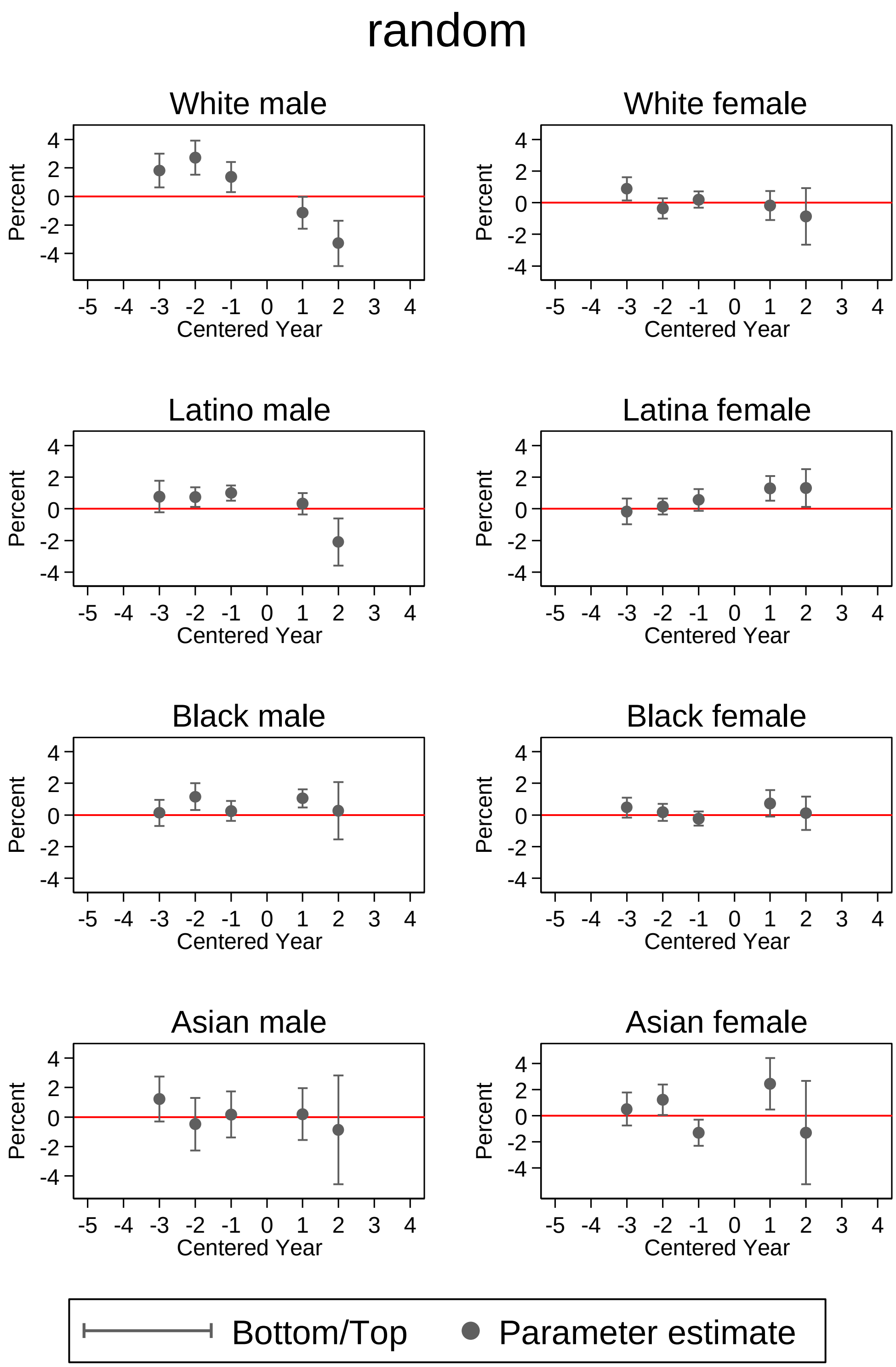



AP Statistics
Asian female
Asian male
Black female
Black male
Latina female
Latino male
White female
White male
Percent
Year
0
10
20
30
40
2002
2004
2006
2008
2010
Lower Bound
Upper Bound
Random


AP Calculus AB
Asian female
Asian male
Black female
Black male
Latina female
Latino male
White female
White male
Percent
0
20
40
60
80
2002
2004
2006
2008
2010
Year
Lower Bound
Upper Bound
Random


AP Calculus BC
Asian female
Asian male
Black female
Black male
Latina female
Latino male
White female
White male
Percent
Year
0
10
20
30
40
2002
2004
2006
2008
2010
Lower Bound
Upper Bound
Random

Appendix A-3. Falsification checks for difference-in-difference: AP CS (random imputation).



Appendix A-4. Triple difference estimates, using 2006 as policy start year.

	By start year			Force 100%		
	LB	UB	Rand	LB	UB	Rand
White male	-0.852	1.388	0.173	-1.214	1.514	0.285
SE	(0.995)	(1.579)	(1.467)	(1.454)	(1.822)	(1.757)
Obs.	4182	4182	4182	4260	4260	4260
White female	-0.686	-1.427	-1.661	-0.784	-1.631	-1.838
SE	(1.406)	(1.591)	(1.431)	(1.754)	(1.847)	(1.705)
Obs.	4124	4124	4124	4205	4205	4205
Latino male	-0.573	0.546	-0.795	-1.215	-0.0468	-1.397
SE	(0.904)	(1.405)	(1.072)	(1.262)	(1.617)	(1.363)
Obs.	4238	4238	4238	4270	4270	4270
Latina female	1.883	1.741	1.677	2.836*	2.643	2.588+
SE	(1.196)	(1.726)	(1.377)	(1.375)	(1.832)	(1.518)
Obs.	4208	4208	4208	4246	4246	4246
Black male	-0.628	-0.294	-0.652	-1.176	-0.845	-1.202
SE	(0.693)	(1.164)	(1.005)	(1.042)	(1.369)	(1.245)
Obs.	3953	3953	3953	3983	3983	3983
Black female	0.0576	-1.534	-0.606	-0.457	-1.981	-1.076
SE	(1.094)	(1.497)	(1.304)	(1.362)	(1.639)	(1.486)
Obs.	3931	3931	3931	3969	3969	3969
Asian male	1.067	3.024	-0.372	-0.0557	1.890	-1.431
SE	(1.669)	(2.575)	(2.379)	(1.930)	(2.723)	(2.536)
Obs.	3476	3476	3476	3553	3553	3553
Asian female	-1.633	-4.795+	-3.234	-1.727	-4.924+	-3.297
SE	(1.287)	(2.531)	(2.155)	(1.639)	(2.650)	(2.348)
Obs.	3416	3416	3416	3496	3496	3496

Note that seniors for 2004-05 have missing data. All regressions use school and all two-way (year-"grade", year-subject, "grade"-subject) fixed effects. Clustered (school) standard errors are used. + p<0.10 * p<0.05 ** p<0.01 *** p<0.001

Appendix A-5. Triple difference estimates, using AP Macro as “treated” subject vs AP Psych.

	By start year			Force 100%		
	LB	UB	Rand	LB	UB	Rand
White male	-0.666	-1.812	-1.324	-0.747	-1.987	-1.473
SE	(1.425)	(1.683)	(1.527)	(1.579)	(1.795)	(1.663)
Obs.	5156	5156	5156	5235	5235	5235
White female	-0.624	-0.237	-1.151	-0.316	0.110	-0.821
SE	(1.351)	(2.020)	(1.739)	(1.656)	(2.193)	(1.957)
Obs.	5109	5109	5109	5223	5223	5223
Latino male	0.315	0.590	0.218	0.934	1.266	0.872
SE	(0.830)	(1.379)	(1.116)	(1.155)	(1.552)	(1.371)
Obs.	5234	5234	5234	5280	5280	5280
Latina female	0.572	1.291	1.007	0.200	1.023	0.718
SE	(1.038)	(1.504)	(1.362)	(1.317)	(1.676)	(1.580)
Obs.	5199	5199	5199	5268	5268	5268
Black male	-0.646	-0.294	0.167	-1.210	-0.861	-0.379
SE	(1.109)	(1.643)	(1.527)	(1.320)	(1.762)	(1.678)
Obs.	4847	4847	4847	4902	4902	4902
Black female	-0.349	-3.644+	-2.066	0.160	-3.131	-1.553
SE	(1.192)	(1.949)	(1.654)	(1.243)	(1.964)	(1.669)
Obs.	4840	4840	4840	4901	4901	4901
Asian male	0.975	0.688	0.0682	0.0749	-0.212	-0.800
SE	(1.750)	(3.001)	(2.604)	(1.888)	(3.035)	(2.658)
Obs.	4169	4169	4169	4267	4267	4267
Asian female	-2.081	-1.976	-3.797	-1.368	-1.616	-3.218
SE	(1.955)	(3.311)	(2.795)	(2.346)	(3.440)	(3.017)
Obs.	4086	4086	4086	4237	4237	4237

All regressions use school and all two-way (year-"grade", year-subject, "grade"-subject) fixed effects. Clustered (school) standard errors are used. + $p<0.10$ * $p<0.05$ ** $p<0.01$ *** $p<0.001$

Appendix A-6. Triple difference estimates, using AP test-taking data.

	Treatment	SE	Observations
White male	0.000531	(0.0750)	546
White female	-0.103	(0.141)	546
Asian male	-0.444	(0.310)	546
Asian female	-0.279	(0.283)	546
Latino male	-0.0895	(0.102)	546
Latina female	-0.465	(0.470)	546
Black male	-0.0774	(0.0771)	546
Black female	-0.0756	(0.0947)	546

Note: Analyses include fully saturated triple difference model (i.e. two-way fixed effects for state-year, state-subject, subject-year) and robust standard errors.

Appendix A-7. Triple difference estimates, using AP Psychology instead of math subject.

	By start year			Force 100%		
	LB	UB	Rand	LB	UB	Rand
White male	1.370	1.175	1.883	3.177	2.965	3.612
SE	(1.573)	(2.264)	(1.986)	(1.996)	(2.576)	(2.297)
Obs.	2123	2123	2123	2165	2165	2165
White female	1.128	0.119	0.540	1.594	1.173	1.265
SE	(2.323)	(2.566)	(2.515)	(2.680)	(2.864)	(2.860)
Obs.	2096	2096	2096	2169	2169	2169
Latino male	1.081	2.445	2.871*	1.861	2.789	3.327*
SE	(0.907)	(2.025)	(1.219)	(1.398)	(2.130)	(1.424)
Obs.	2163	2163	2163	2181	2181	2181
Latina female	0.522	2.459	1.669	1.097	2.288	2.235
SE	(1.040)	(2.057)	(1.932)	(1.475)	(2.317)	(2.169)
Obs.	2151	2151	2151	2183	2183	2183
Black male	1.728	2.460	1.925	1.742	2.701	2.033
SE	(1.164)	(2.646)	(1.982)	(1.761)	(2.935)	(2.339)
Obs.	1984	1984	1984	2017	2017	2017
Black female	2.003	3.289	2.687	2.554	3.157	2.664
SE	(1.516)	(2.797)	(2.514)	(2.139)	(3.033)	(2.797)
Obs.	1992	1992	1992	2024	2024	2024
Asian male	3.573+	8.512*	7.290*	7.261*	11.90**	11.05**
SE	(2.089)	(4.202)	(3.561)	(3.141)	(4.551)	(4.188)
Obs.	1746	1746	1746	1785	1785	1785
Asian female	-0.806	-5.868	-4.379	0.0612	-4.564	-3.235
SE	(2.183)	(4.911)	(4.431)	(3.642)	(5.541)	(5.173)
Obs.	1730	1730	1730	1798	1798	1798

All regressions use school and all two-way (year-"grade", year-subject, "grade"-subject) fixed effects. Clustered (school) standard errors are used. + p<0.10 * p<0.05 ** p<0.01 *** p<0.001

Appendix A-8. Math triple difference estimates, using 2006 as implementation year.

All regressions use all two-way (year-yearinschool, year-school, yearinschool-school) fixed effects. Clustered (school) standard errors are used. + p<0.10 * p<0.05 ** p<0.01 *** p<0.001")

8A. Precalculus

	By start year			Force 100%		
	LB	UB	Rand	LB	UB	Rand
White male	0.686	1.439	1.075	0.0612	0.774	0.419
SE	(0.790)	(1.261)	(1.089)	(0.922)	(1.325)	(1.176)
Obs.	11886	11886	11886	12044	12044	12044
White female	1.533*	1.641	1.714	1.180	1.369	1.453
SE	(0.778)	(1.152)	(1.058)	(0.935)	(1.240)	(1.156)
Obs.	11727	11727	11727	11948	11948	11948
Latino male	-0.180	-1.006	-1.600	-0.140	-0.923	-1.495
SE	(0.717)	(1.035)	(0.985)	(0.878)	(1.140)	(1.092)
Obs.	11649	11649	11649	11773	11773	11773
Latina female	-0.413	0.110	-0.319	0.0726	0.638	0.232
SE	(0.825)	(1.140)	(1.043)	(1.072)	(1.262)	(1.188)
Obs.	11529	11529	11529	11705	11705	11705
Black male	0.648	-0.827	-1.350	0.00713	-1.429	-1.881
SE	(0.882)	(1.621)	(1.393)	(1.095)	(1.686)	(1.491)
Obs.	8896	8896	8896	9059	9059	9059
Black female	0.950	1.566	0.380	1.093	1.400	0.190
SE	(1.181)	(1.719)	(1.404)	(1.299)	(1.759)	(1.480)
Obs.	8698	8698	8698	8856	8856	8856
Asian male	1.947	-1.913	-0.794	2.382	-1.424	-0.297
SE	(1.915)	(2.799)	(2.543)	(1.990)	(2.779)	(2.547)
Obs.	5117	5117	5117	5320	5320	5320
Asian female	0.706	5.458+	7.095*	0.791	4.729	6.527*
SE	(1.869)	(3.022)	(2.876)	(2.037)	(3.052)	(2.928)
Obs.	4990	4990	4990	5234	5234	5234

8B. AP Calculus AB

	By start year			Force 100%		
	LB	UB	Rand	LB	UB	Rand
White male	-1.247	-1.986	-1.490	-1.832	-2.495	-2.304
SE	(1.381)	(2.580)	(1.972)	(1.964)	(2.809)	(2.360)
Obs.	2986	2986	2986	3042	3042	3042
White female	-0.491	-1.510	-0.299	-1.769	-3.155	-1.736
SE	(1.267)	(1.858)	(1.614)	(1.756)	(2.096)	(1.926)
Obs.	2948	2948	2948	3005	3005	3005
Latino male	-0.412	-0.626	0.439	-1.374	-1.529	-0.549
SE	(1.485)	(1.917)	(1.696)	(1.845)	(2.177)	(2.019)
Obs.	3104	3104	3104	3132	3132	3132
Latina female	0.313	-1.025	-0.335	-1.295	-2.607	-2.006
SE	(0.961)	(1.778)	(1.439)	(1.704)	(2.164)	(1.963)
Obs.	3079	3079	3079	3121	3121	3121
Black male	-0.207	-1.890	-0.756	-0.796	-2.476	-1.333
SE	(1.125)	(2.152)	(1.675)	(1.446)	(2.330)	(1.917)
Obs.	2613	2613	2613	2637	2637	2637
Black female	-0.740	-2.773	-0.923	-0.597	-2.927	-1.049
SE	(1.198)	(2.089)	(1.674)	(1.479)	(2.217)	(1.841)
Obs.	2599	2599	2599	2615	2615	2615
Asian male	-5.829+	-10.49+	-3.524	-7.795*	-12.35*	-5.322
SE	(3.313)	(5.878)	(5.086)	(3.479)	(5.874)	(5.136)
Obs.	2013	2013	2013	2067	2067	2067
Asian female	2.289	5.958	6.118	-0.792	3.419	3.325
SE	(2.736)	(6.208)	(5.336)	(3.415)	(6.325)	(5.519)
Obs.	1991	1991	1991	2060	2060	2060

8C. AP Calculus BC

	By start year			Force 100%		
	LB	UB	Rand	LB	UB	Rand
White male	0.639	0.384	-1.482	4.117	4.299	2.011
SE	(1.920)	(2.856)	(2.784)	(3.736)	(4.646)	(4.246)
Obs.	1027	1027	1027	1043	1043	1043
White female	-0.462	-1.781	-0.253	2.426	0.762	2.795
SE	(1.776)	(2.261)	(1.510)	(3.220)	(3.323)	(3.013)
Obs.	1015	1015	1015	1037	1037	1037
Latino male	0.702	0.836	0.745	1.336	1.507	1.402
SE	(0.475)	(1.517)	(1.339)	(2.030)	(2.448)	(2.349)
Obs.	1033	1033	1033	1041	1041	1041
Latina female	-0.364	0.637	0.254	3.144	4.013	3.706
SE	(1.165)	(2.062)	(1.957)	(2.705)	(3.076)	(3.069)
Obs.	1039	1039	1039	1047	1047	1047
Black male	0.388	1.380	1.205	0.429	1.420	1.244
SE	(0.510)	(2.040)	(1.494)	(0.518)	(2.043)	(1.502)
Obs.	967	967	967	973	973	973
Black female	-0.466	-2.173	-1.550	1.547	-0.222	0.422
SE	(0.535)	(1.537)	(1.463)	(2.088)	(2.468)	(2.448)
Obs.	968	968	968	974	974	974
Asian male	-4.300	-11.64	-9.402	-0.653	-7.465	-4.777
SE	(3.863)	(7.856)	(6.970)	(4.563)	(7.842)	(7.141)
Obs.	937	937	937	963	963	963
Asian female	2.547	8.513	4.647	7.656+	12.42+	9.434
SE	(2.357)	(6.756)	(6.002)	(3.920)	(6.992)	(6.527)
Obs.	906	906	906	930	930	930

8D. AP Statistics

	By start year			Force 100%		
	LB	UB	Rand	LB	UB	Rand
White male	0.374	0.1000	0.163	-0.343	-0.468	-0.612
SE	(1.257)	(2.019)	(1.865)	(1.761)	(2.308)	(2.196)
Obs.	2355	2355	2355	2412	2412	2412
White female	-0.303	-1.446	-1.262	-0.982	-2.479	-2.212
SE	(1.181)	(1.634)	(1.422)	(1.896)	(2.225)	(2.072)
Obs.	2332	2332	2332	2401	2401	2401
Latino male	-0.347	0.485	-1.225	0.802	1.522	-0.138
SE	(1.331)	(1.985)	(1.787)	(1.681)	(2.228)	(2.062)
Obs.	2366	2366	2366	2396	2396	2396
Latina female	-1.387	-1.720	-1.930	-1.395	-1.749	-1.924
SE	(0.877)	(1.482)	(1.267)	(1.582)	(1.955)	(1.811)
Obs.	2374	2374	2374	2416	2416	2416
Black male	-0.399	-2.240	-2.200	0.0453	-1.762	-1.722
SE	(0.991)	(2.451)	(1.778)	(1.684)	(2.756)	(2.203)
Obs.	2207	2207	2207	2245	2245	2245
Black female	1.344	0.271	0.413	2.225	1.012	1.176
SE	(1.293)	(2.568)	(2.096)	(1.861)	(2.804)	(2.434)
Obs.	2208	2208	2208	2232	2232	2232
Asian male	1.121	1.871	3.066	-0.605	-0.0506	1.053
SE	(2.524)	(4.117)	(3.960)	(2.706)	(4.235)	(4.080)
Obs.	1903	1903	1903	1952	1952	1952
Asian female	1.658	6.355	9.365*	0.512	4.914	7.912+
SE	(2.576)	(5.291)	(4.517)	(2.953)	(5.328)	(4.584)
Obs.	1790	1790	1790	1839	1839	1839