Working Paper Series
No. 2016-12

Affirmative Action And Human Capital Investment: Theory And Evidence From A Randomized Field Experiment

Christopher S. Cotton, Brent R. Hickman, and Joseph P. Price

April 1, 2016

Keywords: Affirmative Action, College Admissions, Field Experiment, Human Capital, All- Pay Auction

JEL Codes: J15, J24, C93, D82, D44.

Becker Friedman Institute
for Research in Economics

Contact:
773.702.5599
bfi@uchicago.edu
bfi.uchicago.edu

AFFIRMATIVE ACTION AND HUMAN CAPITAL INVESTMENT: THEORY AND EVIDENCE FROM A RANDOMIZED FIELD EXPERIMENT

CHRISTOPHER S. COTTON, BRENT R. HICKMAN†, AND JOSEPH P. PRICE

ABSTRACT. High-School human capital investment occurs within a competitive environment, and Affirmative Action (AA) shapes the relative competition between blacks and whites for admission to high-quality colleges. We present a theory of AA in university admissions and conduct a field experiment to mimic aspects of competition for college. We offer relative incentives to study math and track students' time on a mathematics website. AA increases average human capital investment and exam performance for its target beneficiaries by mitigating "discouragement effects" We also find evidence that AA can promote greater equality of market outcomes and narrow achievement gaps at the same time.

Date: **Original version: October 2013; Current version: April 2016**.

Key words and phrases. Affirmative Action, College Admissions, Field Experiment, Human Capital, All-Pay Auction.

JEL subject classification: J15, J24, C93, D82, D44.

Cotton: Department of Economics, Queen's University, Dunning Hall 230, Kingston, Ontario K7L 3N6; cotton@econ.queensu.ca; phone: (613) 533-2251.

†Hickman (corresponding author): Department of Economics, University of Chicago, 1126 E 59th Street, Chicago, IL 60637; hickmanbr@uchicago.edu; phone: (773) 702-3578.

Price: Department of Economics, Brigham Young University, 162 FOB, Provo, UT 84602; joe_price@byu.edu; phone: (801) 422-5296.

We greatly appreciate comments and helpful input from John List, Andrew Sweeting, Aaron Bodoh-Creed, John-Eric Humphreys, and participants of seminars at the University of Chicago, the 2013 meeting of the American Economic Association, the 2013 North American summer meetings of the Econometric Society, 2014 Canadian Public Economic Group conference, the 2015 Econometric Society World Congress, the 2015 European Meetings of the Economic Science Association, Brigham Young University, University of Missouri-Columbia, Queen's University, Ryerson University, SUNY Binghampton, University of British Columbia-Okanagan, University of Guelph, and London School of Economics. We also wish to acknowledge the outstanding research assistance of Joe Patten, who played a crucial role in executing this project. Hickman acknowledges support for the financial incentives in this experiment provided by a grant from the Social Sciences Division at the University of Chicago. Cotton is grateful for financial support provided by his position as the Jarislowsky-Deutsch Chair in Economic and Financial Policy at Queen's University. Cotton and Price gratefully acknowledge partial funding from the Spencer Foundation which supported an initial pilot study that lead to this research.

1. INTRODUCTION

Affirmative Action (AA) is the practice of granting preferential treatment to under-represented demographic groups when allocating contractual, employment, or educational opportunities. In the US it was first mandated by the Kennedy Administration in the 1960s, and has since been widely implemented in public procurement, education, and hiring. Today, AA is a pervasive fixture of North American university admissions, though it has generated much controversy.[1] AA is also widely implemented outside North America, from Malaysia to Northern Ireland, and in India where *Reservation Law*, a set of racial and ethnic quotas, is imposed by constitutional edict.

In the US university context, the rationale for AA does not stem from concerns over proactive discrimination. Rather, the rationale is that the university market is effectively a competition where black and Hispanic children are at a fundamental disadvantage to White and Asian children due to residual effects of past institutionalized discrimination. That is, on average they attend lower quality schools, are less affluent, have less educated parents, and have less access to other developmental inputs such as health care and tutoring.[2] In turn, AA is an attempt at compensating for the competitive disadvantage by giving special consideration for race. There is a substantial empirical literature studying the effects of AA, focusing mainly on its direct impact at the point when university admission outcomes are determined.[3]

However, the literature has largely ignored the implications of AA for student behavior *prior* to admissions, as they study, participate in extra curricular activities, and otherwise invest in human capital before applying to university. For example, consider a set of students vying for admission to universities which vary by quality. Each student has 24

[1]The US Supreme Court has deliberated on the legality of racial considerations in university admissions at least five times, including the cases of *Schuette v. Coalition to Defend Affirmative Action* (2014), *Fisher v. Texas* (2013), *Grutter v. Bollinger* (2003), *Gratz v. Bollinger* (2003), and *Regents of the University of California v. Bakke* (1978). At the time of this writing, *Fisher v. Texas* (2013) was under deliberation in a second-round hearing before the Court, after having been remanded back to the US 5th Circuit Court.

[2]Lyndon B. Johnson, Kennedy's successor, was the first American president to implement AA. In his 1965 commencement address at Howard University, Johnson articulated this idea as a motivation for AA: "*You do not take a person who, for years, has been hobbled by chains and liberate him, bring him up to the starting line of a race and then say, 'you are free to compete with all the others,' and still justly believe that you have been completely fair. Thus it is not enough just to open the gates of opportunity. All our citizens must have the ability to walk through those gates... To this end equal opportunity is essential, but not enough, not enough. Men and women of all races are born with the same range of abilities. But ability is not just the product of birth. Ability is stretched or stunted by the family that you live with, and the neighborhood you live in–by the school you go to and the poverty or the richness of your surroundings. It is the product of a hundred unseen forces playing upon the little infant, the child, and finally the man.*"

[3] William G. Bowen and Derek Bok [1998], Peter Arcidiacono [2005] and Jessica S. Howell [2010], have attempted estimation of counterfactual racial admissions profiles in a color-blind world. Linda Datcher Loury and David Garman [1995], Richard H. Sander [2004], Mark C. Long [2008], Jesse Rothstein and Albert H. Yoon [2008], and David L. Chambers, Timothy T. Clydesdale, William C. Kidder and Richard O. Lempert [2005], have estimated the impact of AA on graduation rates among UR minority groups.

hours in a day and would prefer to consume time as leisure, but time is a required input for human capital production. Since university seats are allocated according to measured human capital (HC) output (*e.g.*, through an exam score), each student will choose to invest some fraction of her leisure time based on her own ability, the distribution of her competitors, and her prospects for winning a favorable outcome. Since AA alters academic cutoffs required for admittance depending on one's demographic status, it in turn shapes a student's labor-leisure decision. Thus, some important questions arise: is there a trade-off between fostering diversity on university campuses and human capital investment during middle and high school? Does AA increase or decrease incentives for its targeted beneficiaries (under-represented minorities)? How does it impact incentives for other students? Finally, how does AA affect the racial achievement gap?

We begin the paper by presenting a model of human capital investment which is tailored toward capturing important aspects of competitive university admissions. The model allows for predictions about the direction in which AA should impact effort and performance distributions within each demographic group. There are two demographic groups of students, whose learning costs differ, in a competitive investment game running up to a matching market. A fixed distribution of university seats is awarded in rank-order fashion by a centralized admissions board. The model predicts that HC production effort will depend on a student's relative ability. Moreover, if the admissions board bases placement partially on demographic status, then, conditional on a fixed ability level, investment will also differ across groups. Specifically, AA tends to decrease HC accumulation by the highest ability minority students, while increasing investment in HC by intermediate and low ability individuals. Students in the non-minority group respond in a largely opposite way, with the highest ability students in this group increasing their effort and the lowest ability students decreasing it.

Overall, we find that AA has the potential to largely *increase* minority investment activity by mitigating so-called "discouragement effects" which arise from cost asymmetry across demographic groups. Discouragement is a well-known phenomenon in the literature on rank-order competition. The idea is fairly simple: holding fixed gross payoffs and one's own productivity type, if a dominating shift in the distribution of one's competitors occurs so that one's own type falls further behind, eventually effort/output will fall. This logic is independent of one's fixed productivity type, being an artifact of the competition one faces instead. Thus, if AA works to dampen discouragement effects for minorities by shifting their effective competition group, it may actually increase their willingness to re-allocate leisure time toward building HC instead. Moreover, this increase in effort happens on the region of the minority distribution where students are

most concentrated, meaning it generally leads to an increase in average investment activity. The sign of the average effect within the non-minority group is less clear-cut.

After presenting the theoretical framework, we develop a field experiment to investigate how AA shapes the labor-leisure trade-off in mathematics learning, as well as measured proficiency following a period of competitive HC investment. Our experimental design, modeled after the theoretical framework, structures a set of short-run incentives so as to mimic important aspects of university admissions. In our setup, cash prizes stand in for university placement outcomes to motivate test subjects to spend leisure time learning math.[4] Students are individually randomized into two treatments for our math competition, a "color-blind" control treatment, and an AA treatment. Our experiment involves paying middle-school-aged students based on their relative performance on the American Mathematics Competition 8 (AMC8), a national mathematics exam, in similar fashion as outcomes are determined on the university matching market using college entrance exams and high school grades. In order to create a clean test of theory on investment and discouragement effects, we use grade cohort as our demographic delimiter. This distinction mirrors some important racial disparities—our disadvantaged students (*e.g.*, 7^{th} graders) have received, on average, fewer inputs into their math education relative to their counterparts (*e.g.*, 8^{th} graders) with substantial overlap in the ability distribution notwithstanding—while filtering out other factors such as cultural differences or stereotype threat that could confound the effects we seek to investigate.

In the color blind treatment, students compete against others in both grade levels for prizes, with AMC8 score rankings determining each student's pay-off. In the AA (or "representative quota") treatment, we proportionally split the overall distribution of prizes and earmark subsets for each group, after which students compete only with other students from their same grade. In other words, a representative quota reserved a more favorable set of prizes for students in the lower grade, relative to what they would receive when competing with higher grade students. And, of course, the opposite is true for the set of prizes allocated to the higher grade students.

Because we are interested in how AA changes incentives to invest real effort into developing HC, a typical classroom experiment in which students are assigned to a treatment immediately before completing some task is not ideal. Instead, we worked with teachers to incorporate the experiment into the students' schooling over a course of two weeks.

[4]Our experiment covers short-run incentives spanning a matter of weeks. The longer pre-university investment period, which spans years, is of ultimate interest to policy makers, but is difficult to study in an environment where the researcher can experimentally create exogenous identifying variation. Thus, the current study focuses on how AA shapes competitive market forces which drive individual labor-leisure decisions and short-run human capital accumulation. What the magnitude of these forces might be when students are discounting over longer periods of time, with payoffs in mind that are much larger than those we can create, is beyond the scope of current research.

We assessed two in-class AMC8 exams separated by 2 weeks—a pre-exam to measure baseline proficiency and a post-exam to allocate prizes and measure progress—and we provided the students with a 10-day interim period in which they could study math problems ahead of the final AMC8 exam. During this period we provided the students access to a website with practice materials, and tracked their use of this site. By showing that students in the color blind and AA treatments access the site at different rates, we are the first experimental study to identify an effect of AA on real learning effort. We did this in a natural classroom and home learning environment, where the free-time/ study-time trade-offs were familiar to test subjects, and mimic scenarios they face when preparing for the competitive university admissions process during high school.

We find strong evidence that AA favorably alters labor-leisure trade-offs and measured proficiency by our disadvantaged group. Although the highest ability disadvantaged students did decrease performance as predicted by theory, this drop was modest, with the majority of disadvantaged students increasing learning time and test performance under AA. On average, the AA policy increased test scores among the disadvantaged group by $1/5$ of a standard deviation over the course of the experiment. While some portion of this increase may be due to greater focus or mental effort on the day of the exam, we also show that disadvantaged students in the AA treatment are more than twice as likely to use the practice website to prepare for the exam, spend almost triple the amount of total time, and attempt more practice questions relative to their counterparts in the control group. In this way, we are the first experimental paper to show that AA policy can increase real effort exerted by its beneficiaries to acquire HC. One might worry that these gains within the demographic group targeted by the policy come at the cost of commensurate weakening in incentives for other students, but the data do not support this concern. We find no evidence that these gains come at the expense of lower average investment or test scores for higher grade students.

Finally, the theory makes fine predictions at the individual level on how the signs and magnitudes of the incentive effects will vary, depending on one's underlying ability and demographic group. We also perform a test for these finer predictions, based on flexible, semi-nonparametric quantile functions estimated via the method of sieves. We find that the interquartile range (in ability space) within the disadvantaged group displays large and statistically significant improvement in achievement under AA, whereas there is a small but positive mass of the highest ability types within the same group who reduce their achievement in a statistically detectable way. We also find modest evidence for effects in the opposite direction within the advantaged group—a positive mass of high-ability types increase their achievement under AA while a positive mass of low-ability types reduce it—though the magnitudes of these effects are comparatively much smaller.

It is remarkable the degree to which these observed behavioral responses to our exogenous policy variation conform to the theory, both in their signs and relative magnitudes across the two groups. Disadvantaged students react strongly to AA and in a largely positive way because, on average, they begin with relatively high costs of competition and the policy therefore produces a substantial improvement in their standing within the rank-order investment competition. On the other hand, students from the advantaged group respond weakly because, despite seeing their gross payoffs decline due to AA, on average their learning costs were relatively low enough to maintain roughly the same performance under more anemic incentives. These two facts further suggest that AA will predominantly have the effect of narrowing demographic achievement gaps by incentivizing disadvantaged groups of students to "catch up," academically, to other groups with lower overall costs. A further analysis of achievement gap changes in our data lends further support to this conclusion.

We conclude our analysis with an exploration of how our limited experimental design might relate to broader human capital investment over longer periods, and in more diverse settings. First, we find that using a one-year difference in mathematics instruction (*e.g.*, 7^{th} graders vs. 8^{th} graders) understates the average disadvantage that underrepresented minority groups may face in practice. Specifically, we show that the test score gap between American Black and Hispanic children in 8^{th} grade, relative to White and Asian 8^{th} graders, is substantially larger than the gap between our 7^{th} and 8^{th} grade competitors in Utah. Second, we explore some back-of-the-envelope calibrations in order to compare our relatively small test subject payments to incentives associated with university admissions. Our experimental design seeks to engineer a short-term microcosm of the competitive human capital investment process running up to college. If we measure our experimental incentives as an effective hourly wage rate, and then divide estimates of the net present value of a college education by hours worked during 4 years of high school, we find that the short-term incentives within our study were smaller, but of a similar order of magnitude as plausible, long-run time-investment incentives.

The remainder of this paper has the following structure. Section 2 gives an overview of the previous literature and explains how our contribution relates to what has been done before. Section 3 presents a theoretical model of HC investment and university admissions to motivate our experimental design. We also explore model predictions to be tested in our study. Section 4 describes the structure of our field experiment in more detail. Section 5 presents and discusses our experimental results. Section 6 concludes with a discussion of our findings, the inherent limitations within our research design, and directions for future research. An appendix contains mathematical proofs and relevant extensions of the theoretical model. An accompanying online appendix

also presents additional technical details of our empirical analysis, as well as additional tables and graphs.

2. RELATED LITERATURE

The model which guides our experimental study is based on the general theoretical framework in Aaron Bodoh-Creed and Brent R. Hickman [2016]. Both papers depict a model of the university admissions market as an all-pay auction, in which students' bids are investments in HC, and prizes are seats at different universities. We present a simple model focusing on aspects of the market that are most readily testable using experimental methods. Bodoh-Creed and Hickman [2016], on the other hand, provide a general theoretical model and analysis, establishing conditions under which a rich college assignment model with AA can be represented in a tractable form. They begin by fleshing out a finite model, exploring how, when the number of agents and seats is large, the analytically unruly equilibrium is well approximated by a simpler setting where agents and university seats belong to a continuum set. For our purposes in this paper, it is sufficient to focus only on the continuum setting; interested readers are directed Bodoh-Creed and Hickman [2016] for the technical foundations behind this modeling choice.

They derive general results concerning the properties and design of market mechanisms for college placement. In this paper, our focus is more narrow in one way and more broad in another. First, we restrict attention to experimental comparisons between two very simple, canonical mechanisms: a pure rank-order (color-blind) rule, and a demographically representative quota. This allows us to assess the real-world applicability of the theoretical predictions on academic outcomes, and measure their associated magnitudes. Second, we use our novel experimental design with on-line monitoring of study effort to explore background phenomena—*e.g.*, labor-leisure trade-offs—on which the theory is silent. This allows us to confirm whether observed movements in outcomes induced by our treatment are associated with actual learning activities (*i.e.*, HC investment as the general framework assumes), or something less socially valuable such as exam-day effort or purely wasteful signaling. In that sense, this paper and Bodoh-Creed and Hickman [2016] may be seen as complementary to one another.

Our theoretical framework is also related to Stephen Coate and Glenn Loury [1993], which shows how an AA policy can make high value jobs more-easily attainable and can therefore reduce the incentives for minorities to undertake costly investment that make them more qualified for such jobs. The key difference between our model—with many-to-many matching—and theirs—with one-to-one matching—is that there is no scarcity of high value positions in Coate and Loury's framework. In our framework,

students compete with one another for a fixed set of heterogeneous university seats. We show that, although AA will decrease incentives for the highest ability minorities to invest, the opposite effect will predominate for most of them. Although our results are consistent with Coate and Loury [1993] for the highest ability minorities, our predictions are quite different for the overall minority population. We also test our predictions with the experiment, and find evidence largely consistent with our theoretical model.

Previous experimental work has also tested the link between AA and effort. Examples in the laboratory include Andrew Schotter and Keith Weigelt [1992], where asymmetry was exogenously imposed by researchers assigning cost functions to subjects, and Anat Bracha, Alma Cohen and Lynn Conell-Price [2015] which focused on gender-based asymmetry in quantitative problem solving. Caterina Calsamiglia, Jorg Franke and Pedro Rey-Biel [2013] conducted a related field experiment in which 10-13 year old children compete in Sudoku puzzles in a pair-wise (one-to-one) tournament, with asymmetry stemming from previous exposure to Sudoku. Each of these studies found that AA can increase average performance by disadvantaged players.

Our analysis offers a number of advantages over previous experimental studies. First, we include an investment period between assignment to a treatment group and our final exam. During this period we monitor student time usage at home in a non-invasive way. Ours is the first paper to do so, and it provides us with a window into individuals' labor-leisure trade-offs in actual learning, rather than focusing solely on in-class effort during a task. Our experimental design allows us to assess the impact of AA on both interim investment *and* final outcomes. Math learning in preparation for the AMC8 exam is close to learning that enables higher performance on college entrance exams, and therefore provides a relevant measure of HC investment. Second, our experimental competition more closely mirrors important aspects of actual college admissions markets. It is a many-to-many competitive matching contest (with participants numbering in the hundreds) having wide heterogeneity among both competitors and prize values. This heterogeneity is communicated to our test subjects in simple terms, similar to how US university applicants can compare their own standardized test score to mean scores of students who typically enroll in their most preferred universities. By working with test subjects' regular teachers, using materials that were already being used as teaching tools in their schools, and allowing for study choices at home, we are the first experiment to create a natural setting in which test subjects are making decisions similar to those which will lead to their ultimate college placement outcomes. Third, we intentionally chose an experimental design to eliminate concerns about stereotype threat—our test subjects were fairly homogeneous culturally and socioeconomically—so as to be able to measure magnitudes of incentive effects in a clean environment. Fourth, our format also

allows for overlap in ability distributions across demographic groups, rather than assuming no overlap as previous experimental work has done. This allows us to compare the performance distributions of students from asymmetric demographic groups to see whether different qualitative effects at different quantiles appear, as predicted by theory.

There is also a substantial empirical literature studying AA and its impact on college admissions. Bowen and Bok [1998] used student-level applications data to estimate the preference given to minority students by admissions officers at elite schools. Arcidiacono [2005] and Howell [2010] estimated structural models which adjust counterfactuals for changes in minority application behavior induced by policy shifts. A related vein of the literature focuses on mismatching, or the idea that AA may cause damage to black students by assigning them to academically demanding environments for which they are unprepared. Loury and Garman [1995] and Sander [2004] present evidence supporting mismatching. Other empirical work, such as Long [2008], Rothstein and Yoon [2008], and Chambers et al. [2005], suggest that some mismatching may occur, but its magnitude is relatively small and is outweighed by the benefits of placing blacks into higher quality institutions. Throughout this literature SAT—"Scholastic Aptitude Test"—scores are used as a proxy for student ability, and assumed to be fixed. However, students' incentives to invest in HC during middle and high school depend on admission policies they expect to face when applying to college. Test scores are therefore a function of student ability *and* market incentives induced by AA. Assuming SAT scores are independent of the prevailing admissions policy may bias counterfactual estimates. Our objective is to test this theory in a field experimental study of how students adjust their effort (influencing their test scores) in response to changes in AA policy.

This paper contributes to a small but growing literature recognizing that admissions policies shape incentives for pre-college HC accumulation. Bruno Ferman and Juliano Assuncao [2011] found evidence that test scores among black Brazilian high-school students decreased in response to an admissions quota at elite universities in Rio de Janeiro.[5] Hickman [2015] estimated a structural empirical model of the U.S. college admissions market based on the theoretical foundation of Bodoh-Creed and Hickman [2016]. His counterfactual analysis of admissions, investment, and welfare under alternative AA policies found evidence that AA increases the stock of minority HC overall,

[5] While seemingly contradictory to our findings, our model predicts that a quota which affects all points in the quality spectrum of universities will lead to a widening of the demographic achievement gap among the most talented students. Ferman and Assuncao [2011] estimated a negative overall effect of a policy change which applied only to top universities in Rio De Janeiro. Thus, the policy they study would have predominantly impacted high-ability black students who were at less of a disadvantage to begin with. Likewise, Brent R. Hickman [2015] estimated that AA negatively affects investment among a small mass of the highest ability minorities in the US, even though on average the effect is positive.

and provided further comparisons between color-blind admissions, representative quotas, and American-style preference-based AA (score bonuses).

Finally, our study is also related to others concerning performance pay in primary and secondary schools. Recent studies including Eric Bettinger [2012], Edwin Leuven, Hessel Oosterbeek and Bas van der Klaauw [2010] and Roland Fryer [2011] found mixed results about whether paying students can improve their effort and outcomes. Our experimental incentives are different from these in that we paid participants based on their *relative performance* rather than offering a fixed wage contract. Michael Kremer, Edward Miguel and Rebecca Thornton [2009] also studied a program awarding merit scholarships based on relative performance, and found evidence that students respond to changing financial incentives in a competitive environment. Our experimental design has two main differences from these studies: first, we were able to monitor outcomes *and* interim time inputs; and second, our focus is not on the effectiveness of pay-for-performance *per se*, but rather on *differences* between common relative performance incentive schemes. Finally, in a competitive setting some of a subject's behavior may be driven by an intrinsic desire to win, as in Christopher Cotton, Frank McIntyre and Joseph Price [2013] who presented evidence that simply framing a task as a contest can lead to better performance by some participants. In our study, we frame a common task as *two different contests*, which allows us to pick up on differences across alternative allocation mechanisms.

3. Theory

3.1. **MODEL.** Here we present a model of AA in competitive HC investment which will form the basis of our experimental design. It is a special case of Bodoh-Creed and Hickman [2016] who developed the technical foundations of the framework presented here. We model university admissions as a Bayesian game with a continuum of heterogeneous students of mass 1. Each student exerts effort developing human capital (HC). A student's relative HC output affects the quality of university she attends. The competition between students for higher-valued outcomes is a many-player all-pay contest for a continuum of heterogeneous university seats of mass 1. Students differ in their cost of HC accumulation, and university seats differ by quality.

Each student simultaneously chooses a level of HC accumulation, $h_i \geq \underline{h}$, where the quantity $\underline{h}$ represents a minimum pre-requisite for market participation, as when university applicants must first earn a high school diploma in order to be considered. Acquiring HC requires time—which could otherwise be consumed as leisure—and effort, and is therefore thought to be costly. Students differ by their background, abilities, access to help, and other resources that affect the rate at which time is converted into new HC. At the individual level, these factors are summarized by a parameter θ_i, which determines

the marginal costs of student i's HC accumulation. Investment entails a utility cost $c(h_i;\theta_i)$, which is strictly increasing and convex in the amount of output, or $c'(h_i;\theta_i) > 0$ and $c''(h_i;\theta_i) > 0$ for all $h_i \geq \underline{h}$. We also assume $\frac{\partial c}{\partial \theta} > 0$ and $\frac{\partial^2 c}{\partial h \partial \theta} > 0$ so that smaller θ implies a more productive student having lower costs for a given level h, and also a lower marginal cost of increasing output from h to $h + \Delta$. θ_i is privately known to each student, who views costs of her competitors as realizations of a random variable Θ.

There is a continuum of university seats of mass 1. We denote an individual seat by quality level s, and model the relative masses of different seats as following distribution F_S. Throughout the paper, we shall simplify discussion by assuming $F_S = Uniform[\underline{s}, \overline{s}]$ and $[\underline{s}, \overline{s}] = [0, 1]$ so that seat quality index and quantile rank are the same, with higher s indicating a more valuable seat. A student placing at seat s_i experiences benefit $u(s_i) = s_i$. Net utility of student i is

$$U(s_i, h_i; \theta_i) = s_i - c(h_i; \theta_i).^6 \tag{1}$$

For simplicity here, a student's gross match utility $u(\cdot)$ depends only indirectly on her HC, through determining the quality of university to which she is admitted. In the appendix, we explore an extension where HC directly enters match utility and students with higher HC benefit more from admittance to high quality universities. We show that the extended model still admits the qualitative patterns predicted by the version described here, which is simpler and isolates phenomena that we can directly test with our experimental design.[7]

3.1.1. ***Incorporating Demographics***. Each student observably belongs to one of two mutually exclusive demographic subgroups, $\mathcal{A}$ and $\mathcal{D}$, with $\delta \in (0, 1)$ being the mass of the latter. Since costs are privately known to each individual, students view their competitors' types in group $j = \mathcal{A}, \mathcal{D}$ as realizations of a random variable Θ_j with distribution $F_j(\theta)$ and density $f_j(\theta)$ which is strictly positive on a common support $[\underline{\theta}, \overline{\theta}]$. As convenient shorthand, we define the overall distribution and density as $F(\theta) \equiv \delta F_{\mathcal{D}}(\theta) + (1 - \delta) F_{\mathcal{A}}(\theta)$ and $f(\theta) \equiv \delta f_{\mathcal{D}}(\theta) + (1 - \delta) f_{\mathcal{A}}(\theta)$.

[6] We restrict attention to the case where cost type $\overline{\theta}$ chooses to acquire $\underline{h}$ units of human capital. In other words, we assume $U(\underline{s}, \underline{h}; \overline{\theta}) = U(\emptyset, 0; \overline{\theta})$, which guarantees that all students in our game weakly prefer to acquire the minimum HC and attend university. Those with $h_i \geq \underline{h}$ are allocated a seat based upon their relative HC. This means that the model is one of decisions on the intensive margin, being conditional on market participation. The question of how AA may affect university attendance decisions on the extensive margin is left for future research.

[7] Since it is difficult to directly experimentally test how h factors into gross utility, we instead focus on a simpler model in this section of the paper. We study HC accumulation in our experiment through direct observations of time investment into the learning process. For illustrative purposes though, in the appendix we present some examples where gross utility is Cobb-Douglas in s and h, $U(s_i, h_i; \theta_i) = s_i^{\alpha} h_i^{\beta} - \theta_i h_i$. Similar qualitative patterns emerge in response to AA, relative to color-blind admissions, as when $u = s_i$.

We assume group $\mathcal{A}$ is "advantaged" and group $\mathcal{D}$ is "disadvantaged" in that students from $\mathcal{D}$ tend to find HC production more costly, on average. Formally, we assume the distributions of $\Theta_{\mathcal{A}}$ and $\Theta_{\mathcal{D}}$ are ordered by *likelihood ratio (LR) dominance*, where

$$\frac{\partial \left(f_{\mathcal{D}}(\theta) / f_{\mathcal{A}}(\theta) \right)}{\partial \theta} > 0 \quad \text{for all} \quad \theta \in [\underline{\theta}, \overline{\theta}].$$

Intuitively, LR dominance is a strong form of first-order stochastic dominance which implies, among other things, that cost quantiles within $\mathcal{A}$ are all lower than the corresponding quantiles within $\mathcal{D}$. Although some "disadvantaged" students have low costs of acquiring HC, and some "advantaged" students have high costs, the average and median students in $\mathcal{D}$ have higher costs of producing new HC than their counterparts in $\mathcal{A}$, *etc*. This is how we operationalize the idea that, on average, disadvantaged students must invest more time and effort to overcome obstacles which are (imperfectly) correlated with their demographic status. In the context of race, for example, it is well-known that Black and Hispanic children in the United States tend to be less affluent and have less access to crucial childhood inputs like healthcare and high-quality public education; however, some still grow up in affluent environments which are more advantageous to childhood learning. Note also that LR dominance implies there exists a unique point, $\widetilde{\theta} \in \left(\underline{\theta}, \overline{\theta}\right)$, at which the densities $f_{\mathcal{A}}$ and $f_{\mathcal{D}}$ cross. This fact will become useful later when we explore comparative statics under different allocation mechanisms.

3.1.2. ***Seat Allocation Rules***. We consider two distinct university seat allocation rules. Let $H \sim G(H)$ denote the human capital output of a randomly selected student, and let $H_j \sim G_j(H_j)$ be similarly defined for a student from group $j = \mathcal{A}, \mathcal{D}$. The first seat allocation rule is a **color-blind (*CB*)** policy, which ignores demographics when allocating university seats. This rule assigns students to seats in pure rank-order fashion. The mechanism determines the quantile rank of h_i within the overall HC distribution and then matches student i to a seat at the corresponding quantile rank. For example, the 75$^{\text{th}}$ percentile student matches with the 75$^{\text{th}}$ percentile university, *etc*. Formally, under the color-blind rule, student i from group j receives seat assignment

$$S_j^{cb}(h_i) = S^{cb}(h_i) \equiv G(h_i), \quad j = \mathcal{A}, \mathcal{D}. \tag{2}$$

A student's allocation under the *CB* rule depends on her performance relative to *all* other students; thus, investment behavior will not vary by demographic status.

The second seat allocation rule is a **representative quota (*RQ*)** AA policy, which reserves a similar distribution of university seats for each group, $\mathcal{A}$ and $\mathcal{D}$, and then allocates the seats *within* each group by rank ordering. By "similar distribution" we mean that fraction δ of all seats at each point in the quality spectrum are earmarked *ex*

ante for group $\mathcal{D}$, thus splitting the set of university seats into two subsets having mass δ and $(1-\delta)$, but with both subsets still following the original quality distribution F_S.[8] Formally, under the RQ rule, student i receives seat

$$S_j^{rq}(h_i) \equiv G_j(h_i), \quad j = \mathcal{A}, \mathcal{D}. \tag{3}$$

A student's seat assignment under the RQ policy depends only on her performance relative to other members of her own demographic group. It splits the university admissions market into two separate contests, with identical distributions of prizes, but different distributions of player ability. From a student's perspective, the distinguishing characteristic of the RQ mechanism is that it alters the distribution of one's competitors, while leaving all other aspects of the contest the same as under a CB rule.

3.1.3. ***Solution Concept:*** The strategic environment presented above will form the basis for our experimental design in the next section. Bodoh-Creed and Hickman [2016] develop a general model of competitive HC investment which nests the framework presented here as a special case. They begin with a finite set of players and university seats, and demonstrate existence of a unique symmetric Bayes-Nash equilibrium. They then establish general conditions under which equilibria of the unweildy finite model can be well approximated by the solution to a decision problem where the finite sets on both sides of the matching market are replaced with continua.[9] This solution they refer to as an *approximate equilibrium*, and it lends analytic tractability to an otherwise complicated model. For the understanding we wish to develop in this paper, it suffices to limit discussion to the continuum representation of the model.

Our main goal in this section is to illustrate qualitative model predictions which are testable through experimental methods. The primary contribution of this paper derives from empirical comparisons of the distributions of effort and outcomes generated under the alternative policies CB and RQ. Thus, under each of these mechanisms we now characterize the approximate Bayes-Nash equilibrium of the game played between students who each choose how much HC to acquire prior to execution of the school assignment mechanism. To avoid tedious repetition, we abstract from the intricacies of the finite

[8] A subtle but crucial detail to note here is that our RQ policy calibrates δ to the *fraction of group $\mathcal{D}$ market participants*, which may not be the same as the mass of that group within the population at large. For example, South Africa mandates employment quotas for skilled professions, where quotas are pegged to the fraction of blacks in the overall population. One problem in the implementation of this law has been that not enough blacks exist within the skilled labor market (*e.g.*, individuals with pre-requisite post-secondary degrees) to fill mandated quotas. Such a rule would be a more extreme version of the AA policy studied in this paper, and may therefore entail different consequences.

[9] Alternatively, one may also interpret the continuum simplification as the basis of a behavioral strategy employed by a cognitively constrained agent. Rather than tracking probabilities over all the order statistics of large, complicated sets of competitors and seats, the agent may drastically reduce computational burden at little cost by pretending as if she was operating within a continuous world instead of a discrete one.

game and simply refer to our derivation of the approximate equilibrium as "the equilibrium." The interested reader is directed to Bodoh-Creed and Hickman [2016] for a full exposition of the technical foundations underlying the continuum model approximation.

3.2. **ANALYSIS.** Let $h^*(\theta)$ denote the common equilibrium investment function under the benchmark CB mechanism, and let $h^*_{\mathcal{A}}(\theta)$ and $h^*_{\mathcal{D}}(\theta)$ denote the group-specific investment functions under the alternative RQ mechanism. Moreover, let θ^*, $\theta^*_{\mathcal{A}}$, and $\theta^*_{\mathcal{D}}$ denote the relevant inverses, so that $\theta^* \equiv h^{*-1}$ and $\theta^*_j \equiv h^{*-1}_j$, $j = \mathcal{A}, \mathcal{D}$.

3.2.1. *CB **Allocations**.* Student i chooses HC to maximize her net payoff, given that other students play according to h^*. Equilibrium HC investment is strictly decreasing in θ, so equation (2) can be re-written as $S^{cb}(h) = 1 - F\left[\theta^*(h)\right]$, and student i's objective as

$$\max_{h_i \geq \underline{h}} \left\{ \left(1 - F\left[\theta^*(h_i)\right]\right) - c(h_i; \theta_i) \right\}. \tag{4}$$

Taking a first-order condition (FOC), we get $-f(\theta_i)\theta^{*\prime}(h_i) = c'(h_i; \theta_i)$. In equilibrium, $\theta^*(h_i) = \theta_i$ for all i, and since h^* is the inverse of θ^*, it follows that $h^{*\prime}(\theta_i) = 1/\theta^{*\prime}(h_i)$. Therefore, through a change of variables we can rearrange the FOC to get

$$h^{*\prime}(\theta_i) = \frac{-f(\theta_i)}{c'\left[h^*(\theta_i); \theta_i\right]}, \text{ with boundary condition } h^*(\overline{\theta}) = \underline{h}. \tag{5}$$

Given the assumptions on f and c, it is easy to see that h^* is strictly decreasing in θ. Equation (5) allows us to compare investment under two alternative cost distributions.

Theorem 1. *Consider two color blind contests,* 1 *and* 2*, which differ only by their cost distributions, and assume competition is more fierce under contest* 2 *in the sense that F_1 LR dominates F_2. Let $\tilde{\theta} \in (\underline{\theta}, \overline{\theta})$ denote the unique crossing point of the density functions where $f_1(\tilde{\theta}) = f_2(\tilde{\theta})$. There exists a unique interior crossing point $\ddot{\theta} \in (\underline{\theta}, \tilde{\theta})$, such that $h^*_1(\theta) < h^*_2(\theta)$ for all $\theta < \ddot{\theta}$ and $h^*_1(\theta) > h^*_2(\theta)$ for all $\theta > \ddot{\theta}$.*

A formal proof is left to the Appendix. The theorem provides useful insight into how competition shapes incentives. It says that, holding fixed the set of all universtiy seats, an increase in the degree of competition will cause the most able students to invest more aggressively, increasing their HC output, and less talented students to withdraw somewhat, decreasing their HC output. This second shift is a common feature of contests known as the *discouragement effect*. The intuitive idea is that since investment costs must be sunk before assignments are made to a set of fixed outcomes, then, holding one's own cost type θ fixed, if the distribution of competitors shifts so that one's quantile rank falls low enough, investment incentives fall. As we will see below, there is much insight to be had from Theorem 1 when comparing alternative allocation mechanisms.

3.2.2. *RQ **Allocations**.* Similarly as before, under a RQ equation (3) can be re-written as $S_j^{rq}(h) = 1 - F_j\left[\theta_j^*(h)\right]$, $j = \mathcal{A}, \mathcal{D}$, and the objective for student i from group j is now

$$\max_{h_i \geq \underline{h}} \left\{ \left(1 - F_j\left[\theta_j^*(h_i)\right]\right) - c(h_i; \theta_i) \right\}. \tag{6}$$

Note the main difference here is that the group-specific distribution enters decision making rather than the unconditional one. That leads to the following FOC

$$h_j^{*\prime}(\theta_i) = \frac{-f_j(\theta_i)}{c'\left[h_j^*(\theta_i); \theta_i\right]}, \text{ with boundary condition } h_j^*(\overline{\theta}) = \underline{h}, \quad j = \mathcal{D}, \mathcal{A}. \tag{7}$$

Recall our assumption that the random variable $\Theta_{\mathcal{D}}$ LR dominates $\Theta_{\mathcal{A}}$, or in other words, the ratio $f_{\mathcal{D}}(\theta)/f_{\mathcal{A}}(\theta)$ is strictly increasing in θ. This implies that Θ LR dominates $\Theta_{\mathcal{A}}$, and that $\Theta_{\mathcal{D}}$ LR dominates Θ as well. To see why, note that

$$\frac{f(\theta)}{f_{\mathcal{A}}(\theta)} = \frac{\delta f_{\mathcal{D}}(\theta) + (1-\delta) f_{\mathcal{A}}(\theta)}{f_{\mathcal{A}}(\theta)} = \delta \frac{f_{\mathcal{D}}(\theta)}{f_{\mathcal{A}}(\theta)} + (1-\delta),$$

from which it follows that $f(\theta)/f_{\mathcal{A}}(\theta)$ is strictly increasing in θ. Likewise,

$$\left(\frac{f_{\mathcal{D}}(\theta)}{f(\theta)}\right)^{-1} = \frac{\delta f_{\mathcal{D}}(\theta) + (1-\delta) f_{\mathcal{A}}(\theta)}{f_{\mathcal{D}}(\theta)} = \delta + (1-\delta)\frac{f_{\mathcal{A}}(\theta)}{f_{\mathcal{D}}(\theta)},$$

so $f_{\mathcal{D}}(\theta)/f(\theta)$ is strictly increasing in θ as well. Therefore, switching between allocation mechanism CB—where the competition group is all students—and mechanism RQ—where competition occurs only within one's own group—entails an effective LR dominance shift in the distribution of competitors while holding the distribution of seats fixed. Thus, Theorem 1 directly leads to the following corollary on HC accumulation under an RQ rule, relative to CB.

Corollary 2. *Assume $F_{\mathcal{D}}$ LR dominates $F_{\mathcal{A}}$ and let $\tilde{\theta} \in (\underline{\theta}, \overline{\theta})$ denote the unique crossing point of the cost densities where $f(\tilde{\theta}) = f_{\mathcal{A}}(\tilde{\theta}) = f_{\mathcal{D}}(\tilde{\theta})$. Then there exist crossing points $\ddot{\theta}_{\mathcal{A}}, \ddot{\theta}_{\mathcal{D}} \in (\underline{\theta}, \tilde{\theta})$, such that*

(i) $h_{\mathcal{D}}^*(\theta) < h^*(\theta)$ *for all* $\theta < \ddot{\theta}_{\mathcal{D}}$ *and* $h_{\mathcal{D}}^*(\theta) > h^*(\theta)$ *for all* $\theta > \ddot{\theta}_{\mathcal{D}}$, *and*
(ii) $h_{\mathcal{A}}^*(\theta) > h^*(\theta)$ *for all* $\theta < \ddot{\theta}_{\mathcal{A}}$ *and* $h_{\mathcal{A}}^*(\theta) < h^*(\theta)$ *for all* $\theta > \ddot{\theta}_{\mathcal{A}}$.

In words, the best and brightest students within group $\mathcal{D}$ *decrease* HC investment, as competition for the top university seats becomes less intense. At the same time, higher-cost individuals exert greater effort and *increase* HC accumulation, as RQ mitigates discouragement effects by placing them in a competition group where they are less far behind the curve. For the same reasoning, the opposite effects apply to group $\mathcal{A}$. The best and brightest *increase* HC investment under the RQ, as competition for the

top university seats becomes more intense, and others become discouraged as they find themselves further behind the curve in their new competition group.

Since theory predicts behavioral responses in opposite directions for students of different abilities within each group, the result above begs the question of which effect will dominate. Once again, Theorem 1 can shed some light on the answer. Using the density crossing point $\tilde{\theta}$, we can partition the cost support into two subsets,

$$T_{\mathcal{A}} \equiv (\underline{\theta}, \tilde{\theta}) \quad \text{and} \quad T_{\mathcal{D}} \equiv (\tilde{\theta}, \overline{\theta}),$$

which we refer to as the *typical cost sets* for each group. Intuitively, T_j is the region of the support where group j's density is strictly higher, and therefore it is overrepresented relative to its share in the overall population. LR dominance implies some interesting properties for these sets. Since both densities integrate to 1, it follows that

$$\int_{T_{\mathcal{A}}} [f_{\mathcal{A}}(\theta) - f_{\mathcal{D}}(\theta)] = \int_{T_{\mathcal{D}}} [f_{\mathcal{D}}(\theta) - f_{\mathcal{A}}(\theta)] .$$

In words, the degree of over-representation of group $\mathcal{D}$ in the high-cost set $T_{\mathcal{D}}$ is the same as the degree of over-representation of group $\mathcal{A}$ within the low-cost set $T_{\mathcal{A}}$.

Corollary 3. *Assume the same conditions as in Corollary 2. Then under the RQ policy (relative to the CB policy), typical disadvantaged students—that is, group $\mathcal{D}$ students with costs $\theta \in \{(\ddot{\theta}_{\mathcal{D}}, \tilde{\theta}] \cup T_{\mathcal{D}}\}$—exert higher effort and accumulate more HC. Moreover, if we define $\Delta : (\ddot{\theta}_{\mathcal{D}}, \tilde{\theta}] \cup T_{\mathcal{D}} \to \mathbb{R}$ as the difference on this set between group $\mathcal{D}$ investment under RQ versus CB, or*

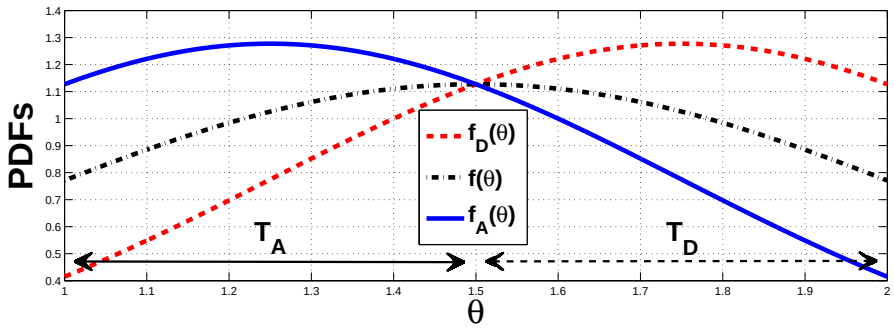
$$\Delta(\theta) \equiv \left(h^*_{\mathcal{D}}(\theta) - h^*(\theta)\right),$$

then $\Delta(\theta)$ is strictly positive and attains a maximum on the interval $(\ddot{\theta}_{\mathcal{D}}, \tilde{\theta}]$. Moreover, if investment costs are strictly convex in h, then $\Delta(\theta)$ attains its maximum on the open interval $(\ddot{\theta}_{\mathcal{D}}, \tilde{\theta})$.

Proof: The reasoning behind this result is simple. Corollary 2 directly implies the first part and that $\Delta(\cdot)$ tends toward zero at its endpoints but is strictly positive everywhere else. Moreover, equation (A.2) in the proof of Theorem 1 (see appendix) establishes that $\Delta'(\theta) < 0$ for all $\theta \in T_{\mathcal{D}}$, meaning the difference between the two investment functions becomes steadily wider as one moves toward the density crossing $\tilde{\theta}$ from the left. If costs are strictly convex, then (A.2) shows that $\Delta'(\theta) < 0$ for all $\theta \in T_{\mathcal{D}} \cup \tilde{\theta}$, so the maximum cannot occur at $\tilde{\theta}$. ■

Corollary 3 implies that, relative to their disadvantage in the overall population, a disproportionate share of group $\mathcal{D}$ students increase HC investment. To gain an appreciation for the strength of this result, the corollary also shows that, not only do we

FIGURE 1. NUMERICAL EXAMPLE:
Truncated Normal Cost Densities



see a positive effect for all group $\mathcal{D}$ cost types in their typical set, but the improved incentives extend well beyond $T_{\mathcal{D}}$ as well. In fact, the largest improvement of investment incentives by type (*i.e.*, where $\Delta(\cdot)$ attains its maximum) actually occurs on the interval $(\ddot{\theta}_{\mathcal{D}}, \tilde{\theta}]$. Thus, the result implies an increase of investment activity for a large fraction of the disadvantaged group. It is worth noting that analogous results to the three above can be proven in an extension of the model where h directly enters a student's gross match utility; these are Theorem 4 and Corollaries 5 and 6 in the Appendix.

On the other hand, the situation is less clear for the advantaged group. Although an analogous statement can be made—that all group $\mathcal{A}$ cost types $\theta \in \{(\ddot{\theta}_{\mathcal{A}}, \tilde{\theta}] \cup T_{\mathcal{D}}\}$ will reduce HC output under RQ—the statement is less informative, because investment also increases on part of the typical set $T_{\mathcal{A}}$ as well. To further illustrate the intuition of model predictions, we present three numerical examples, one below and two additional ones in the Appendix.

3.3. **NUMERICAL EXAMPLE.** Here, we explicitly solve the model using functional forms to illustrate the qualitative patterns predicted by theory. In the example, $\delta = 0.5$ and both $\Theta_{\mathcal{D}}$ and $\Theta_{\mathcal{A}}$ follow normal distributions truncated to a common support $[\underline{\theta}, \overline{\theta}] = [1, 2]$ with variance parameter $\sigma_{\mathcal{A}} = \sigma_{\mathcal{D}} = 0.5$. The mean parameters differ, with $\mu_{\mathcal{D}} = 1.75$ and $\mu_{\mathcal{A}} = 1.25$, which ensures that the distributions are ordered by LR dominance (see Figure 1). We specify costs as a linear function $c(h;\theta) = \theta h$, so that the maximum distance between investment under RQ and CB occurs at the boundary between $T_{\mathcal{A}}$ and $T_{\mathcal{D}}$. Finally, we set $\underline{h} = 0.5$; this makes type $\overline{\theta}$ indifferent to market participation when the outside option is normalized to a value of zero.

Figures 2 and 3 depict a comparison of the RQ investment functions for each group (solid and dashed lines), as well as the common investment function for both groups under a CB allocation rule (dotted line). For each group there is crossing point of the investment functions, with the upper bounds of the HC distributions being different by

FIGURE 2. NUMERICAL EXAMPLE:
Investment, *CB* vs AA

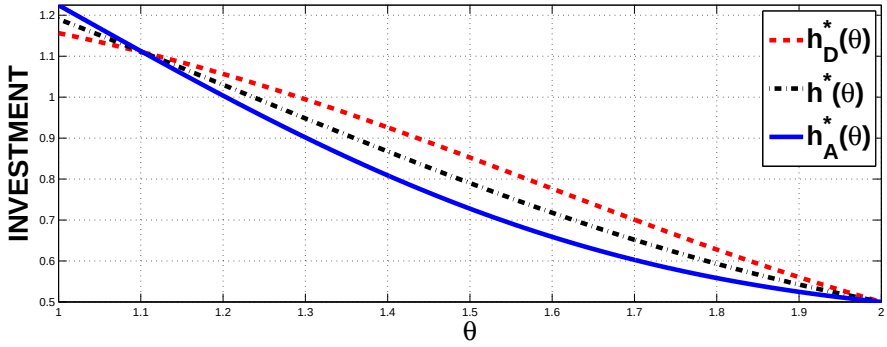
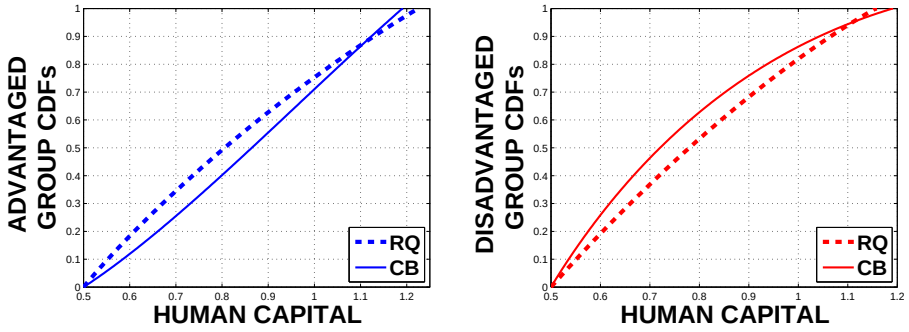


FIGURE 3. NUMERICAL EXAMPLE:
HC Distributions, *CB* vs AA



seat allocation rule. For the disadvantaged group, a positive mass of the top students reduce investment, while middle- and high-cost students increase it. Intuitively, the policy aids the top students from $\mathcal{D}$, but since they were already placing close to the upper bound their outcomes cannot be commensurately improved and they rationally reduce effort. For other students in $\mathcal{D}$, the policy alleviates discouragement effects by placing them in a competition group where their own type is not as far behind, making them more competitive for higher quality outcomes. In turn, they respond with an increased willingness to engage in costly investment. Similar logic holds for group $\mathcal{A}$, but in reverse. A key aspect of our experiment will be to engineer this sort of counterfactual scenario, where similar ability students find themselves facing different levels of competition for the same set of outcomes. Within this experimental counterfactual we test whether AA leads to changes in performance distributions. We also investigate intermediate inputs behind investment costs depicted in the model: time and effort.

4. EXPERIMENTAL DESIGN

We built our incentives and learning exercise around the American Mathematics Competition 8 (AMC8) exam, sponsored by the Mathematical Association of America, for students in 8^{th} grade and below. It consists of 25 multiple choice questions (five choices each) in 40 minutes, and the questions become progressively more difficult from start to finish. The AMC8 is an early precursor to math competitions such as the International Mathematical Olympiad and the Putnam. The AMC8 website explains, "[it] provides an opportunity to apply concepts [to] high level to problems which... are designed to challenge and offer problem solving experiences beyond those provided in... junior high school."

4.1. **SAMPLE POPULATION.**

Our total sample includes 992 middle school and junior high students from 10 schools in Utah County, Utah, including both charter schools and regular public schools. Within these schools, participation in our study was at the classroom level. Academically and socioeconomically, our sample population was fairly average within the US.[10] Most of our partner schools had previously participated in the AMC8 before partnering with us for this study.

Our partner schools exhibited a high degree of racial and cultural homogeneity. Less than 1% of students were black and only 7.5% were Hispanic, compared to nationwide averages of 15.25% and 22.2% for blacks and Hispanics, respectively. All schools in this study serve suburban populations. However, given that the goal of this paper is to cleanly test theory of incentives, the demographic homogeneity in our sample may be an advantage. We chose grade level as our demographic delimiter for groups $\mathcal{A}$ and $\mathcal{D}$, which ensures that they only differ in observable ways (*i.e.*, age and grade) but are otherwise similar. Focusing on a relatively homogeneous subject pool allows us to largely rule out other cultural/behavioral phenomena (such as stereotype threat) which might confound the pure incentive effects arising from our exogenous policy variation. At the same time, however, we must recognize that AA policies typically apply in real world settings with a number of confounding factors which we minimize in our

[10]Our test subjects are drawn from a population which is somewhat more affluent than the rest of the country, but academically comparable. Based on figures from the National Center for Education Statistics, we estimate a median household income of $59,800 for the sample population, compared to a nationwide median of $53,046 (information downloaded from http://nces.ed.gov/surveys/sdds/framework/tables.aspx). In 2012, approximately 33% of test subjects were eligible for free or reduced-price lunch, compared to a national average of 48%. That year our partner schools housing 5^{th} and 6^{th} grade subjects (20% of our sample pool) performed significantly better than other Utah schools in terms of meeting state math standards (approximately 91% vs 76%), while schools housing our 7^{th} and 8^{th} graders (80% of our sample pool) performed slightly worse than other Utah schools (81% vs 83% meeting state standards). Utah is a state which typically ranks at or near the median for nationally measured academic outcomes such as NAEP scores and enrollment rates in Advanced Placement programs.

experiment. We discuss this point further in Section 6 where we compare achievement gaps within our sample pool to race gaps within the US population.

4.2. **TREATMENT GROUPS AND INCENTIVES.** Participants in our study first took a practice AMC8 test from a previous year. We used this as a baseline measure of each student's ability. Individuals were randomized into either a control group—with a "color-blind" competition—or an AA treatment—with a representative quota competition. Following the notational convention from the previous section, we will henceforth refer to these treatments as CB and RQ, respectively. For treatment CB, we ran competitions involving students in two adjacent grades; that is, 7^{th} and 8^{th} graders competing together, and 5^{th} and 6^{th} graders competing together. For each of the two age cohort pairings, students in the lower grades (5^{th} or 7^{th}) are henceforth referred to as the "disadvantaged" demographic $\mathcal{D}$, and students in the higher grades (6^{th} or 8^{th}) are referred to as the "advantaged" demographic group $\mathcal{A}$, since the latter are one year older and have received one more year of mathematics education on average.[11] For treatment Q, students competed only within their own grade level, but for a proportionally equivalent set of prizes (relative to CB), as described below. We ran separate competitions for 5^{th}, 6^{th}, 7^{th}, and 8^{th} graders.

The top 30 percent of students within each competition group received cash prizes, which were uniformly distributed between \$4 and \$34 in \$2 increments. Prizes were awarded assortatively within competition groups, according to final exam scores. For example, 7^{th} grade students in treatment CB needed to score within the top 30 percent of all 7^{th} and 8^{th} grade students in their treatment to receive a prize. In treatment Q, students competed against others in their own grade only, but for a representative set of prizes. More specifically, we began with the same aggregate prize distribution as for treatment CB, and then earmarked prizes at each different level in proportion to the mass of lower-grade students in each age cohort pairing (note that the mass of groups $\mathcal{A}$ and $\mathcal{D}$ were also identical across treatments CB and Q). This ensured that the moments of the prize distribution (including the 70% mass of zeros) were the same across all competition groups, with each one vying for the same number and variety of prizes on a per capita basis. For example, 7^{th} grade students in treatment Q only had to score within the top 30 percent of 7^{th} graders in their treatment to receive a prize. Moreover, their distribution of prizes, conditional on winning something, was the same as for 8^{th}

[11] As we discuss in Section 6, the difference in average preparation between our disadvantaged and advantaged groups likely understates difference between black and white students at the same age in the US. Using national figures from the NAEP exam, we estimate that the 7^{th}–8^{th} grade gap in our sample pool is roughly 60% of the gap between black and white 8^{th} graders in terms of math proficiency.

graders in treatment Q, and also the same as for all $7^{th}/8^{th}$ grade students in treatment CB. Thus, for an advantaged or disadvantaged student of a given ability level, the only difference across the two treatments is the distribution of one's competitors.

Each student received an information sheet describing their assigned group, how many students from which grade(s) they would be competing against, and the score distribution within their group based on the practice test. Students received their own practice score back at the same time so they could see where they fit within their competition group. The sheet also contained a table describing the prize structure. We printed information relative to each competition group on a different color of paper so that students could visually see in their classroom that roughly half of the students were assigned to each treatment. Altogether, there were six different groups: four groups for the quota treatment (one for each grade) and two groups for the neutral treatment (one for $5^{th}/6^{th}$ grade and one for $7^{th}/8^{th}$ grade). In a web appendix, we provide an example of the information sheet given to each group.[12]

4.3. **MATH LEARNING WEBSITE.** At the bottom of the information sheet was the url of a website we set up with practice problems drawn from five past AMC8 exams. At 25 questions each, this made for 125 total practice problems covering six different math subjects: Arithmetic, Algebra, Combinatorics, Geometry, Logic, and Probability. Problems were divided into a set of 31 total quizzes. Each year, the 25 AMC8 exam questions are numbered in increasing order of difficulty. For each of the previous five year's exams, the website included one quiz covering problems 1-10, a second quiz covering problems 11-20, and a third covering problems 21-25. Test subjects were notified that each grouping of 3 same-year quizzes were ordered by their difficulty level. We also arranged this same battery of math problems into an additional set of 16 quizzes, each containing 5 subject-specific math problems. These subject quizzes were also ordered by their difficulty level.

Students could attempt each quiz as many times as they liked, or move on to additional materials they had not yet tried. After completing each quiz, our software displayed an instructional page which reported to each student her score, the correct answers for each problem, and step-by-step solutions published by the developers of the AMC8. Students were provided with a web page that contained links for all of the quizzes we offered, but in order to access the quizzes, they had to input their name, grade, and school on the first page of the web form. This allowed us to track online activities for each quiz session, including which students visited the website, how many

[12]Copies of the information sheets given to test subjects are available for download at http://home.uchicago.edu/~hickmanbr/uploads/CHP2014_WEB_APPENDIX.zip

different subjects they tried, how much time they spent, how many questions they attempted, what they answered on each attempt at each question, and how much time they spent viewing the instructional page.

Within each quiz, questions were separated on different web pages in blocks of 3, 4, or 5 questions per page, and the instructional page at the end displayed feedback for all questions on a single page.[13] Time on our website was measured at the page level, meaning that we got a time measure for blocks of either 3, 4, or 5 questions. In order to convert this information into a time spent per question measure, we divided each block-level time observation by the number of questions within that block. Instructional page times for 10-question quizzes were split into two observations a piece by dividing by two in order to make them comparable to 5-question instructional page view times.

4.3.1. ***Time Measurement.*** One difficulty arose in that there were clear instances where students left the website in the middle of a quiz for several hours or more. To adjust for this problem we chose truncation points on the domains of time per question and instructional page view time, and we replaced each observation above that point with the appropriate student-specific censored mean.[14] In selecting our truncation point we looked for occurrences of "holes" in the support of the distribution of times per question.[15] For our time per question data, this leads to a truncation point of 26.14 min/question (the 99.35^{th} percentile), and for instructional page views, 108.39 min/page view (*i.e.*, 21.68 min/solution, or the 98^{th} percentile). In the Appendix we display a histogram of (uncensored) time per question and instructional page view times.

At the end of the day, the time monitoring capability on our website is not perfect, and it is impossible to directly observe work stoppages in the middle of a quiz question. In particular, it may still be the case that smaller work stoppages occur below the truncation points. Therefore, in terms of time per question we are effectively interpreting work stoppages of less than 27 minutes as time which comes at a positive cost to the child. We argue that 27 minutes is a reasonable truncation point for several reasons. First,

[13]Each 10-question quiz was broken into three pages with 3 questions on page one, 3 questions on page two and 4 questions on page three. Each 5-question quiz displayed all 5 problems on a single page.

[14]To illustrate this rule, suppose that Tommy attempted three 5-question quizzes for a total of 15 questions. Suppose further that we observed times of 5 minutes each for seven questions, 15 minutes each for another seven questions, and 2000 minutes for the last one. Then if the truncation point were, say 30 minutes, the last observation of 2000 is replaced by Tommy's idiosyncratic censored mean time of 10 minutes (for all other questions he attempted). As a robustness check, we also ran our analysis by simply dropping truncated observations instead, and results are very similar to those we present below.

[15]More specifically, a hole in the distribution support was defined as the minimum point at which a full-support condition fails, which we estimated as a point where a kernel-smoothed density estimate hit zero. The idea behind this rule is that if the type distribution has full support, then the distribution of times per question should have full support as well since the choice of how much time to spend is continuous. For a more complete description of our truncation point selection rule, see the Appendix.

work stoppages for our uncensored time observations (most of which were less than 10 minutes) would serve as a poor substitute for longer, unbroken leisure spells. Second, since this potential problem is the same across both treatment groups, there is no reason to believe that our results are being aided by it.

Third, the AMC8 contains fairly challenging material that may require significant time inputs for some students. Table 2 in the next section displays the mean and variance of time spent per question attempt, using the censored sample of times. The most difficult subject appears to be combinatorics, with a mean time of 2.839 minutes and a standard deviation of 3.532. Given that the censored distribution of time per question is right skewed, and 10 minutes (the 98^{th} percentile of the un-censored sample) is roughly two standard deviations above the mean for combinatorics, it is plausible that roughly 1.5% of our sample could exist on the interval between 10 minutes and 26 minutes.

As for instructional page view times, it is informative to consider a particular student whom we will rename "Kate" to protect her identity. Kate, a 7^{th} grader in the quota treatment, spent more time than anyone else on the website (after time adjustments), averaging roughly 53 minutes per day during the study period. At 55 question attempts, Kate was also above the 94^{th} percentile on that dimension as well. She attempted 11 quizzes of 5 questions each, averaging 2.24 minutes per question attempt and 7.49 minutes per question on the instructional page with solutions. She spent an hour or more on 5 of her quiz attempts, each time spending the majority of her time on the instructional page. For each of Kate's 11 quiz attempts we see that she clicked through to the quiz termination screen herself. Kate is an example of a student who displayed consistent patterns of substantial time inputs into many of the quizzes she took—particularly on the instructional page—while never having left a quiz session open overnight.

4.4. **TESTING.** Students took the actual AMC8 test in their regular classrooms, under all of the normal conditions in which students around the country take the AMC8. Most of the students in our study attended schools where participation in the AMC8 was already being offered to students by their teachers, but on a voluntary basis. The schools that cooperated in this study administered the test to all students within each participating classroom on an opt-out basis, so that all students participated in the study, except those whose parents proactively signed and returned an opt-out form. The study involved two in-class exam sessions: the practice test was the AMC8 exam for the previous year, and the final exam was the AMC8 for the current year. The cash prizes were delivered to each school shortly after the final exam, and handed out to each student in an envelope. The outcome measures that we use in the next section include both the

effort-based measures with website data, as well as a performance-based measure using students' scores on the AMC8.

5. EXPERIMENTAL RESULTS

The theoretical model predicts that some students in each demographic group will increase their effort and HC accumulation under AA, while others in the same group will decrease their efforts. Our empirical investigation takes the next step by testing for the average effects of AA on effort and performance by demographic group and estimating magnitudes. We also use a non-parametric analysis to consider the effects of AA on the performance distributions and achievement gaps.

5.1. **DESCRIPTIVE STATISTICS.**

5.1.1. ***Pre- and Post-Exams***. Table 1 contains descriptive statistics on AMC8 exam scores. Roughly three quarters of our sample were $7^{th}/8^{th}$ graders. The difference between these and our $5^{th}/6^{th}$ grade test subjects is that the latter all came from accelerated classes, whereas the former are representative of the overall student body within their schools. This difference is born out in the data: while 8^{th} grade students did best on the pre-test with an average score of 9.04, 6^{th} graders as a group came in second at 8.12 on average. 7^{th} and 5^{th} grade average pre-test scores are close, at 7.58 and 7.19, respectively.[16]

We have also broken down test scores by two groups that we refer to as ***investors***—students who logged on to our website at least once during the study period—and ***non-investors***—those who did not. Students who did better on the pre-test were more likely to be investors (*i.e.*, they were more likely to use the math practice website), although some students who did not do as well also chose to be investors, and many students who did quite well on the pre-test chose to be non-investors. For the group of investors, we also present summary statistics concerning their activities on the website. Investors' times ranged between a few minutes and 8.92 hours, or an average of about 53 minutes per day over the study period. Number of questions attempted ranged between 1 and 120, with mean and standard deviation of roughly 19 and 23, respectively. *Subjects* represents the number of different subject categories a student attempted, using the subject-specific quizzes, being about two on average.

[16]The national AMC8 population in 2013 (see https://amc-reg.maa.org/reports/generalreports.aspx) had mean and median of 10.69 and 10 out of 25, with standard deviation of 4.44. These figures illustrate the difficulty of the exam. The AMC8 is predominantly administered through opt-in participation, whereas our experiment was on an opt-out basis. This accounts for the lower mean among our sample pool.

TABLE 1. STUDENT DESCRIPTIVE STATISTICS

	Mean	Median	Std. Dev.	N
Pre-Exam Scores				
All	8.45	8	2.90	992
5th Grade	7.19	7	2.39	48
6th Grade	8.12	8	2.47	155
7th Grade	7.58	7	2.84	275
8th Grade	9.04	9	2.82	396
Investors	9.46	10	3.19	118
Non-Investors	8.32	8	2.83	874
Final Exam Scores				
All	8.64	8	2.88	895
5th Grade	7.40	7	2.22	42
6th Grade	9.17	9	2.82	133
7th Grade	8.12	8	2.90	233
8th Grade	8.75	9	2.80	374
Investors	9.20	9	3.06	113
Non-Investors	8.56	8	2.84	782
Human Capital Investment (Investors Only)				
Total Time	43.65	26.85	64.65	118
Problem Solving Time	32.99	19.31	41.43	118
Instructional Time	10.66	3.37	38.85	118
Questions	18.89	10.00	22.53	118
Subjects	1.94	1.00	1.43	118

Notes: All time figures are post-censoring as described in Section 4.3 and quoted in minute units. *Investors* are defined as students who logged on to the math learning website at least once during the investment period. *Non-Investors* are those who did not.

TABLE 2. Time Per Question Attempt by Subject

Subject	**Censored Mean Minutes Per Question**	**Std. Dev. (minutes)**	Mean + 2×Std. Dev.
Algebra	2.422	2.563	7.548
Arithmetic	1.398	1.238	3.874
Combinatorics	2.839	3.532	9.903
Geometry	2.183	2.577	7.337
Logic	1.807	1.742	5.291
Probability	1.996	1.137	4.27

TABLE 3. EFFORT AND PERFORMANCE BY TREATMENT

	Investment				**Performance**
	Used Website	# Subjects Attempted	Total Time	# Questions Attempted	Final Exam Score
Quota	0.154	0.284	6.634	2.729	8.680
Std. Err.	(0.015)	(0.037)	(1.216)	(0.456)	(0.139)
Neutral	0.088	0.189	3.932	1.817	8.604
Std. Err.	(0.014)	(0.035)	(1.149)	(0.431)	(0.133)
N	992	992	992	992	895

Notes: Each cell provides the mean of the measure listed in each column. Standard errors are provided in brackets. Estimates under each of the four effort variables are intended to capture the effect of a treatment on human capital investment for the total study population, and are therefore averaged over both investors and non-investors.

5.1.2. ***Subject-Specific Practice Question times.*** Table 2 displays the mean and variance of time spent per question attempt on our website, using the censored sample of non-truncated times (once again, for investors only). Some math subjects appeared more challenging in terms of the time students took to solve problems. The most difficult subject was combinatorics, with a mean time per question attempt of 2.839 minutes and a standard deviation of 3.532. The least difficult subject appeared to be arithmetic, with mean and standard deviation of 1.398 and 1.238, respectively.

5.2. **EMPIRICAL ANALYSIS.**

5.2.1. ***Testing Overall Differences by Treatment.*** Tables 3 and 4 investigate the effect of a quota on the overall population, including both advantaged *and* disadvantaged groups.

The first column of Table 3 displays the mean of a binary variable, being the fraction of investors from each treatment group (*i.e.*, subjects who logged on to our website at least once to practice math). As for the other investment variables, the reader should keep in mind that Tables 3 – 5 aim to measure a treatment effect of a policy on an entire group, including both the intensive and extensive margins of investment. This is why the effort numbers in Table 3 and afterward appear small: they are averaged over both investors *and* non-investors. The results indicate that subjects in the quota treatment, including students from all age groups, were 75% more likely to have visited the website than students in the color-blind treatment. They also tried out more subjects, spent more time on the website and answered more questions. Table 3 indicates that students in both treatments scored roughly the same on the final exam. This is allowed for by the theory, where predictions for the overall population are qualitatively ambiguous, but later on we will see a different story when we separate these measures by demographic group.

Table 4 provides statistical tests for the raw differences displayed in table 3. In the first row we run a simple regression using a dummy for the quota treatment, meaning it represents the experimental difference between an RQ rule and CB allocations at the population level (*i.e.*, including both demographic groups). Each cell in the table represents a separate regression with the outcome variable labeled in the column header. We report the point estimate and p-value for a test of the hypothesis that there is no difference by treatment group. From the table we see strong evidence that AA increases the fraction of students willing to invest at least some time. We also see evidence that it induces them to experiment with more subjects, as well as increase the total time invested and number of questions attempted. Although these last two differences are only marginally significant, the estimated magnitudes are large, with quota students logging an estimated 57% and 70% more inputs of time and question attempts, respectively. Additional controls (pre-test score and/or school fixed effects) are added in the bottom two rows as a check on the effectiveness of our randomization. It seems to have worked well, as adding these additional 11 variables caused no significant shifts in point estimates.

5.2.2. ***Testing Differences by Treatment Within Demographic Groups***. Recall that the theory allows for AA to have differential effects by ability (*i.e.*, cost of competition) and demographic group. In Table 5, we add a demographic dummy to investigate this claim. The first five columns each present estimates for a regression equation of the form

$$Outcome = \beta_0 + \beta_1 Quota + \beta_2 Advantaged * Quota + \beta_3 Advantaged + \beta_4 Pre\text{-}Test + \varepsilon,$$

where $Quota$ is a dummy for treatment status, $Advantaged$ is a demographic dummy, $Pre\text{-}Test$ is a student's pre-test score in standard deviation units, and the specific $Outcome$

TABLE 4. TESTING DIFFERENCES BY TREATMENT

	Investment				**Performance**
	Used Website	# Subjects Attempted	Total Time	# Questions Attempted	Final Exam Score
Quota − Neutral	0.066***	0.095*	2.701	0.912	0.076
P-Value:	*[0.001]*	*[0.061]*	*[0.107]*	*[0.146]*	*[0.693]*
(*Controls: none*)					
Quota − Neutral	0.065***	0.093*	2.650	0.884	0.097
P-Value:	*[0.002]*	*[0.067]*	*[0.113]*	*[0.158]*	*[0.576]*
(*Controls: pre-test scores*)					
Quota − Neutral	0.058***	0.078	2.404	0.773	0.164
P-Value:	*[0.005]*	*[0.130]*	*[0.158]*	*[0.224]*	*[0.346]*
(*Controls: pre-test scores, school FEs*)					
N	992	992	992	992	895

Notes: Each cell represents a separate regression. The number reported is the coefficient for the quota treatment. Row 1 includes no controls and provides a statistical test of the differences in Table 1. Row 2 includes control for practice test score. Row 3 includes school fixed effects. P-values for a two-sided test of the null hypothesis of zero difference are italicized and in brackets. Estimates under each of the four effort variables are intended to capture the effect of a treatment on human capital investment for the total study population, and are therefore averaged over both investors and non-investors.

variable is labeled in the column header.[17] With the inclusion of the interaction term $Advantaged * Quota$, the coefficient β_1 represents the average effect of AA specifically on the disadvantaged group. The effect of the policy on the advantaged group is represented by the sum $\beta_1 + \beta_2$. The last column in the table moves pre-test score to the left-hand side, to estimate an alternative measure of the treatment on math improvement:

$$FinalExamScore - Pre\text{-}TestScore = \beta_0 + \beta_1 Quota + \beta_2 Advantaged * Quota + \beta_3 Advantaged + \varepsilon.$$

For completeness, all regressions include controls for school-level fixed effects, and for the primary effects of interest we report p-values in brackets.

[17]One potential concern with including pre-test score is that it provides a noisy measure of initial human capital, and may therefore introduce an attenuation bias. As a robustness check we re-estimated the regressions in Table 5, omitting *Pre-Test*, and nearly all coefficient estimates and standard errors—except for the ones connected to β_3, the multiplier on higher-grade cohort status—remained virtually unchanged. See Table 8 in the online appendix for coefficient estimates with *Pre-Test* omitted as a control.

TABLE 5. TESTING DIFFERENCES BY DEMOGRAPHICS AND TREATMENT

	Investment				**Performance**	
	Used Website	# Subjects Attempted	Total Time	# Questions Attempted	Final Exam Score	Exam Score Change
Constant ($\widehat{\beta}_0$)	0.073***	0.144**	3.033	1.316*	8.147***	0.096
Std. Err.	(0.024)	(0.059)	(1.982)	(0.741)	(0.209)	(0.240)
Quota ($\widehat{\beta}_1$)	0.087***	0.146*	5.517**	1.312	0.624**	0.591*
Std. Err.	(0.033)	(0.083)	(2.757)	(1.030)	(0.287)	(0.330)
P-Value:	*[0.009]*	*[0.077]*	*[0.046]*	*[0.203]*	*[0.030]*	*[0.074]*
Advantaged $*$ *Quota* ($\widehat{\beta}_2$)	-0.047	-0.111	-5.034	-0.866	-0.712**	-0.546
Std. Err.	(0.042)	(0.105)	(3.506)	(1.310)	(0.360)	(0.415)
Advantaged ($\widehat{\beta}_3$)	0.028	0.083	1.613	0.877	0.488*	-0.324
Std. Err.	(0.030)	(0.076)	(2.545)	(0.951)	(0.264)	(0.300)
Pre-Test (*standardized*) ($\widehat{\beta}_4$)	0.029***	0.045*	1.009	0.572*	1.280***	*N/A*
Std. Err.	(0.011)	(0.027)	(0.893)	(0.334)	(0.092)	*N/A*
School Fixed Effects	yes	yes	yes	yes	yes	yes
N	992	992	992	992	895	895
Additional Test: Effect of Quota on Advantaged Group						
$\widehat{\beta}_1 + \widehat{\beta}_2$	0.040	0.035	0.483	0.446	-0.089	0.045**
P-Value:	*[0.123]*	*[0.586]*	*[0.823]*	*[0.581]*	*[0.684]*	*[0.030]*

Notes: Each column is a separate regression. Advantaged is an indicator variable for whether the student is a 6^{th} or 8^{th} grader (the older group in each school type). We also include each student's standardized pre-test score, where standardization is based on the mean and variance within each school type (*i.e.*, $5^{th}/6^{th}$ or $7^{th}/8^{th}$), in order to control for differences in starting human capital. Standard errors are in parentheses; p-values for a two-sided test of the null hypothesis of zero effect are italicized and in brackets. Estimates under each of the four effort variables are intended to capture the effect of a treatment on human capital investment for the total study population, and are therefore averaged over both investors and non-investors.

For disadvantaged group students we find evidence of large and positive effects across all four investment measures. First, we see a highly significant 8.7 percentage point increase in disadvantaged students' willingness to spend at least some time on the website, under AA. To put this in perspective, we can compute a within-demographic percent change for the disadvantaged group by $100 \times (\beta_1/\beta_0)\%$, which amounts to an increase of 119% on the extensive margin, relative to their disadvantaged counterparts under the

color-blind treatment. We also see a significant and even larger increase in terms of time investment: disadvantaged students under treatment RQ increased investment by 181%. The other two measures capture specific tasks done during time spent on the website: number of subjects attempted and number of questions attempted. Although the latter is only marginally significant, both render large point estimates for increases of 101% and 100%, respectively.

Another striking feature of the table are the performance measures. We find a large and significant difference in final exam scores and exam score changes for disadvantaged students under AA: both measures estimate that they lifted their scores by roughly a fifth of a standard deviation, relative to their disadvantaged counterparts in the control group. Although some portion of this effect may also be due to increased effort and concentration on the day of the final exam, we interpret this result and the other columns in Table 5 as evidence that treatment RQ altered labor-leisure trade-offs to induce additional study effort at home for $\mathcal{D}$ students, which in turn lead to more learning and increased math proficiency.

One concern is that this strengthening of incentives for disadvantaged students may come at the cost of weakening incentives for advantaged students. However, Table 5 also shows that this concern is not supported by the evidence. For 4 out of 5 outcome measures, point estimates for the effect of RQ on group $\mathcal{A}$, given by $\beta_1 + \beta_2$, was actually positive, but insignificant. In column 5, the sum of the two coefficients is slightly negative (representing about 3% of a standard deviation) but with a large p-value. The two outcome measures under which $\beta_1 + \beta_2$ is most significant are the binary measure of investment (column 1), with a p-value of 0.123, and the change in exam score (column 6), with a p-value of 0.03. This implies an estimated percent change of $100 \times (\beta_1 + \beta_2)/(\beta_0 + \beta_3) = +39.6\%$ on the extensive margin for advantaged students studying under a quota, with an estimated gain of about 1.5% of a standard deviation in their test scores. Thus, we do not find evidence that there is a trade-off between average human capital investment across demographic groups; if anything the data seem to slightly favor a small increase of investment for the advantaged demographic as well. Figures 13 – 16 in the Appendix contain graphical depictions of the distributional shifts of inputs and outputs by demographic and treatment group.

5.2.3. ***Selective Attrition.*** One potential source of bias in our results concerning the performance measure (final exam score) is that 97 of the students who took the practice test and were randomly assigned to a competition group (9.8%) did not show up on the day

of the final test.[18] We find that among the disadvantaged students, those assigned to the quota group were less likely to miss the final exam (10.6% vs. 16.7%). We also find that among the students who did not show up for the final test, the disadvantaged students assigned to the quota group had higher practice scores than the disadvantaged students not assigned to the quota group (7.16 vs 6.35). However, the practice scores among the students who did show up for the final test were nearly the same across these two groups (7.91 vs 7.79). These comparisons all point in a direction opposite of our main results and suggest that the effect of the quota on final performance for disadvantaged students may have been greater in the absence of this selective attrition.

5.2.4. ***Differing Policy Responses by Ability Level.*** The theoretical model predicts that if the underlying cost types for the disadvantaged group stochastically dominate those in the advantaged group, then qualitative patterns like those displayed in Figure 3 should appear. Namely, for group $\mathcal{D}$ the test score distributions under a quota and color-blind mechanism should have a unique interior crossing point, with the quota CDF strictly above the color-blind CDF to the right of the crossing point, and strictly below to the left. In other words, there should be a positive mass of the best students in group $\mathcal{D}$ who decrease output, while students of medium and low ability from that group increase output. Moreover, the upper bound on the output distribution may also be higher for $\mathcal{D}$ under a quota. The theory makes opposite predictions for group $\mathcal{A}$.[19]

While it is impossible to directly observe the distributions of cost types, we can take queues from the distribution of pre-test scores by demographic group, since they reflect initial math proficiency, as well as how much progress each student will need to increase her payout. We can then examine the distributions of final exam scores within demographic groups under different treatments to see whether our experimental data seem to be consistent with the theory of incentive effects under AA. Figures 4 – 6 depict these comparisons in three plots of empirical cumulative distribution functions (CDFs) for pre-test and final exam scores for grades 7 and 8. For the sake of comparability, we have limited our sample in these figures to include only students for whom we have both test scores. Therefore, Figure 4 plots empirical pre-test CDFs only for 7^{th} and 8^{th} graders who took the final exam.

[18]This problem does not arise with the four investment measures, which did not require observing a final score for us to estimate them. In Table 5 the sample size for the first four columns represent the full sample of test subjects.

[19]Although the figures we present in this paper explore distributions of both inputs and outputs, it is important to remember that the predictions of the theory only directly apply to exam score, as this is the variable on which prize allocations are based. The mapping from inputs to outputs may vary by student if each one differs by raw math talent and leisure preference.

FIGURE 4. PRE-TEST SCORES: 7^{th} GRADE VS 8^{th} GRADE

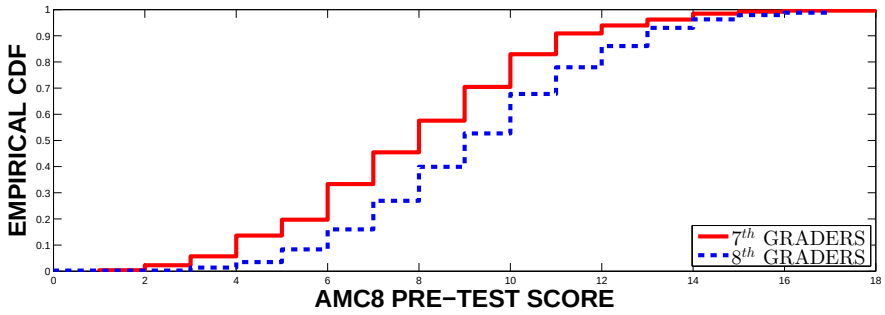


FIGURE 5. SEVENTH GRADE FINAL EXAM SCORES

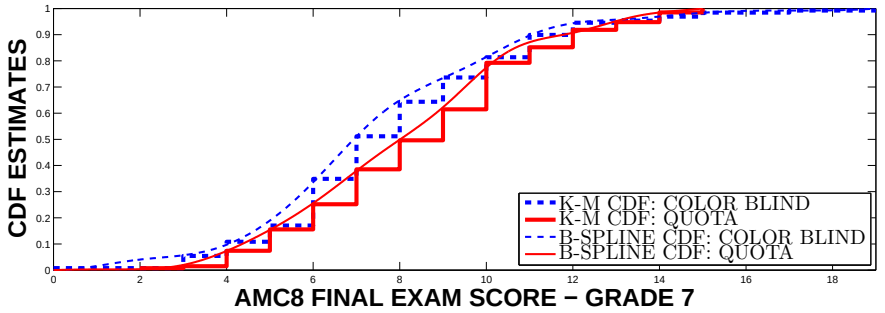
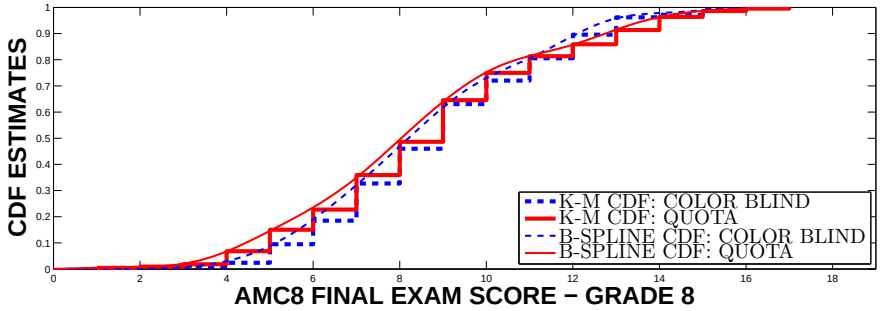


FIGURE 6. EIGHTH GRADE FINAL EXAM SCORES



Notes: For the sake of comparability, Figures 4 – 6 above use only data for 7^{th} and 8^{th} graders who took both the pre-test *and* final exam. See figures 15 – 16 in the appendix for a discussion on the role of selective attrition.

Figure 4 strongly supports stochastic dominance of initial math proficiency levels across demographic groups. A two-sample Kolmogorov-Smirnov (KS) test rejects the null hypothesis that the 7^{th} and 8^{th} grade distributions are the same, against a one-sided alternative that the latter stochastically dominates with a p-value of 1.03×10^{-5}. This means group $\mathcal{D}$ students, on average, had to achieve more progress in order to be competitive for a prize. This is not the same as observing costs, but the two are certainly related and the hypothesis of stochastic dominance in cost types appears plausible.

We find evidence in Figure 5 that by the end of the study period the score distribution within the 7th grade subsample had diverged by treatment status. As a first attempt at investigating the difference, we perform two-sided KS test for group $\mathcal{D}$ final exams across treatments, which results in a p-value of 0.1047, providing marginally significant evidence that the two distributions were not the same.[20] There appears to be a hint of divergence by treatment group among 8th graders in Figure 6 as well, though the effect is weak and we lack sufficient power for a KS test to distinguish between the two distributions (the p-value for a two-sided test is 0.889).

For our application, the KS test has important limitations. Aside from being a relatively low-power test, it also does not provide a readily interpretable measure for the significance and magnitude of effects at a given point in the distribution. In order to evaluate theoretical predictions concerning how AA shapes incentives differently for students who vary by ability levels, we wish to make within-group, cross-treatment comparisons which allow for the magnitude, and even the sign, of the effect to vary across different quantiles. One challenge in making such comparisons is that the set of quantile ranks $q \in [0,1]$ attained by the empirical CDFs for two different samples will rarely overlap. To overcome this problem, we construct smoothed quantile functions based on a flexible B-spline CDF estimator.[21] Our smoothed quantile function estimator allows for comparisons at arbitrary quantile ranks q, with differences that are easily interpretable, being in the original test score units.

Specifically, within each treatment $t \in \{RQ, CB\}$ and group $j \in \{\mathcal{A}, \mathcal{D}\}$ we partition the support of final exam scores $[0, \overline{h}_j^t]$ into K uniform subintervals and we specify the breakpoints between these intervals as a vector of $K+1$ knots (including endpoints). These knots define a set of $K+3$ B-spline basis functions, denoted $\mathcal{B}_{j,k}^t : [0, \overline{h}_j^t] \to \mathbb{R}$, $k =$

[20] Figures 15 and Figure 16 in the Appendix contain additional plots comparing pre-test scores by treatment within demographic groups. The differences in the pre-test distributions are due to selective attrition after we omitted students for whom we have no final exam. The figures suggest that in general selective attrition is working against our results presented here. For group $\mathcal{D}$ the pre-test color-blind distribution is below the pre-test quota distribution for values at or below the median, and the upper bound of the pre-test distribution for 7th grade quota students is highest. Both characteristics of these distributions are substantially reversed by the final exam. Figure 16 suggests that selective attrition within the 8th grade is also working slightly against finding a significant difference in the direction predicted by theory.

[21] B-splines constitute a parametric family of functions which combines flexibility, stability, and computational convenience. They behave similarly as piece-wise, local polynomials, but afford greater numerical convenience, being made up of globally defined basis functions, like orthogonal polynomials. Moreover, incorporating shape restrictions (*e.g.*, monotonicity and terminal conditions for a CDF function), is quite simple within the B-spline family. A standard text on B-splines is Carl de Boor [2001]. For a brief primer on B-splines and their advantages for empirical economics, see Brent R. Hickman, Timothy P. Hubbard and Harry J. Paarsch [2016, Appendix].

$1, \ldots, K+3$, with which we parameterize the CDF of final exam scores as

$$\hat{G}_j^t(h; \boldsymbol{\alpha}_j^t) = \sum_{k=1}^{K+3} \alpha_{j,k}^t \mathcal{B}_{j,k}^t(h).$$

Finally, we estimate the parameter vector $\boldsymbol{\alpha}_j^t$ to achieve a constrained, least-squares, best fit to the empirical CDF, where constraints ensure that $\hat{G}_j^t$ is monotone and attains a value of 0 at the lower bound, and 1 at $\overline{h}_j^t$, or the maximum observed score for group j in treatment t. Our method belongs to the class of *sieve estimators* since, if we allow the number of knots $(K+1)$ to grow with the sample size, a B-spline can accommodate arbitrary shapes for any well-behaved CDF. We chose a partition of $K = 10$ uniform subintervals for each support, which results in a flexible CDF estimator with 11 free parameters once the terminal conditions are enforced. Figures 5 and 6 display a comparison of the Kaplan-Meier empirical CDFs (thick lines) and the smoothed, B-spline CDFs (thin lines) for each treatment and age group.

With these smoothed CDF estimates in hand, we can construct quantile functions, which we denote by $\hat{h}_j^t(q) = (\hat{G}_j^t)^{-1}(q),\ q \in (0,1)$. Following the notation of Corollary 2 in Section 3, from this we can compute the quantile difference function,

$$\hat{\Delta}_j(q) \equiv \hat{h}_j^{RQ}(q) - \hat{h}_j^{CB}(q),\ q \in [0,1],\ j = \mathcal{A}, \mathcal{D}.$$

This is the primary object of interest, as it allows us to make behavioral comparisons across policies at specific quantiles. Moreover, in order to investigate the role of sampling variability we execute a bootstrap routine, wherein for each age group $j = \mathcal{A}, \mathcal{D}$, we re-sampled from the quota and color-blind samples (with replacement) and re-estimated $\hat{\Delta}_j(\cdot)$ 50,000 times. The left panels of Figures 7 and 8 display point estimates and 95% confidence bounds from this exercise; for each, quantile ranks $q \in [0,1]$ are on the horizontal axis, and differences displayed in final exam score units are on the vertical axis. Point estimates for the functions $\Delta_{\mathcal{A}}(\cdot)$ and $\Delta_{\mathcal{D}}(\cdot)$ conform well to the patterns predicted in the theory: both have a unique interior point where they cross the zero line. For group $\mathcal{D}$, achievement under a quota within the interquartile range improves substantially, by a margin of 17%–43% of a standard deviation. There is a difference in the opposite direction for $\mathcal{A}$, but it is much smaller, with a point estimate from 6%–12% of a standard deviation within the interquartile range.

The right panels in the two figures illustrate the strength of the statistical evidence for the quantile-specific differences by evaluating the null hypothesis $\mathcal{H}_0 : \Delta_{\mathcal{D}}(q) = 0$ at each point, against the two alternative hypotheses where the sign is negative, denoted by

FIGURE 7. SEVENTH GRADE FINAL EXAM SCORES

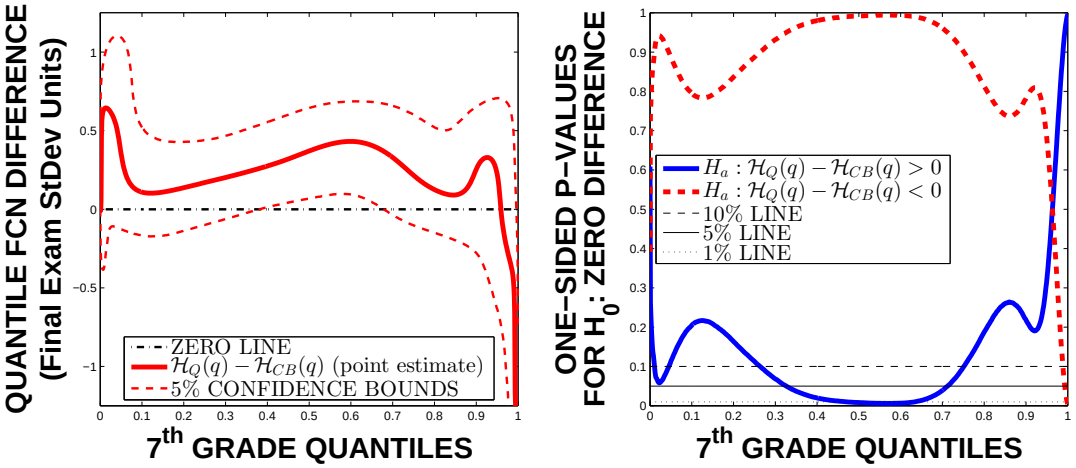
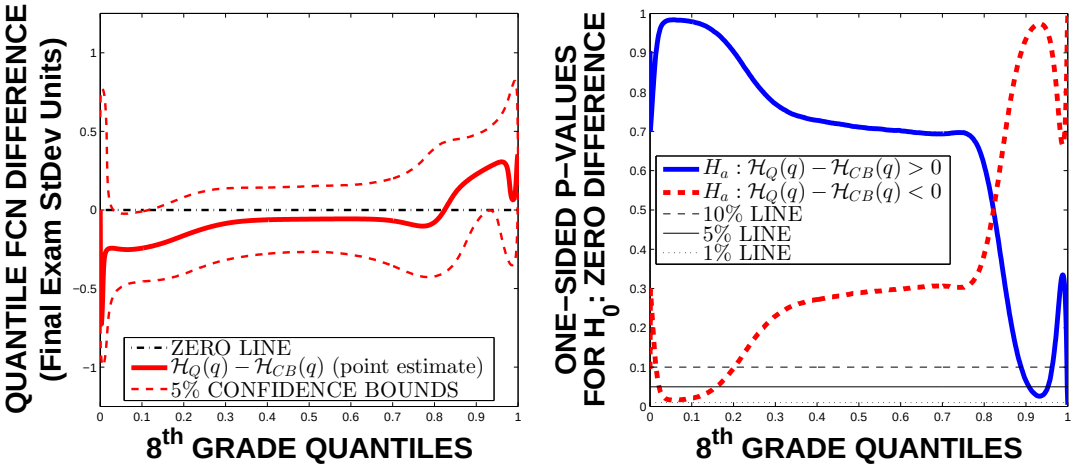


FIGURE 8. EIGHTH GRADE FINAL EXAM SCORES



$\mathcal{H}_a^-$, or positive, denoted by $\mathcal{H}_a^+$. At each point $q \in [0,1]$ we computed point-wise one-sided P-values as the fraction of bootstrapped samples resulting in a positive difference (solid lines), or a negative difference (dashed lines) at q. Each panel also displays thin horizontal lines for the 10%, 5%, and 1% thresholds for comparison. The figures show strong evidence for a positive interquartile range difference in group $\mathcal{D}$, and they also show some evidence of the opposite effect near the upper bound of the support. The evidence for policy-induced shifts in group $\mathcal{A}$ is much weaker, but there is some evidence of statistically significant changes near the extremes of the sample, and once again, in directions consistent with those predicted by theory.

5.2.5. ***Narrowing Achievement Gaps.*** We now conclude analysis of our experimental data with a look at the tendency for AA to narrow achievement gaps across demographic groups. Table 6 displays summary statistics on standardized test scores for the pre-test

TABLE 6. NARROWING GAPS

	Mean	Median	Std. Dev.	N
Achievement Gaps for All Treatments				
Standardized Pre-Score (GRADE 7)	-0.295	-0.267	0.996	264
Standardized Pre-Score (GRADE 8)	0.181	0.071	0.960	431
Standardized Final Score (GRADE 7)	-0.138	-0.205	1.005	264
Standardized Final Score (GRADE 8)	0.085	0.142	0.987	431
Achievement Gaps for Quota Treatment				
Standardized Pre-Score (GRADE 7)	-0.211	-0.199	1.061	135
Standardized Pre-Score (GRADE 8)	0.231	0.135	0.920	220
Standardized Final Score (GRADE 7)	-0.030	0.142	0.961	135
Standardized Final Score (GRADE 8)	0.055	0.142	1.050	220
Achievement Gaps for Color-Blind Treatment				
Standardized Pre-Score (GRADE 7)	-0.243	-0.199	0.904	129
Standardized Pre-Score (GRADE 8)	0.259	0.135	0.982	211
Standardized Final Score (GRADE 7)	-0.251	-0.552	1.041	129
Standardized Final Score (GRADE 8)	0.115	0.142	0.921	211

Notes: There are three separate panels in the table, each containing standardized scores on the pre-test and post-test. Standardization was performed within each panel-test grouping, excluding scores for students who missed the final exam. For example, pre-test scores for the quota treatment were standardized using the mean and standard deviation of pre-test scores for all 7^{th} and 8^{th} graders in the quota treatment who took both the pre-test and post-test.

and final exam, for grades 7 and 8. In the top panel of the table scores were standardized within each exam by subtracting the mean and dividing by the standard deviation for all grade 7 and 8 students.[22] Therefore, the means indicate distance between the population average and grade cohort average, in standard deviation units. Without accounting for treatment status we see that 7^{th} grade students were roughly half of a standard deviation behind their 8^{th} grade counterparts on average—or $-0.295 - 0.181 = -0.476$ standard

[22]Once again, in order to make the pre-test and final exam figures comparable, we excluded from the analysis any students whose final scores were missing due to attrition.

deviations to be exact—but by the final exam, the gap between 7^{th} and 8^{th} graders had narrowed by about half—to $-0.138 - 0.085 = -0.223$ standard deviations.

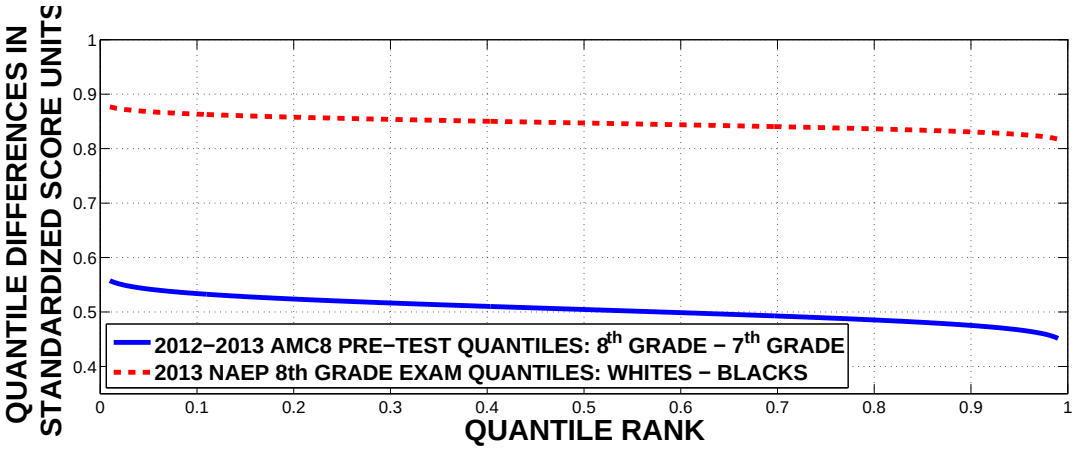
In the lower two panels of Table 6 we break out this effect by treatment group, displaying the same numbers within treatments, but where score standardization now happens within each exam-treatment cell. Part of the test score convergence had to do with differences in the conditions of the pre-test and final exams (likely due to incentives or slightly different content): within the color-blind treatment, about a quarter of the gap disappeared but remained relatively high at 0.367 standard deviations on the final exam. However, the achievement gap under AA closed substantially more, by about 80%, beginning at 0.442 standard deviations, and ending at only 0.085 standard deviations on the final exam.

The medians tell a slightly stronger story, with the median gap beginning about the same within both treatments, closing virtually to zero under AA, and closing only slightly otherwise. Finally, observing that the within-treatment-demographic standard deviations are all close to one suggests that the narrowing of gaps within the two treatments was due predominantly to mean/median shifts in test scores. We interpret these findings as evidence that AA can actually help to narrow achievement gaps by catching up its target demographic to the rest of the population.

6. DISCUSSION AND CONCLUSION

6.1. **EXTERNAL VALIDITY AND DIRECTIONS FOR FUTURE RESEARCH.** While interpreting the results from this study, it is important to keep in mind what it can and cannot say. The ultimate question of interest is the effect that real world AA has on the human capital incentives of American under-represented minorities. There are several reasons why this study cannot fully address this policy question. First, in executing field experiments with actual schools, a significant hurdle involves forming partnerships with school administrators whose primary job it is to educate, rather than to produce research. Our school partners in this experiment serve a fairly homogeneous population. On one hand, this produces a clean exploration of theoretical predictions, using grade level as a proxy for race, so that learning costs and market institutions systematically differ, but all else is held fixed. On the other hand, this proxy is obviously imperfect. As Figure 9 illustrates, it likely understates the competitive difference in math proficiency between blacks and whites vying for actual college seats. The graph compares the difference between black and white 8^{th} grade NAEP quantiles—expressed in standardized NAEP score units (dashed line)—to the difference in 7^{th} and 8^{th} grade quantiles from our sample—expressed in standardized AMC8 score units (solid line). At most NAEP

FIGURE 9. GRADE-LEVEL DIFFERENCES VS BLACK/WHITE DIFFERENCES



(A) This figure first standardizes AMC8 and NAEP exam scores by subtracting the mean and dividing by the standard deviation for the relevant sample population. Then, the distribution of standardized test scores for each sub-sample–*i.e.*, AMC8 7th grade, AMC8 8th grade, NAEP 8th grade whites, and NAEP 8th grade blacks–is approximated by a normal distribution, which appears to fit the data quite well. The resulting quantile functions are subtracted from each other in order to illustrate relative disparities between groups.

quantiles, black 8th graders trail their white counterparts by about 0.85 standard deviations, whereas the gap among our test subjects is closer to 0.5 standard deviations. Another way to interpret the figure is that wherever the dashed line lay above the solid line indicates that the corresponding quantiles of American Blacks are *more than* one full grade level behind their white, 8th Grade counterparts.

This picture suggests real world discouragement effects minorities face in the competitive college market may be even larger than those found in our experimental study, opening the possibility that AA may have a larger impact. On the other hand, our homogeneous subject pool does not account for important differences in home environments, school quality, or cultural factors which may push the results in other, less predictable directions. Certainly though, in order to understand how and why a policy works, it is important to cleanly understand the incentive dimension in addition to other sociocultural factors. The results here demonstrate that AA can be used as an effective tool to improve performance incentives and promote investment activity among systematically disadvantaged populations.

The second major limitation in the current analysis is that, although it creates a set of incentives which mirror human capital competition in key ways, due to feasibility constraints said incentives are only engineered on a small scale and measured over a short-run horizon. We tracked our students over a period of 10-days during which they

were given the opportunity to invest leisure time into their math proficiency. For students in our sample who logged positive amounts of study time, their average expected wage was \$10.73/hour, measured as total combined earnings divided by total observed time spent studying on the website, across all students. Junior high and high school students preparing for college optimize labor-leisure division over a much longer horizon (4-5 years until college applications are due), with much larger payoffs that come far in the future.[23] The question of quantifying incentives running up to the US college market is complex and involves many factors, including time preferences, cost of a college education, and many other things that vary widely across individuals. However, we provide some rough calculations here to get a basic sense of how the expected hourly wage within our experiment might compare to the effective hourly wage for a high school student studying for college. To do so we calibrate payoffs using market aggregate figures taken from various sources.

Bodoh-Creed and Hickman [2016] use data from the Current Population Survey (CPS) from 2003 to calibrate the average net present value (NPV) of lifetime, post-college-age, personal income for workforce participants with different levels of education. They separate income by age for individuals 23–65 and take averages within each age-education group, to get $I_{a,e},\ a \in \{23,\dots,65\},\ e \in \{cg, sc, hsg\}$, where cg denotes a college graduate (with no additional, advanced degrees), sc denotes a high-school graduate who enrolled in college but did not graduate, and hsg denotes a high-school graduate with no college experience. They then assume various values of the annual time discount factor, δ, and compute $NPV_e^{18} = \sum_{a=23}^{65} \delta^{(a-23+5)} I_{a,e}$, where discounting begins 5 years into the future so as to capture the perspective of a recent high-school graduate, at 18 years of age, who is considering going to college for four years.[24] We use their numbers as a rough estimate of the average gross payoff to the three education options. As for costs, the College Board estimated that the average cost of attendance during the 2002-2003 academic year was \$28,090 at four-year, private, non-profit universities and \$12,376 at four-year publics, which we denote by $AnnualCOA$. We shall assume that these same

[23] A large literature has focused on the wage return to higher college quality, including Stacy Berg Dale and Alan B. Krueger [2002], Dan A. Black and Jeffrey A. Smith [2006], and Long [2008]. Although there is some disagreement as to the magnitude for all students, there is broad consensus that the return to attending higher quality colleges is economically significant for poor students.

[24] Bodoh-Creed and Hickman [2016] omit the comparison during college years as it is unclear what the appropriate comparison would be. College attendees forgo 4 years of labor force income, but most also remain their parents' dependents until graduation, meaning their consumption during college to some degree is still a function of their parents' income. It is also unclear the extent to which recent high school graduates who do not attend college still derive consumption from their parents' income stream. Therefore Bodoh-Creed and Hickman [2016] assume that by age 23 individuals of all education levels become independent adults.

costs prevailed over all four previous years of study.[25] The National Center for Education Statistics reports the 6-year college graduation rate, which we denote by *GradRate*, for the entering class of 1996 (*i.e.*, all those who had graduated by AY2002-2003) as 0.631 for private universities, and 0.517 for publics.[26] We shall assume that college dropouts incur 2 years of college attendance cost before dropping out.

In order to estimate how many hours, on average, high school students devote to academics, we turn to the Schools and Staffing Survey (SASS), published by the National Center for Education Statistics. SASS reports that in 2003 the average number of hours in the school day for American K-12 schools was 6.6, and that the average number of school days in their academic year was 178.7. If we assume that each school day is coupled with 3.5 hours of homework for high schoolers, then this gives us a measure of $HSHoursWorked = 4 \times 178.7 \times (6.6 + 3.5) = 7,219.48$.[27] If we double the number of daily homework hours to 7 then this number rises by a third. Finally, we put these various measures together for a rough measure of how 18-year-old college enrollees are compensated for the time they invested into learning during 4 years of high school:

$$\begin{aligned} HSHourlyWage = \Big[& GradRate \times \left(NPV_{cg}^{18} - \textstyle\sum_{t=1}^{4} \delta^t AnnualCOA \right) \\ & + (1 - GradRate) \times \left(NPV_{sc}^{18} - \textstyle\sum_{t=1}^{2} \delta^t AnnualCOA \right) - NPV_{hs}^{18} \Big] \Big/ HSHoursWorked \end{aligned} \tag{8}$$

Intuitively, the numerator is a measure of the lump-sum transfer an average college-bound high school graduate would need in order to forgo college, and the denominator normalizes this transfer by the number of total hours spent preparing for college. Since this number will vary by assumptions of discount factor, college option (*i.e.*, public vs. private), and daily homework time inputs, Table 7 presents various hourly wage calculations covering several different scenarios. A commonly assumed annual discount factor in the macroeconomics literature is $\delta = 0.96$, and a student with these time preferences who studies 3.5 hours per school day will garner an effective return of \$33.72 per hour if she attends a private university, and \$34.90 per hour if she attends a public. Depending on the various assumptions, this number could range from \$10 and \$55 per hour.

[25] This is a conservative assumption, given that education costs were steadily on the rise during this period. The cost of attendance number includes tuition, fees, room and board, books/supplies, transportation, and other living expenses. Figures taken from Table 1 of report *Trends in College Pricing*, downloaded at http://www.collegeboard.com/prod_downloads/press/cost03/cb_trends_pricing_2003.pdf on April 29, 2016.

[26] Numbers taken from *The Digest of Education Statistics*, downloaded from https://nces.ed.gov/programs/digest/d13/tables/dt13_326.10.asp on April 29, 2016.

[27] Figures downloaded from SASS reports at https://nces.ed.gov/surveys/sass/tables/sass0708_045_s1n.asp and https://nces.ed.gov/surveys/sass/tables/sass0708_046_d1n.asp on April 29, 2016.

TABLE 7. Projections of Effective Hourly Wage for College-Bound High Schoolers

		PUBLIC COLLEGE HS Homework Hours		PRIVATE COLLEGE HS Homework Hours	
		3.5/school day	7/school day	3.5/school day	7/school day
ANNUAL DISCOUNT FACTOR	$\beta = 0.92$	\$16.99	\$12.62	\$13.77	\$10.23
	$\beta = 0.96$	\$34.90	\$25.92	\$33.72	\$25.04
	$\beta = 0.98$	\$54.10	\$40.18	\$55.10	\$40.92

At the end of the day, the rough calculations in Table 7 cannot speak directly to the question of how much AA *changes* expected economic outcomes from competitive human capital investment, and how these changes interact with individual ability and forward-looking behavior. How these various forces balance out in the long run is a challenging question that deserves additional attention from researchers. Answering this question directly using experimental methods would would require an ability to track time use decisions and exogenously shifting market incentives over a period of years or even decades. However, the numbers above suggest that our experimental incentives may not be entirely un-representative of the magnitudes of incentives that high-school students face on a regular basis when making decisions about leisure-study tradeoffs. Moreover, our experimental results at least inform us that students who are about to enter high school are capable of the kind of complex optimization that would imply AA having a nontrivial, positive impact on minority human capital, and narrowing achievement gaps. In that sense, we believe that our short-run evidence, while suggestive in nature, provides an important step toward a more complete picture of the role of AA in American higher education.

6.2. **CONCLUSION.** We designed a field experiment in which middle school and junior high students compete for heterogeneous cash prizes and are paid by their relative performance on a nationwide math exam. Our experimental design is intended to give insight into students' pre-college human capital investment followed by the college admissions market; namely, voluntary labor-leisure decisions, mathematics learning, and affirmative action. Within this context, the rank-order prize allocation rule approximated either a color-blind system, or a representative quota AA policy. We tracked student study effort during a 10-day investment period prior to the final exam. Although our

experiment involves relatively low-value prizes and a relatively low-stakes exam, we still find sizable and significant effects of AA on motivation and learning.

Our results are clearly suggestive that students adjust their labor supply in response to AA considerations. Although some of this effect may come through changes in effort or focus on the day of the exam, we also find evidence that the disadvantaged students are more likely to study for the exam under AA. To the extent that our experiment is informative about American university admissions, the findings suggest two things: (i) that AA may improve a minority student's real effort to build human capital in preparation for college, and (ii) that it can actually *narrow* achievement gaps while promoting diversity in higher education.

References

Arcidiacono, Peter. 2005. "Affirmative Action in Higher Education: How do Admission and Financial Aid Rules Affect Future Earnings?" *Econometrica*, 73(5): 1477–1524.

Bettinger, Eric. 2012. "Paying to Learn: The Effect of Financial Incentives on Elementary School Test Scores." *Review of Economics and Statistics*, 94: 686–698.

Black, Dan A., and Jeffrey A. Smith. 2006. "Estimating the Returns to College Quality Using Multiple Proxies for Quality." *Journal of Labor Economics*, 24(3): 701–728.

Bodoh-Creed, Aaron, and Brent R. Hickman. 2016. "College Assignment as a Large Contest." *working paper, University of Chicago*.

Bowen, William G., and Derek Bok. 1998. *The Shape of the River: Long-Term Consequences of Considering Race in College and University Admissions.* Princeton, NJ: Princeton University Press.

Bracha, Anat, Alma Cohen, and Lynn Conell-Price. 2015. "Affirmative Action and Stereotype Threat." working paper.

Calsamiglia, Caterina, Jorg Franke, and Pedro Rey-Biel. 2013. "The incentive effects of affirmative action in a real-effort tournament." *Journal of Public Economics*, 98: 15–31.

Chambers, David L., Timothy T. Clydesdale, William C. Kidder, and Richard O. Lempert. 2005. "The Real Impact of Eliminating Affirmative Action in American Law Schools: An Empirical Critique of Richard Sander's Study." *Stanford Law Review*, 57: 1855–1897.

Coate, Stephen, and Glenn Loury. 1993. "Anti-Discrimination Enforcement and the Problem of Patronization." *American Economic Review*, 83(2): 92–98.

Cotton, Christopher, Frank McIntyre, and Joseph Price. 2013. "Gender differences in repeated competition: Evidence from school math contests." *Journal of Economic Behavior and Organization*, 86: 52–66.

Dale, Stacy Berg, and Alan B. Krueger. 2002. "Estimating the Payoff to Attending a More Selective College: An Application of Selection on Observables and Unobservables." *Quarterly Journal of Economics*, 117: 1491–1528.

de Boor, Carl, ed. 2001. *A Practical Guide to B-Splines, Revised Edition.* New York: Springer-Verlag.

Ferman, Bruno, and Juliano Assuncao. 2011. "Does Affirmative Action Enhance or Undercut Investment Incentives? Evidence from Quotas in Brazilian Public Universities." *Typescript, Massachusetts Institute of Technology Department of Economics*.

Fryer, Roland. 2011. "Financial Incentives and Student Achievement: Evidence From Randomized Trials." *Quarterly Journal of Economics*, 126: 1755–1798.

Hickman, Brent R. 2015. "Human Capital Investment and Affirmative Action: A Structural Policy Analysis of US College Admissions." *Typescript, University of Chicago Department of Economics*.

Hickman, Brent R., Timothy P. Hubbard, and Harry J. Paarsch. 2016. "Investigating the Economic Importance of Pricing-Rule Mis-Specification in Empirical Models of Electronic Auctions." *typescript, University of Chicago*.

Howell, Jessica S. 2010. "Assessing the Impact of Eliminating Affirmative Action in Higher Education." *Journal of Labor Economics*, 28(1): 113–66.

Kremer, Michael, Edward Miguel, and Rebecca Thornton. 2009. "Incentives to Learn." *Review of Economics and Statistics*, 91: 437–456.

Leuven, Edwin, Hessel Oosterbeek, and Bas van der Klaauw. 2010. "The Effect of Financial Rewards on Students' Achievement: Evidence from a Randomized Experiment." *Journal of the European Economic Association*, 8: 1243–1265.

Long, Mark C. 2008. "College Quality and Early Adult Outcomes." *Economics of Education Review*, 27: 588–602.

Loury, Linda Datcher, and David Garman. 1995. "Selectivity and Earnings." *Journal of Labor Economics*, 13(2): 289–308.

Rothstein, Jesse, and Albert H. Yoon. 2008. "Affirmative Action in Law School Admissions: What do Racial Preferences Do?" *University of Chicago Law Review*, 75(2): 649–714.

Sander, Richard H. 2004. "A Systemic Analysis of Affirmative Action in American Law Schools." *Stanford Law Review*, 57: 367–483.

Schotter, Andrew, and Keith Weigelt. 1992. "Asymmetric Tournaments, Equal Opportunity Laws and Affirmative Action: Some Experimental Results." *Quarterly Journal of Economics*, 107(2): 511–539.

7. APPENDIX

7.1. THEORETICAL PROOFS APPENDIX.

7.1.1. ***Proof of Theorem 1***. The result of the theorem follows from Bodoh-Creed and Hickman [2016, Theorems 3 and 4], which establish a claim that is both broader (it allows for shifts in F_S as well) and holds under a more general set of conditions. Because of this their proof is quite complicated, so for completeness we provide a proof for the simplified version of the model from Section 3. For the reader's convenience Theorem 1 is restated below:

Theorem 1: *Consider two color blind contests,* 1 *and* 2*, which differ only by their cost distributions, and assume competition is more fierce under contest* 2 *in the sense that F_1 LR dominates F_2. Let $\tilde{\theta} \in (\underline{\theta}, \overline{\theta})$ denote the unique crossing point of the density functions where $f_1(\tilde{\theta}) = f_2(\tilde{\theta})$. There exists a unique interior crossing point $\ddot{\theta} \in (\underline{\theta}, \tilde{\theta})$, such that $h_1^*(\theta) < h_2^*(\theta)$ for all $\theta < \ddot{\theta}$ and $h_1^*(\theta) > h_2^*(\theta)$ for all $\theta > \ddot{\theta}$.*

Proof: As mentioned above, the LR dominance property implies first-order stochastic dominance. Therefore, not only do the densities have a unique crossing point, but it must also be true that $f_1(\theta) > f_2(\theta)$ for $\theta > \tilde{\theta}$, and $f_1(\theta) < f_2(\theta)$ for $\theta < \tilde{\theta}$. Since the same boundary condition applies to both contests, $h_1^*(\overline{\theta}) = h_2^*(\overline{\theta}) = \underline{h}$, then equation (5) implies the initial trajectories at the boundary point are ordered in the following way:

$$h_1^{*\prime}(\overline{\theta}) = \frac{-f_1(\overline{\theta})}{c'(\underline{h}; \overline{\theta})} < \frac{-f_2(\overline{\theta})}{c'(\underline{h}; \overline{\theta})} = h_2^{*\prime}(\overline{\theta}). \tag{A.1}$$

This in turn means that $h_1^*(\theta) > h_2^*(\theta)$ within a neighborhood of $\overline{\theta}$ since the investment functions are continuous and everywhere differentiable. Note that because slopes are negative $h_1^{*\prime}(\theta) < h_2^{*\prime}(\theta)$ means h_1^*, rises in the leftward direction and is *more steep* at θ.

Now suppose there exists at least one point where h_1^* and h_2^* cross, and let $\ddot{\theta} \in (\underline{\theta}, \overline{\theta})$ denote the maximum of all such possible points, with $\ddot{h} \equiv h_1^*(\ddot{\theta}) = h_2^*(\ddot{\theta})$, if any exist. Since h_1^* crosses h_2^* from above at $\ddot{\theta}$, it must be that $h_1^{*\prime}(\ddot{\theta}) \geq h_2^{*\prime}(\ddot{\theta})$ (*i.e.*, h_1^* is *less steep* at the crossing point). However, since $f_1(\theta) > f_2(\theta)$ on $(\widetilde{\theta}, \overline{\theta}]$ by LR dominance, and since $h < h'$ implies $c'(h'; \theta) \geq c'(h; \theta)$ by convexity, the following must be true for any $\theta \in (\tilde{\theta}, \overline{\theta})$:

$$h_1^{*\prime}(\theta) = \frac{-f_1(\theta)}{c'\left[h_1^*(\theta); \theta\right]} < \frac{-f_2(\theta)}{c'\left[h_2^*(\theta); \theta\right]} = h_2^{*\prime}(\theta) \quad \Rightarrow \quad h_1^*(\theta) > h_2^*(\theta). \tag{A.2}$$

Therefore, $\ddot{\theta} < \tilde{\theta}$, if such a point exists. Similarly, since $f_1(\theta) < f_2(\theta)$ on $[\underline{\theta}, \widetilde{\theta})$, then any crossing point would have to obey $h_1^{*\prime}(\ddot{\theta}) = -f_1(\ddot{\theta})/c'(\ddot{h}; \ddot{\theta}) > -f_2(\ddot{\theta})/c'(\ddot{h}; \ddot{\theta}) = h_2^{*\prime}(\ddot{\theta})$.

This means that h_1^* can only cross h_2^* from above (*i.e.*, at points where it is *less steep*) and so there can be at most one such crossing.

Finally, to see why a crossing point must exist, suppose for a contradiction that for all $\theta \in (\underline{\theta}, \overline{\theta})$ we have $h_1^*(\theta) > h_2^*(\theta)$. In that case, it follows that

$$\begin{aligned} h_1^*(\underline{\theta}) &= \int_{\underline{\theta}}^{\overline{\theta}} \frac{1}{c'\left[h_1^*(u); u\right]} f_1(u)du + \underline{h} \\ &< \int_{\underline{\theta}}^{\overline{\theta}} \frac{1}{c'\left[h_1^*(u); u\right]} f_2(u)du + \underline{h} \\ &\leq \int_{\underline{\theta}}^{\overline{\theta}} \frac{1}{c'\left[h_2^*(u); u\right]} f_2(u)du + \underline{h} \;=\; h_2^*(\underline{\theta}), \quad \rightarrow\leftarrow . \end{aligned}$$

The strict inequality follows because the first and second line depict expectations over the decreasing function $1/c'\left[h_1^*(u); u\right]$ and f_2 places more weight on strictly higher values of u (or lower values of the function). The weak inequality follows from the supposition and from c being convex and having a positive cross-partial derivative. Thus we have a contradiction, so a unique crossing $\ddot{\theta}$ exists on the open interval $(\overline{\theta}, \tilde{\theta})$ and the theorem is proved. ■

7.2. **THEORY EXTENSION: INTRINSICALLY VALUED HC.** In the body of the paper, we assume that students only indirectly benefit from their HC accumulation. Here, we consider an extension to the model where h is intrinsically valued to the student, and not just a means of securing a high-quality seat assignment. Naturally, with a more flexible model comparative statics become more difficult to prove. Therefore, we explore some simple examples based on a Cobb-Douglas utility form $u(s,h) = s^\alpha h^\beta$, in order to argue that the qualitative model predictions explored in Section 3 are non-pathological.

7.2.1. ***Example 1: Linear Utility***. If $\alpha = \beta = 1$ and costs are linear so that $u(s,h) = sh$, $c(h;\theta) = \theta(h - \underline{h})$, and $[\underline{\theta}, \overline{\theta}] \subset (1, \infty)$, then proving statements analogous to Theorem 1, Corollary 2, and Corollary 3 is relatively straightforward.[28] In this scenario, under CB student i's objective is now

$$\max_{h_i \geq \underline{h}} \left\{ h_i \left(1 - F\left[\theta^*(h_i)\right]\right) - \theta_i(h_i - \underline{h}) \right\}. \tag{A.3}$$

Taking an FOC, we get $-h_i f\left[\theta_i^*(h_i)\right] \theta^{*\prime}(h_i) + \left(1 - F\left[\theta^*(h_i)\right]\right) = \theta_i$. In this world, student i can be thought to choose HC production in two parts. First, she raises investment to the level where she optimally benefits from complementarity of her own input h_i and

[28]The assumption of $[\underline{\theta}, \overline{\theta}] \subset (1, \infty)$ is required because the gross utility and cost function are both linear in h. Since the direct marginal benefit of more human capital is $u_2(s,h) = s$ and the direct marginal cost is θ, we must have $\underline{\theta} > \overline{s}$, in order to rationalize all students choosing finite HC production.

the university's input s; this calculation involves the direct marginal benefit of a unit of HC, or $(1 - F[\theta^*(h_i)])$, and Bodoh-Creed and Hickman [2016] refer to it as the *productive channel* of incentives. Above that base level of investment, competitive incentives play an additional role: in a monotone equilibrium i invests further so as to make her higher-cost competitors prefer not to try and jump her. This calculation involves the indirect placement benefit of more HC, or $-h_i f\left[\theta_i^*(h_i)\right]\theta^{*\prime}(h_i)$, and it produces a tendency for over-investment which is referred to as the *competitive channel* of incentives.

If we now define $\eta(\theta) \equiv \log(h^*(\theta))$ then through a similar change of variables as before we can rearrange the FOC to get a differential equation

$$\eta'(\theta_i) = -\frac{f(\theta_i)}{\theta_i + F(\theta_i) - 1}, \text{ with boundary condition } \eta(\overline{\theta}) = \log(\underline{h}). \tag{A.4}$$

Once again, η is strictly decreasing in θ and we can now prove the analog of Theorem 1 above:

Theorem 4. *Assume HC production costs and gross utility are linear in h, so that $c(h;\theta) = \theta(h - \underline{h})$, and $u(s,h) = sh$. Moreover, consider two cost distributions, $F_1(\theta)$ and $F_2(\theta)$, where competition is more intense under F_2 in the sense that F_1 LR dominates F_2 (i.e., the ratio $\frac{f_1(\theta)}{f_2(\theta)}$ is strictly increasing). Then, letting $\tilde{\theta}$ denote the unique crossing point of f_1 and f_2, there exists a unique interior crossing point $\ddot{\theta} \in (\underline{\theta}, \widetilde{\theta})$ such that $h_1^*(\theta) < h_2^*(\theta)$ for $\theta < \ddot{\theta}$ and $h_1^*(\theta) > h_2^*(\theta)$ for $\theta > \ddot{\theta}$.*

Proof: Recall that strict LR dominance implies $f_1(\theta) \geq f_2(\theta)$ and $F_1(\theta) < F_2(\theta)$, for $\theta \in [\widetilde{\theta}, \overline{\theta})$. This with $\eta_1(\overline{\theta}) = \eta_2(\overline{\theta})$ and equation (A.4) together mean that $\eta_1'(\theta) < \eta_2'(\theta)$ and $\eta_1(\theta) > \eta_2(\theta)$, for each $\theta \in [\widetilde{\theta}, \overline{\theta})$. Thus, if η_1 and η_2 cross, the crossing must be on the interval $[\underline{\theta}, \widetilde{\theta})$.

Now, equation (A.4) can be expressed in integral form by $\eta_j(\theta) = \int_\theta^{\overline{\theta}} \frac{f_j(x)}{x + F_j(x) - 1} dx + \log(\underline{h})$, $j = 1, 2$. Moreover, if we impose a change of variables $y = F_j(\theta)$ within the integral, we get

$$\eta_1(\underline{\theta}) = \int_0^1 \frac{1}{F_1^{-1}(y) + y - 1} dy + \log(\underline{h}) < \int_0^1 \frac{1}{F_2^{-1}(y) + y - 1} dy + \log(\underline{h}) = \eta_2(\underline{\theta}),$$

where the inequality follows from LR dominance. Therefore, by continuity at least one crossing point exists on the open interval $(\underline{\theta}, \tilde{\theta})$. Let $\ddot{\theta}$ denote the maximum point at which $\eta_1(\ddot{\theta}) = \eta_2(\ddot{\theta})$, and note that since the two functions are negatively sloped and η_1 crosses η_2 from above (moving in the leftward direction) at $\ddot{\theta}$, the following must be true

when $\theta = \ddot{\theta}$:

$$\begin{aligned} \eta_1'(\theta) = &- \frac{f_1(\theta)}{\theta + F_1(\theta) - 1} > -\frac{f_2(\theta)}{\theta + F_2(\theta) - 1} = \eta_2'(\theta) \\ &\Leftrightarrow \quad f_1(\theta)\left[\theta + F_2(\theta) - 1\right] < f_2(\theta)\left[\theta + F_1(\theta) - 1\right] \end{aligned} \tag{A.5}$$

Now consider approaching $\underline{\theta}$ from above, beginning at $\ddot{\theta}$ and once again moving leftward. Since $\ddot{\theta}$ is to the left of the density crossing $\tilde{\theta}$, then by the LR dominance property, we know that if we begin at $\ddot{\theta}$ and approach $\underline{\theta}$ from above, then $f_2(\theta)$ becomes steadily larger relative to $f_1(\theta)$ as we move leftward. This also implies that $[\theta + F_1(\theta) - 1]$ becomes steadily larger relative to $[\theta + F_2(\theta) - 1]$ in the leftward direction (beginning from $\ddot{\theta}$) as well.[29] Therefore, the ordering between the right-hand and left-hand sides of inequality A.5 only becomes more pronounced as we move leftward from $\ddot{\theta}$. From this fact it follows that $\eta_1'(\theta) > \eta_2'(\theta)$ for each $\theta \in [\underline{\theta}, \ddot{\theta}]$, and the crossing point $\ddot{\theta}$ is therefore unique. ■

We can build on the above result to prove comparative statics for different seat allocation rules just as we did before. Corollary 5 and Corollary 6 below build on Theorem 4 in precisely the same way that Corrolary 2 and Corollary 3 build on Theorem 1 above. In particular, Corollary 6 follows naturally from the proof of Theorem 4: we now know that $\eta_{\mathcal{D}}'(\theta) < \eta'(\theta) < \eta_{\mathcal{A}}'(\theta)$ for each $\theta \in [\tilde{\theta}, \overline{\theta}]$, and since the log transformation preserves ordering, it follows that $h_{\mathcal{D}}^{*\prime}(\theta) < h^{*\prime}(\theta) < h_{\mathcal{A}}^{*\prime}(\theta)$ on that same interval as well. This also produces a slight strengthening of the result since the derivatives are strictly ordered at the density crossing $\tilde{\theta}$. Together, the final two results demonstrate that the model still predicts a large fraction of the disadvantaged group increasing investment under RQ:

Corollary 5. *Let HC production costs and gross utility be linear in h, so that $c(h;\theta) = \theta(h - \underline{h})$, and $u(s,h) = sh$. Moreover, assume $F_{\mathcal{D}}$ LR dominates $F_{\mathcal{A}}$ and let $\tilde{\theta} \in (\underline{\theta}, \overline{\theta})$ denote the unique crossing of the cost densities where $f(\tilde{\theta}) = f_{\mathcal{A}}(\tilde{\theta}) = f_{\mathcal{D}}(\tilde{\theta})$. Then there exist crossing points $\ddot{\theta}_{\mathcal{A}}, \ddot{\theta}_{\mathcal{D}} \in (\underline{\theta}, \tilde{\theta})$, such that*

(i) $h_{\mathcal{D}}^*(\theta) < h^*(\theta)$ for all $\theta < \ddot{\theta}_{\mathcal{D}}$ and $h_{\mathcal{D}}^*(\theta) > h^*(\theta)$ for all $\theta > \ddot{\theta}_{\mathcal{D}}$, and
(ii) $h_{\mathcal{A}}^*(\theta) > h^*(\theta)$ for all $\theta < \ddot{\theta}_{\mathcal{A}}$ and $h_{\mathcal{A}}^*(\theta) < h^*(\theta)$ for all $\theta > \ddot{\theta}_{\mathcal{A}}$.

Corollary 6. *Assume the same conditions as in Corollary 5. Then under the RQ policy (relative to the CB policy), typical disadvantaged students—that is, group $\mathcal{D}$ students with costs $\theta \in \{(\ddot{\theta}_{\mathcal{D}}, \tilde{\theta}] \cup T_{\mathcal{D}}\}$—exert higher effort and accumulate more HC. Moreover, if we define $\Delta : (\ddot{\theta}_{\mathcal{D}}, \tilde{\theta}] \cup T_{\mathcal{D}} \to \mathbb{R}$ as the difference on this set between group $\mathcal{D}$ investment under RQ*

[29] To see why, recall that $F_1(\underline{\theta}) = F_2(\underline{\theta}) = 0$ and $f_1(\theta) < f_2(\theta)$ for each $\theta < \tilde{\theta}$. Thus, F_2 becomes steadily larger relative to F_1 when moving in the rightward direction from $\underline{\theta}$ to $\ddot{\theta}$, which is the same as saying that F_1 becomes steadily larger relative to F_2 when moving in the leftward direction from $\ddot{\theta}$.

FIGURE 10. NUMERICAL EXAMPLES:
Cost Densities

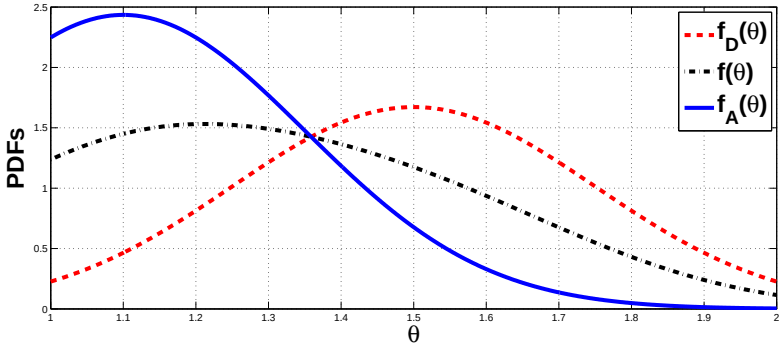
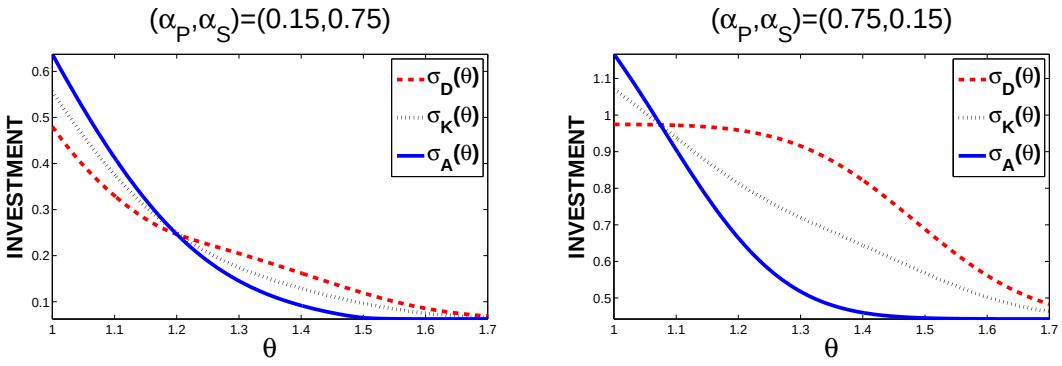


FIGURE 11. NUMERICAL EXAMPLES:
Investment, CB vs AA



(A) Example 2 (B) Example 3

versus CB, or $\Delta(\theta) \equiv \left(h^*_{\mathcal{D}}(\theta) - h^*(\theta)\right)$*, then* $\Delta(\theta)$ *is strictly positive and attains a maximum on the interval* $(\ddot{\theta}_{\mathcal{D}}, \tilde{\theta})$.

7.2.2. ***Example 2: Curved Utility.*** Proving qualitative predictions for the entire Cobb-Douglas utility family is more difficult, so in this section we present some simple numerical examples to further illustrate the model. $\Theta_{\mathcal{D}}$ and $\Theta_{\mathcal{A}}$ once again follow normal distributions truncated to a common support $[\underline{\theta}, \bar{\theta}] = [1, 2]$. They both have the same variance parameter, $\sigma_{\mathcal{A}} = \sigma_{\mathcal{D}} = 0.25$, but different mean parameters, $\mu_{\mathcal{A}} = 1.1$ and $\mu_{\mathcal{D}} = 1.5$, which ensures LR dominance (see Figure 10). The group $\mathcal{D}$ mass is $\delta = 0.5$.

We numerically solve for equilibria in two examples. In the first (Example 2), HC factors relatively heavily into match utility, with $\alpha = 0.15$ and $\beta = 0.75$. In the second (Example 3), students care less about their own HC and more about the quality of the institution they attend, with $\alpha = 0.75$ and $\beta = 0.15$. Equilibrium outcomes are summarized in figure 11.

The patterns arising from these examples are similar to predictions under the linear model. There continues to exist a single crossing point between the investment functions of the two groups under AA vs. CB. Under AA, the highest ability disadvantaged students and the lower ability advantaged students decrease HC investment, while the lower ability disadvantaged students and the higher ability advantaged students increase HC investment. In both examples average investment in group $\mathcal{D}$ rises; for group $\mathcal{A}$ it rises in Example 2 and falls in example 3.