Building Holistic Human Capital: Long-run Randomized Evidence on Athletic and Academic Mentorship*

Noam Angrist† Jonathan Tebes‡§

July 3, 2026

Abstract

We evaluate the long-run effects of a mentorship program combining athletic coaching in swimming and rowing with academic tutoring for disadvantaged high school students. Program access was determined by lottery in oversubscribed semesters, enabling randomized evaluation. We track administrative outcomes for more than a decade after application, including high school achievement and graduation, post-secondary attainment, and leading labor market indicators from credit-bureau records. We find large and persistent effects: admission offers increased high school GPA by 0.48–0.58 standard deviations and graduation by 17–18 percentage points. Bachelor's degree completion rose by 17–21 percentage points through greater persistence, not increased college entry. In early adulthood, labor market indicators also improved substantially: an aggregate credit index rose by 0.37–0.38 standard deviations, while a poor-financial-health index fell by 0.40–0.43 standard deviations. The program was unusually low-cost because it leveraged college-student volunteers and university facilities, placing it among the most cost-effective youth interventions. Mechanism analyses show persistent gains in both cognitive and non-cognitive skills, suggesting large and lasting returns to programs that build holistic human capital.

Acknowledgments: We thank the students, mentors, and tutors who participated in the program. We are grateful for excellent research assistance from Alex Bacon, Catherine Lawlor, and Annabelle Sauer, and for formative support early in this project from CJ Angrist and Camila Caballero. We are grateful for comments from Caroline Chin, Bill Evans, Chloe Gibbs, David Phillips, Patrick Turner and seminar participants of ND Applied Micro Seminar and the ND PIER Seminar. We thank founding leaders of the program – Ron Rosenberg and Camila Caballero – the founding cohort of mentors including Corinne Carland and Kimberly Aziz, as well as the next generation of leaders: Alice Huang and Shilpa Agarwal. Angrist is a co-founder of Amphibious Achievement and set up the randomized lottery structure to facilitate future rigorous evaluation. Tebes was the former head of data and research, and was a mentor and tutor in the program. Tebes is especially grateful to Amy Finkelstein for her support and guidance on his undergraduate thesis from which this paper originated. We dedicate this research to the memory of Camila Caballero, a beloved mentor, teacher, and friend whose generosity and care shaped the Amphibious Achievement community.

†Academic Director of the What Works Hub for Global Education, Oxford University & Co-founder, Youth Impact

‡Department of Economics, University of Notre Dame & The Wilson Sheehan Lab for Economic Opportunities (LEO).

§Corresponding author: jtebes@nd.edu.

I Introduction

Poverty exposes children to a bundle of disadvantages that reinforce each other over time. Children from poor families are more likely to attend lower-performing schools, live in lower-opportunity neighborhoods, face higher-crime environments, and grow up in households with fewer parental resources (Chetty and Hendren, 2018; Oreopoulos et al., 2017). These early disadvantages translate into large and persistent achievement gaps. American high school students born to parents near the bottom of the income distribution score roughly one standard deviation below students born to parents near the top, a gap that has remained strikingly stubborn for three decades (Hanushek et al., 2019). A central challenge for social policy is therefore to identify interventions that address several barriers at once and shift long-run human-capital trajectories while remaining inexpensive enough to scale.

Two classes of interventions have shown particular promise. The first is intensive academic support. High-dosage tutoring is among the most consistently effective interventions for raising academic achievement, but it is expensive, personnel-intensive, and difficult to implement at scale (Davis et al., 2017; Nickow et al., 2020; Guryan et al., 2023; Kraft et al., 2022; List, 2022; Carlana and La Ferrara, 2025). A second is strengthening non-cognitive skills, such as conscientiousness, persistence, and confidence. Non-cognitive skills have been shown to shape both educational attainment and labor-market outcomes (Jackson, 2018; Alan et al., 2019). Conceptually, interventions that combine academic support with non-cognitive skill development may generate complementarities: for example, stronger persistence can increase the returns to academic effort, while improved academic performance can reinforce persistence and educational investment (Cunha and Heckman, 2007).[1]

This paper reports evidence from a randomized evaluation of a novel enrichment program that combined athletic coaching, academic tutoring, and mentorship to build both cognitive and non-cognitive skills, fostering holistic human capital among disadvantaged students. We present long-run results from more than a decade after application, drawing on a wide array of administrative outcomes that span high school achievement and graduation, college attainment, and leading labor-market indicators measured with credit-bureau records.

The program, Amphibious Achievement, served disadvantaged students from public high schools across Boston through weekly programming on MIT's campus. Each Sunday session combined two hours of swimming or rowing instruction, lunch with mentors and peers, and two hours of academic tutoring in small groups. Across both the athletic and academic components, students set goals, tracked progress, and received repeated feedback from mentors. The program was designed to make effort and improvement visible: students competed

[1] We use 'non-cognitive' as conventional shorthand for socio-emotional skills.

against their own prior performance, aimed to set new personal records ('PRs') each week, and practiced applying the same goal-setting framework to academic work. By matching high school students with college-student mentors on a selective university campus, the program also aimed to expand students' information, aspirations, and sense of feasible educational pathways. Reliance on volunteer college-student mentors substantially reduced labor costs and enabled the program to operate at unusually low cost, strengthening scalability.

We estimate causal effects using admissions lotteries from oversubscribed semesters. Students applied to participate in each academic semester, and in four oversubscribed semesters, admission offers were randomly assigned conditional on lottery strata. We focus on the first lottery each student entered and estimate intention-to-treat effects of being offered a spot in the program. We also estimate local average treatment effects (LATE) using initial lottery assignment as an instrument for ever attending Amphibious Achievement. This distinction is important because many initial lottery losers later reapplied and participated: 31 percent of control-group students eventually attended the program, compared with more than 98 percent of initial lottery winners. As a result, LATE estimates are roughly 46 percent larger than the corresponding ITT estimates.

We assemble an unusually rich dataset capturing a wide array of human capital outcomes over the long run. We combine program records with several administrative data sources that follow students from high school through their late twenties. High school transcript records measure short-run academic performance, while Boston Public Schools administrative records measure cumulative GPA, credits earned, graduation, and standardized test scores. National Student Clearinghouse records measure post-secondary college enrollment, persistence, degree receipt, and institutional quality. We merge Experian credit-bureau records from Q3 2021 through Q3 2025, capturing leading labor-market indicators and early-adult financial outcomes, including credit access, credit scores, asset-related borrowing, financial distress, and student-loan debt. We focus on the final eight quarters, when applicants are in their late twenties and predicted earnings measures are available. We combine these measures into two standardized summary indices: an aggregate credit index and a poor financial health index, which summarizes adverse left-tail outcomes such as low predicted income, poor credit, delinquency, collections, and student-loan debt. Together, these data allow us to trace the program's effects on long-run human capital trajectories, from immediate academic achievement to post-secondary college completion to labor market indicators and financial health.

Results show that an offer of admission generated large improvements in high school outcomes. Lottery winners earned semester-level GPAs that were 0.43–0.49 standard deviations higher than those of lottery losers, corresponding to an increase of roughly 0.34–0.40

grade points. These gains persisted through high school: cumulative GPA rose by 0.48–0.58 standard deviations, credits earned increased by 1.3-1.6 courses, and high school graduation increased by 17–18 percentage points. The implied effects of program participation are larger. Among program participants ('compliers'), ever attending the program increased semester-level GPA by 0.68 standard deviations, equivalent to moving from roughly a C to a B-minus average, and increased high school graduation by 26 percentage points.

These high school gains translate into substantial improvements in post-secondary attainment. An offer of admission increased a standardized post-secondary schooling index by 0.45–0.49 standard deviations and raised bachelor's degree completion by 17–21 percentage points. Scaling by compliance, our LATE estimate implies that attending the program closes 80% of the low- versus high-SES gap in bachelor's degree receipt (Cohen et al., 2024). These gains appear to operate through persistence rather than college entry: lottery winners were no more likely to enroll in a bachelor's-degree program, but were substantially more likely to complete one. This pattern highlights the potential role of non-cognitive skills in helping students persist through demanding educational pathways.

An offer of admission also improved leading labor market outcomes in early adulthood. Across credit bureau outcomes, lottery winners scored 0.37–0.38 standard deviations higher on an aggregate credit index and 0.40–0.43 standard deviations lower on a poor financial health index. Individual outcomes all move in a consistent direction. Lottery winners were less likely to have poor credit, bills at least 30 days delinquent, student-loan debt, or other adverse credit-market outcomes. These gains are notable because credit-score gaps typically emerge early, widen over adulthood, and remain sticky throughout the lifecycle (Bakker et al., 2025). Our results show that a high school intervention can shift labor market trajectories before disadvantage compounds: program participation generated a credit-score gain equal to 31 percent of the low- versus high-SES credit-score gap.

Our results are robust to several checks: lottery winners and losers look similar at baseline, omnibus balance tests fail to reject equality across groups, and randomization-inference p-values closely track conventional asymptotic p-values. Estimates are also stable across specifications, including our preferred LASSO-selected controls and richer models that add all available baseline covariates. A joint randomization-inference test across our main outcomes rejects the null of no treatment effects with a p-value of 0.003, suggesting that the overall pattern of results is unlikely to have arisen by chance.

In terms of mechanisms, several results are consistent with complementarities between cognitive and non-cognitive skill formation. First, we find large effects on both GPA and standardized test scores on the Massachusetts Comprehensive Assessment System (MCAS), suggesting that the program improved both non-cognitive skills, more heavily weighted in

GPAs (Jackson, 2018), and cognitive skills, more heavily weighted in standardized tests. Second, gains were larger in STEM courses than in humanities and social science courses, and lottery winners were substantially more likely to earn Bachelor of Science degrees. This pattern is consistent with mentors at a technical university influencing not only students' academic skills, but also their aspirations and perceptions of feasible educational pathways. Third, post-secondary effects are weak on college entry, but strong on college persistence, likely a function of combined cognitive and non-cognitive skills.

One of the most striking elements of the program is its cost-effectiveness. Amphibious Achievement costs just $170 per enrolled student per year in 2025 dollars, well below prevailing tutoring models (Guryan et al., 2023). This is driven in part by the fact that labor costs – often the largest expense of social programs – approach zero by leveraging college volunteers. But the program's high returns reflect not only low costs, but also large and persistent impacts. Across two approaches to estimating present-discounted lifetime earnings gains from program participation, we value private benefits to participants at $113,000 to $275,000 per participant and public benefits from higher tax revenue at $26,000 to $63,000 per participant. These gains yield some of the highest Benefit-Cost Ratio (BCR) estimates in the literature (Angrist et al., 2023). Large effects seem to be driven by tackling multiple human capital constraints at once. Moreover, by including athletics, not just academics, the program selects low-achieving students who need the program most. Consistent with this notion and related theoretical predictions on the marginal applicant (Walters, 2018), we find that effects are concentrated among students with weaker predicted long-run trajectories at baseline, for whom enrichment programs are likely to be especially valuable.

This paper makes several contributions. A frontier of the education literature is focused on identifying lower-cost, scalable tutoring programs (Davis et al., 2017; Nickow et al., 2020; Kraft et al., 2022; List, 2022; Angrist et al., 2022; Carlana and La Ferrara, 2025). We identify an unusually effective, low-cost intervention for disadvantaged high school students, with long-run benefits comparable to some of the highest-return social programs in the literature, such as the Perry Preschool Project (Hendren and Sprung-Keyser, 2020; Kautz et al., 2014). Observed effects are even larger than some of the leading tutoring approaches (Guryan et al., 2023), suggesting that combining tutoring with athletics and mentoring support could be high leverage. Low costs, in addition to the high returns, make the program model especially promising for scale. College students are a ripe and ready pool of tutors at scale: roughly 15 million students attend four-year post-secondary institutions, and about one-third already volunteer each year (National Center for Education Statistics, 2026; U.S. Census Bureau, 2024).[2] Universities could serve as platforms for delivering low-cost, high-impact support to

[2]We multiply the number of students attending a four-year institution, including both undergraduate and

disadvantaged adolescents in surrounding communities.

Second, this paper contributes to a growing literature showing that non-cognitive skill development in adolescence can complement cognitive skill formation and shape long-run human capital accumulation. Long-run effects are challenging to attain even for some of the most effective programs (Dobbie and Fryer, 2020). Our findings help reconcile recent studies documenting gains in college attainment despite fadeout in short-run measures of cognitive skills (Gray-Lobe et al., 2023; Cohodes and Pineda, 2026). Persistent effects may depend on the development of 'holistic human capital,' where non-cognitive skills are a crucial complement to cognitive skills for long-run human capital accumulation (Jackson, 2018). The intervention we study emphasized both non-cognitive skills such as goal-setting and perseverance through athletics, in addition to fostering cognitive skills through academic sessions. Influencing both margins simultaneously resulted in persistent long-run effects. These findings complement research on adolescent interventions that combine academic and socioemotional support (Cook et al., 2014; Heller et al., 2017) and evidence that athletics may foster growth in valuable non-cognitive skills (Stevenson, 2010; Heckman et al., 2025).

Third, we contribute to a growing literature on mentorship as a vehicle for social mobility among disadvantaged teenagers. Recent high-profile randomized studies in Germany show that mentorship can improve educational outcomes for low-SES students: Falk et al. (2020) find that exposure to high-academic-track mentors closes roughly one-third of the SES gap in entry into the high track, while Resnjanskij et al. (2024) show that university mentors improve short-run skills and behaviors predictive of later labor-market success. These short-run effects are promising, yet studies with long-run outcomes remain scarce and effects to date have been mixed, as summarized in Tables A.5 and A.6. The closest analogue to our program is *Pathways to Education*, which combined tutoring, mentoring, and financial support for high school students in Toronto public housing. Oreopoulos et al. (2017) find large effects on high school completion and post-secondary enrollment using differences-in-differences analysis. By contrast, the *Quantum Opportunity Program* in urban American high schools generated small short-run academic gains that did not translate into long-run improvements in bachelor's degree receipt or employment in early adulthood (Rodriguez-Planas, 2012). These mixed findings underscore the need for long-run randomized evidence on whether mentorship programs can generate lasting gains in human capital and economic well-being. Our paper fills this gap with long-term experimental estimates on post-secondary attainment and rich administrative indicators of labor-market success in adulthood.

The rest of the paper proceeds as follows. Section II describes the Amphibious Achieve-

graduate students, by the volunteer rate from the Census Bureau's CPS Civic Engagement and Volunteering Supplement among respondents enrolled in college or university.

ment program. Section III describes the program records and administrative data sources. Section IV presents the randomized research design. Section V reports effects on high school outcomes, post-secondary college outcomes, and leading labor market indicators in adulthood. Section VI discusses mechanisms, heterogeneity, effects of participation, as well as cost-effectiveness and welfare implications. Section VII concludes.

II The Amphibious Achievement Program

Amphibious Achievement is a low-cost enrichment program for Boston-area high school students that combines college-student mentorship, athletics-based instruction, and academic tutoring. The program began in 2010 and served students from Boston public high schools.[3] Students applied to participate each academic semester, and in oversubscribed semesters admission was determined by lottery. Admitted students, referred to as Achievers, were offered weekly Sunday programming on MIT's campus.

The program was developed in response to the multifaceted barriers facing students in Boston Public Schools (BPS), reflected in their persistent academic under-performance. In 2014, BPS tenth-graders' MCAS proficiency rates averaged 17 percentage points below statewide rates across English language arts, mathematics, and science, and the district's four-year cohort graduation rate was 66 percent, compared with state and national public-school adjusted cohort graduation rates of 83 and 79 percent, respectively (Massachusetts Department of Elementary and Secondary Education, 2014; Boston Public Schools, 2015; National Center for Education Statistics, 2020). Against this backdrop, Amphibious Achievement sought to improve academic engagement and achievement by combining individualized academic support with structured opportunities to build non-cognitive skills, including persistence, self-regulation, confidence, and future orientation.

The program was designed to leverage complementarities between academic skill formation and the development of non-cognitive skills through athletic coaching (Cunha and Heckman, 2007; Cunha et al., 2010; Kautz et al., 2014). The athletic portion focused on rowing and swimming – typically expensive sports which the program made accessible to low-income students through the facilities on the MIT college campus. During the athletic portion of the program, Achievers could observe measurable improvement over short periods and aimed to exceed their previous week's "Personal Records." Academic tutoring then asked students to apply a similar structure to schoolwork: set a proximal goal, attempt a task, receive mentor feedback, evaluate progress, and try again. This design drew on evidence that persistence toward long-term goals, self-regulation, and motivation can strengthen

[3] Although most students attended one of three large public high schools, enrollment was open to students across BPS, and the sample includes students from ten public high schools.

academic performance (Duckworth et al., 2007; Alan et al., 2019). Athletics was designed as a complement, not a substitute, for academic support. Athletics provided a setting in which effort, feedback, and improvement were highly visible, while tutoring helped channel the same non-cognitive skills and self-regulatory practices into academic performance. This design also echoes the Becoming a Man intervention studied by Heller et al. (2017), which included an athletic component that supported program engagement and provided concrete opportunities to practice emotional regulation, reflection, and persistence.

Each semester consisted of approximately 8-10 five-hour Sunday sessions, with up to 50 hours of programming: 20 hours of athletic instruction, 20 hours of academic tutoring, and 10 hours of informal mentor contact during lunch and transitions. A typical Sunday began with a large-group warm-up and stretching circle, continued with swimming or rowing instruction, followed with lunch, and ended with academic tutoring. All sessions took place at MIT, with most athletic sessions held in the pool and boathouse and academic sessions held in a large classroom. Students mostly traveled to campus by public transit, and the program provided bus and subway passes. Upon arrival, program staff took attendance while Achievers and mentors gathered for the opening circle before breaking into their respective sports. Mentors also used this time to call Achievers who were not present. The opening circle reinforced a central message of the program: by coming to MIT on a Sunday morning to work on improving themselves, Achievers were already demonstrating commitment, persistence, and willingness to take on difficult tasks. Repeated exposure to MIT also made the college environment more familiar.

The program was delivered by volunteer college-student mentors, primarily from MIT, although a few mentors attended Northeastern University. Prospective mentors completed a selective application process that included a written application and interview. Mentor acceptance was competitive and became increasingly prestigious. Once accepted, mentors attended weekly one-hour training sessions, during which they prepared for Sunday programming and received guidance on tutoring, youth mentorship, classroom management, and serving as role models. By matching BPS students with college-student mentors from a selective university, the program sought to leverage role models to enhance aspirations and inspire future ambition (Bernard et al., 2026; La Ferrara, 2016) and also to create cross-SES relationships that could expand students' information, aspirations, and perceived educational opportunities (Chetty et al., 2022).

Goal-setting and progress monitoring were common features of both the athletic and academic components. At the beginning of each semester, Achievers worked with mentors to set short-term and long-term goals and signed commitment cards, informal contracts intended to make these goals salient. Each week, mentors set aside time for Achievers

to review progress, identify obstacles, and define next steps. Both components also used personal records (PRs) to make improvement observable, as shown in Figure A.2, which documents a sample set of personal records in athletics as well as in Figure A.3, which shows a sample academic schedule with periodic PRs. In athletic sessions, Achievers attempted to improve their PR times in swimming or rowing. In tutoring sessions, Achievers completed short academic PR exit tickets that allowed mentors and students to track mastery of the day's material. This common structure was designed to routinize goal-setting, personal reflection, and the connection between effort and improvement.

The use of swimming and rowing was central to the program design. Figure A.4 shows a typical athletic schedule. These sports require specialized facilities and are often inaccessible to urban public school students. Because most Achievers had little or no prior experience with either sport, the athletic component allowed students to learn a new and aspirational skill, receive repeated feedback from mentors, and compete against their prior performance. For semesters with complete PR data, Achievers improved their PR times in swimming and rowing by approximately 15 percent over the course of the semester, on average. Although Amphibious Achievement used swimming and rowing, the model could in principle be adapted to other sports, such as track and field, where individual progress is measurable and lends itself to repeated personal goal-setting. The program's mission was to promote athletic and academic success, reflected in its tagline, "success in and out of the water."

The academic component took place after lunch and consisted of tutoring in mathematics, reading comprehension, and writing. Excerpts from the curriculum are included in the Appendix. Each Sunday, the program selected academic topics. In critical reading sessions, Achievers read abridged versions of articles from outlets such as *The Economist*, the *New York Times*, or other current-events sources and then answered SAT-style questions with their group. In mathematics and writing sessions, a lead mentor first introduced the day's concepts to the full group, and then Achievers worked with mentors on practice problems, discussed mistakes, and completed short exit tickets to assess mastery of the material. Several times per semester, Achievers also participated in group activities, such as math or writing Jeopardy or class debates based on the critical-reading article. These activities preserved the team-based and feedback-rich structure of the athletic component while focusing directly on school-relevant skills.

Achievers were matched to small groups of Achievers and mentors based on preferences and interests reported in their application materials. More experienced mentors tended to be paired with students who had more difficulty in academic activities. On most Sundays, the program operated with approximately a 3:2 ratio of Achievers to mentors, allowing mentors to provide individualized instruction while preserving the team-based structure used during

athletics. This structure also allowed mentors to provide one-on-one or two-on-one tutoring when Achievers entered activities with different levels of understanding.

Participation was voluntary, but admitted students were expected to attend Sunday sessions regularly. Program staff tracked attendance and followed up with absent students through texts and calls, and reached out to parents when students were not responsive. Unlike some comprehensive youth-support programs, Amphibious Achievement did not encourage attendance through financial incentives. Instead, participation was supported by transportation assistance, access to swimming and rowing facilities, meals, sustained time on a college campus, and the opportunity to receive individualized attention from college-student mentors. Program participation was extremely high: over 98 percent of students who won their initial lottery attended at least once, and over 87% attended at least five sessions. On average, lottery winners attended 21 sessions, roughly one and a half years of the program at a 70% attendance rate.

The program was inexpensive to operate. Direct operating costs were low, largely because the program relied on volunteer mentors. Since labor is often the highest cost component of social programs, this design of the program is particularly scalable. The program reduced costs further through discounted or donated food from local restaurants and subsidized use of university facilities. Program records indicate that direct operating costs averaged roughly $170 per enrolled student per year in 2025 dollars. These costs primarily covered transportation, supplies, athletic-facility expenses, and meals not covered by donations. These costs are more than an order of magnitude cheaper than several of the leading effective academic tutoring alternatives in the United States (Guryan et al., 2023).

III Data

We build a rich dataset covering a comprehensive set of outcomes across multiple stages of the life cycle. We conduct intensive longitudinal tracking of students for more than a decade, following students into early adulthood, from their initial application in high school through their late twenties. Administrative records from four data sources cover high-school achievement and graduation, college enrollment and completion, and leading labor market indicators such as credit-market outcomes. These data allow us to track students' human-capital trajectories through post-secondary education as well as to understand labor market success in early adulthood. Program records provide detailed information on baseline characteristics, lottery assignment, and program participation.

A Administrative Records and Outcomes

A.1 High school data: GPA, credits, standardized test scores, and graduation

We use two sources of high school records. First, we collected high school transcript data directly from the three public schools most represented in our sample. These records provide semester-level measures of academic performance and progress, including grades and courses taken.[4] Because the transcripts were collected in early 2014, they primarily measure short-run academic achievement. We use course grades and credits to construct semester-level GPA and a standardized GPA measure that is normalized by the control-group mean and standard deviation.[5]

Second, we obtained Boston Public Schools administrative records in 2024 to measure longer-run secondary-school outcomes. These records include cumulative weighted GPA, credits earned, high school graduation, Massachusetts Comprehensive Assessment System (MCAS) scores, and demographic characteristics. Because the records were provided in de-identified form, we could not link them to the full set of baseline controls. We therefore instead show robustness when controlling for available demographic characteristics, including race and sex, as well as linked historical eighth-grade achievement measures, including GPA, attendance, and MCAS scores. The transcript and BPS records provide complementary coverage of secondary-school outcomes. The transcript sample is limited to students from the three most represented schools, while the BPS records exclude students from one highly represented charter school. As a result, the transcript and BPS sources omit 25 percent and 33 percent of lotteried students, respectively, but together cover 92% of the sample.

A.2 Post-secondary outcomes: college enrollment, graduation, selectivity

We link students to National Student Clearinghouse (NSC) records to measure post-secondary college enrollment and completion. The NSC receives regular submissions from nearly 3,600 colleges and universities, which together enroll more than 97% of students in public and private U.S. post-secondary institutions (National Student Clearinghouse, 2026). Prior work documents high overall coverage, especially for public two- and four-year institutions, while noting potential measurement issues from institutional non-coverage, FERPA-related directory-information blocks, and matching error (Dynarski et al., 2015). We linked students to NSC records using name and date of birth. In the two cases with multiple plau-

[4]These data were collected as part of Tebes's undergraduate senior thesis (Tebes, 2014). One school reports annual rather than semester-level grades before Fall 2013. For these observations, we assign the annual grade to each semester, introducing measurement error that should attenuate estimated effects.

[5]As explained in the next section, we standardize outcomes using a weighted control-group mean and standard deviation that reweights the control group to match the strata composition of the treatment group.

sible matches, we adjudicated matches by hand using publicly available information online. The NSC data were pulled in 2025, when the median student was 28 years old and student age ranged from 25 to 31 years old.

An outcome of interest is bachelor's degree receipt.[6] We also examine whether students enrolled in a bachelor's program but did not complete their degree, completed an associate's degree, or enrolled in but did not complete an associate's or bachelor's degree.[7] The enrolled-but-did-not-complete outcomes help distinguish between effects on access and effects on persistence along high-stakes post-secondary margins.

We also link institution names to the average *U.S. News & World Report* rankings for Best National Universities and Best Liberal Arts Colleges from 2016 to 2022. These rankings provide proxy measures of institutional quality and selectivity. We construct indicators for attending a top-100 national university or top-50 liberal arts college, and for attending a top-30 national university or top-5 liberal arts college.

Our other post-secondary outcome of interest is a standardized summary index constructed following Kling et al. (2007). The index averages standardized indicators for bachelor's degree completion; enrollment in a bachelor's program without completion, coded negatively; enrollment in an associate's degree without completion, coded negatively; attendance at a top-100 national university or top-50 liberal arts college; and attendance at a top-30 national university or top-5 liberal arts college. Each component is normalized using the control-group mean and standard deviation. We exclude the receipt of an associate's degree because its interpretation is ambiguous: it can represent reduced attainment for students on the margin of a four-year degree but positive attainment for students on the margin of no post-secondary education (Mountjoy, 2022).

A.3 Leading labor market indicators: credit outcomes and financial health

We use Experian credit bureau records to measure leading labor market indicators in early adulthood. These records cover more than four years of credit-bureau activity, from Q3 2021 through Q3 2025, and provide measures of labor-market standing and financial well-being. Our primary sample focuses on the final two years, when median applicant age

[6] One control student obtains a Doctor of Pharmacy degree by attending a six-year undergraduate and graduate program. We mark this student as completing a bachelor's degree; not doing so would *increase* the reported ITT effect by 1 percentage point.

[7] NSC records directly identify whether most enrollment spells are associated with bachelor's- or associate's-degree programs. When program-level information is missing, we classify enrollment using institutional characteristics, including whether the institution is a two- or four-year college and the degrees it typically awards.

ranges from 27 to 29 and predicted earnings information is available.[8] We link students using name, date of birth, and address information, including updated addresses from public records when available. The linked records capture credit access, credit scores, predicted wage earnings, asset-related borrowing, financial distress, and student-loan debt.

Credit access is measured by an indicator for having an Experian record. Credit quality is captured by VantageScore credit scores and indicators for prime credit, defined as a score above 660, and poor credit, defined as a score below 600. We also use Experian's Income Insight W2 measure, which predicts annual earnings using a credit-based model trained on merged credit-bureau and verified W-2 income data (Experian, 2026). We interpret this variable as a noisy but complementary measure of financial standing. Both predicted annual earnings and credit scores are also reported in logs to improve precision. We define an indicator for left-tail predicted earnings as predicted earnings below $30,000.

Indicators for having a mortgage or auto loan capture access to credit for major durable goods and asset-building purchases. Financial distress is measured by indicators for having a positive collections balance, any bills at least 30 days delinquent, and any bills at least 90 days delinquent. Student-loan debt is measured as the total balance on non-derogatory student-loan tradelines reported in the prior three months, including joint, indeterminate-status, and deferred loans (Experian, 2019).

Our primary financial outcomes include two standardized summary indices. First, we present an aggregate credit index which combines standardized versions of all credit-bureau outcomes, with signs oriented so that higher values correspond to better financial health and labor market outcomes.[9] Second, the poor financial health index focuses on adverse outcomes and includes standardized indicators for not having an Experian record, having poor credit, having predicted earnings below $30,000, having a positive collections balance, having any bills at least 30 days delinquent, having any bills at least 90 days delinquent, and having any student-loan balance.

B Program Records

Program records provide information on lottery assignment, program participation, and baseline characteristics. We collect administrative records for each oversubscribed admissions lottery, which include lotteries run in Fall 2011, Fall 2012, Spring 2013, and Fall 2013. These records identify all lottery entrants, whether each applicant was randomly selected to receive an enrollment offer, and the variables used to construct lottery weights.

[8] In an appendix table, we demonstrate results are both quantitatively and qualitatively comparable when using the full sample.

[9] For credit scores and predicted annual earnings, we include standardized *logged values* in the index rather than standardized raw values to reduce residual variance.

Our analysis sample is defined by the first lottery each applicant entered. This lottery sample includes 130 students, with roughly half of students receiving a randomized offer. Because male applicants received twice the lottery weight, offer rates were higher for male applicants than for female applicants. Our preferred specifications account for these differences by including lottery-strata fixed effects, defined by lottery cohort, student sex, and sibling status. After reweighting the control group to match the treatment-group strata composition, the male share is balanced across study groups.

Although our sample is modest, the study builds on a tradition of influential randomized evaluations that follow human-capital accumulation over the long run with similarly moderate sample sizes, including the Perry Preschool Project (Kautz et al., 2014) and the Jamaica early-childhood stimulation experiment (Gertler et al., 2014). We address potential concerns about sample size and finite-sample inference in several ways, including by reporting randomization-inference p-values that leverage the experimental assignment mechanism directly (Young, 2019).

In addition to lottery records, we use semester-level attendance records from Fall 2011 through Spring 2019 to measure realized program participation. These data provide information on participation for every lotteried student, including the total number sessions attended, allowing us to observe take-up and treatment intensity for both lottery winners and losers.

Finally, initial application surveys, collected before lottery assignment, measure students' baseline characteristics, including sex, race, grade, self-reported GPA, and comfort with academic subjects. Beginning in Fall 2012, the application also collected information on grade repetition, parental education, college expectations, single-parent household status, academic engagement, and self-belief.[10] We also geocode student home addresses from the initial application to measure pre-treatment neighborhood characteristics at the Census tract level. We link these tracts to measures of residential racial composition and population from the 2010 Census, measures of median household income, poverty rate, and educational attainment from the 2010 American Community Survey five-year file, and a measure of land area from the 2010 Census Gazetteer file (Schroeder et al., 2025; U.S. Census Bureau, 2010).

[10]Academic comfort captures reported comfort with standardized testing, math, reading, writing and grammar, and thesis construction. Academic engagement captures reported concern for grades, school performance, teacher opinions, pride in academic success, interest in out-of-class learning, and perceived grade fairness. Self-belief captures affinity for challenges, self-satisfaction, self-attitude, and self-respect.

IV Research Design

There were four semesters where the program was oversubscribed and lotteries were conducted to determine the opportunity to enroll in the program. To equalize the sex-ratio of attendees, male applicants were given twice the weight as female applicants. In similar fashion, later lotteries doubled the chances of students who had previously applied to the program or those who had a sibling already enrolled. Conditional on these factors, however, the offer to participate was randomly chosen. While students may apply to the program multiple times, we focus on the outcome of the first lottery that a student entered.[11]

A Intention to Treat (ITT)

Information from these lotteries allows us to implement a straightforward randomized controlled trial design, randomized at the individual level. Our primary specification estimates the *intention-to-treat* (ITT) effect or the impact of being offered a spot in Amphibious Achievement on student outcomes. As given in Equation 1, for student i, let post-assignment outcome Y_i be a function of first lottery assignment (Z_i), baseline characteristics X_i, and lottery-specific strata ($\alpha_{1s(i)}$). Strata fixed effects interact lottery indicators with indicators for each characteristic that affected a student's chance of selection (i.e., male, has sibling enrolled). Within each strata bin, students faced an equal chance of being offered a spot in the program.

$$Y_i = \beta Z_i + \delta_1 X_i + \alpha_{1s(i)} + \epsilon_{1i} \tag{1}$$

The ITT effect is captured by β in Equation 1. It measures the regression-weighted average difference in outcomes between first-lottery winners and losers within lottery strata. Our preferred specification includes school fixed effects to account for school-level differences in quality, grading standards, and other institutional factors, as well as self-reported GPA at the time of initial application and an indicator for missing such GPA information.[12] Baseline GPA is the only covariate selected by LASSO and substantially improves precision because it has strong explanatory power across outcome domains. For transparency, we also report results using the full set of available baseline controls.

We report robust standard errors when there is one observation per student and cluster standard errors at the student level when outcomes are observed over multiple post-assignment periods, such as semester-level GPA records or quarterly credit-market records

[11] 92% of lotteried students applied only once, 7% applied twice, and 1% applied three times.

[12] For the 9% of applicants who did not report GPA on their initial application, we set self-reported GPA equal to the sample mean to avoid dropping these observations. The missing-GPA indicator accounts for this imputation.

in adulthood. We also assess the sensitivity of our inference using randomization-inference p-values. Specifically, we compare the observed treatment t-statistic to the distribution of simulated t-statistics obtained by re-randomizing treatment 1,000 times and re-estimating Equation 1. We find that the resulting randomization-inference p-values are both quantitatively and qualitatively similar to conventional asymptotic p-values.

B Local Average Treatment Effect (LATE)

We next estimate an instrumental-variables model that uses a student's first lottery assignment (Z_i) as an instrument for program participation, defined as attending at least one Amphibious Achievement session (P_i). The first-stage and second-stage equations are:

$$P_i = \pi Z_i + \delta_2 X_i + \alpha_{2s(i)} + \epsilon_{2i} \tag{2}$$

$$Y_i = \gamma P_i + \delta_3 X_i + \alpha_{3s(i)} + \epsilon_{3i}. \tag{3}$$

Following Angrist et al. (1996), the causal interpretation of γ relies on the standard IV assumptions: random assignment, relevance, exclusion, and monotonicity. Random assignment is guaranteed by the lottery design, and relevance follows from the large first-stage effect of winning the initial lottery on program participation. The exclusion restriction requires that winning the lottery affects outcomes only through program participation. That is, the offer has no average effect on students who do not attend, and losing the lottery does not affect outcomes except through its effect on participation. Monotonicity requires that winning the lottery weakly increases the probability of participation for all applicants.

Under these assumptions, γ identifies the Local Average Treatment Effect (LATE): the average causal effect of program participation for compliers, defined as students induced to participate by winning their first lottery. This distinction is important because 31% of the control group reapplied and eventually participated in the program despite losing their initial lottery.[13] If always-takers stand to benefit more than compliers, as their demand for the program may indicate, then the LATE we estimate is less than the average program effect on all students who participated, or the treatment-on-treated effect. By defining participation as having ever attended Amphibious Achievement, we provide a conservative estimate of the LATE, as students who only attend the program once are counted as having participated.[14]

[13] Also, note that γ is simply the intention-to-treat parameter (β) divided by the regression-adjusted take-up rate (π).

[14] If instead, for example, we had assumed students who attended four or fewer sessions did not benefit from the program, then our LATE estimates would be 19% larger.

C Balance and Program Participation

Table I displays mean characteristics and survey responses at the time of initial application for each student. Column 1 reports baseline means for students who lost their first lottery (Control), while Column 2 reports means for students who won their first lottery (Treatment). Using Equation 1, Column 3 provides the p-value for the coefficient on the treatment indicator.

Table I shows that the treatment and control groups are well balanced on baseline characteristics. As expected, males were more likely than females to win their first lottery, yielding a higher share of males in the treatment group. Because student sex was used to define lottery strata, any differences for the male indicator are absorbed by the strata fixed effects.[15] Otherwise, the two groups are similar across a broad set of baseline characteristics. These include demographic and family background measures, such as race and ethnicity, grade level, school attended, and single-parent household status, as well as baseline academic characteristics, including prior grade repetition and parental college attendance, and Census tract characteristics linked to applicants' home addresses. Notably, self-reported baseline GPA is nearly identical across groups and, because of its strong explanatory power, we will include it in our preferred specification as a control variable.

Consistent with random assignment, only one of the twenty-five covariates differs significantly at the 5 percent level, namely the academic comfort index. In contrast, several other differences, including the academic engagement index and neighborhood measures such as poverty rate, median household income, and Bachelor's degree share, point in the opposite direction. We find no evidence that the covariates jointly predict initial lottery assignment. A standard F-test fails to reject joint significance ($p = 0.556$), and a stacked regression likewise indicates no systematic differences across groups. If anything, the stacked regression indicates that lottery winners came from slightly more disadvantaged backgrounds at baseline ($\beta = -0.043$ SD; $p = 0.514$).

Figure I and Table A.1 show that initial lottery assignment generated large differences in program participation. More than 98 percent of treatment-group students ever attended the program, compared with 31 percent of control-group students. Winning the first lottery also roughly tripled the average number of sessions attended, from 7 to 21, as well as the likelihood of attending at least 10 sessions. These differences also persist in the upper tail of the participation distribution: 42 percent of initial lottery winners attended at least 20 sessions, compared with only 18 percent of initial lottery losers.

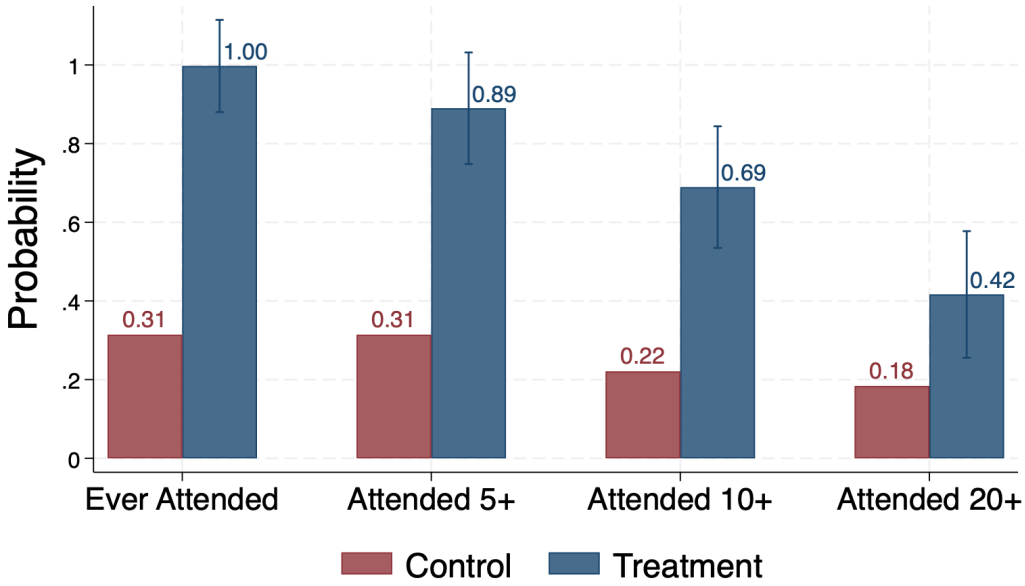
[15] When estimating adjusted differences for high-school indicators, we omit school fixed effects because these indicators are collinear with them.

Table I: Baseline Balance

	Control Mean	Treatment Mean	P-value
	(1)	(2)	(3)
Panel A: Demographics and Personal Background			
Male	0.284	0.444	—
Black	0.567	0.556	0.625
Hispanic	0.254	0.238	0.940
White or other ethnicity	0.179	0.206	0.478
Freshman	0.552	0.492	0.455
Sophomore	0.328	0.302	0.908
Junior or Senior	0.119	0.206	0.402
Attends High School 1	0.179	0.143	0.321
Attends High School 2	0.403	0.381	0.600
Attends High School 3	0.299	0.333	0.656
Attends other high school	0.119	0.143	0.665
Single parent household^{+}	0.511	0.568	0.379
No close adult family members or friends^{+}	0.085	0.135	0.679
Panel B: Academic Background			
Baseline GPA	3.183	3.075	0.378
Ever repeated a grade^{+}	0.234	0.270	0.507
Parent attended college^{+}	0.511	0.486	0.870
College very likely^{+}	0.809	0.703	0.157
Academic Comfort Index [1-5]	3.553	3.819	0.034
Academic Engagement Index [1-5]$^{+}$	4.617	4.498	0.178
Self-Belief Index [1-4]$^{+}$	3.108	3.189	0.741
Panel C: Census Tract Information			
Population per square mile	19,675	22,085	0.208
Black population share	0.409	0.455	0.444
Hispanic population share	0.222	0.234	0.421
Median household income	$49,153	$41,971	0.069
Poverty rate	0.222	0.262	0.067
Bachelor's degree share	0.276	0.233	0.182
Panel D: Tests on Joint Distribution			
F-Test			0.556
Stacked Regression		-0.043	0.514
		(0.065)	

Notes: Columns 1 and 2 report mean baseline characteristics of students who lose or win their initial lottery, respectively. Column 3 reports the p-value for the treatment coefficient from estimating Equation 1 with the baseline characteristic as the outcome, including strata and school fixed effects and robust standard errors. We omit school fixed effects when exploring balance across school indicators. Academic Comfort: comfort with standardized testing, math, reading, writing/grammar, and thesis construction. Academic Engagement: care for grades, school performance, or teacher opinions, pride in academic success, taste for out-of-class learning and perceived grade fairness. Self-Belief: affinity for challenges, self-satisfaction, self-attitude, and self-respect. Using student address information, neighborhood characteristics are from the 2010 Census and 2006-2010 five-year ACS estimates. Panel D reports two omnibus balance tests. The F-test reports the joint significance of all baseline covariates in a regression of treatment assignment on these variables. For the stacked regression, we expand the data by the number of baseline variables, standardize each variable using the control-group mean and standard deviation, and assign a direction based on correlation of a variable with baseline GPA, so that higher values correspond to more favorable baseline characteristics. Missing values are imputed to the sample mean. We then estimate our main ITT specification using the stacked vector of standardized baseline variables as the outcome, including strata and school fixed effects and clustering standard errors at the individual level. $^{+}$ are omitted from joint tests since they were not collected in one lottery sample. $N = 130$; $N^T = 63$, $N^C = 67$.

Figure I: Program Participation by First Lottery Outcome



Notes: This figure plots mean program participation for applicants who lost their first lottery, shown in maroon, and estimated participation for applicants who won their first lottery, shown in navy. Control-group means are weighted to match the distribution of lottery strata among treatment-group applicants. Treatment-group participation is constructed by adding the estimated first-stage effect of an initial lottery win to the weighted control-group mean. The first-stage effect is estimated using Equation 1, which includes lottery-strata and school fixed effects and controls for self-reported baseline GPA, and an indicator for missing baseline GPA. Table A.1 confirms raw group means yield virtually identical results. Whiskers reflect 95% confidence intervals using robust standard errors.

V Results

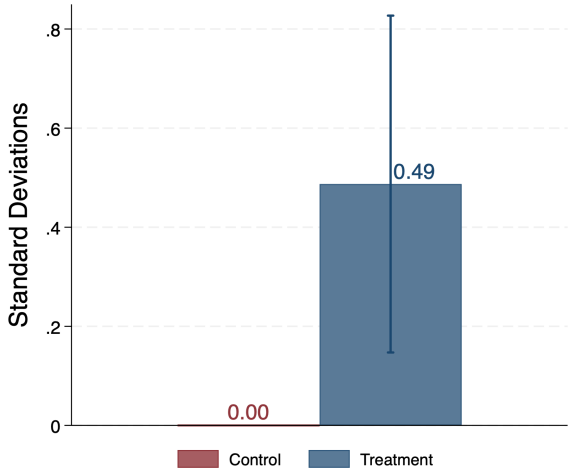
A High School: GPA, test scores, credits, and graduation

We start by studying short-term impacts on high school grade point averages (GPAs), an outcome that reflects both cognitive and non-cognitive skills and is predictive of later-life outcomes (Fryer Jr, 2011; Jackson, 2018). Panel A of Table II uses transcript information to estimate the impact of the opportunity to enroll in the program on students' semester-level GPA, in both its raw and standardized form. Column 1 reports the average outcome of students who lost their initial lottery, weighted to match the strata composition of the treatment group.[16] Column 2 reports intention-to-treat (ITT) effects on GPA controlling for strata and school fixed effects and baseline GPA and an indicator for missing baseline GPA. Column 3 adds the full set of complete covariates from students' initial applications in Panel A and available eighth-grade controls from BPS records in Panel B.

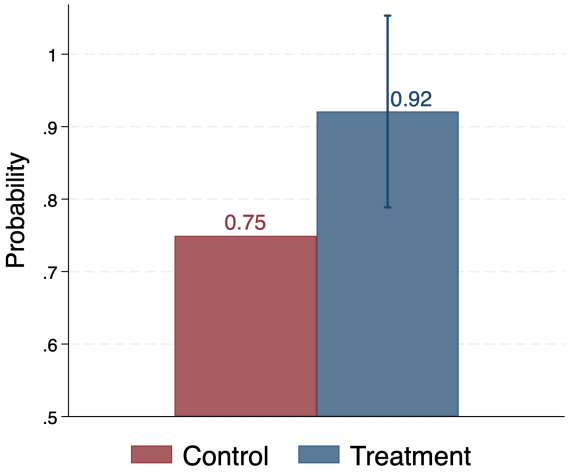
[16]This approach re-weights the control-group mean so that the sex composition matches that of the treatment group. This is important because sex is strongly correlated with baseline achievement.

Figure II: Intention-to-Treat Effects on High School Outcomes

(a) Short-term Semesterly GPA



(b) High School Graduation



Notes: This figure plots ITT effects on high-school outcomes for those who won their first lottery in navy (Treatment), and lost their first lottery in maroon (Control). Panel A displays short-term GPA effects using transcript information and Panel B reports high school graduation. Control-group means are weighted to match the distribution of lottery strata among treatment-group applicants. Treatment-group outcomes are constructed by adding the estimated ITT effect of an initial lottery win to the weighted control-group mean. ITT effects are estimated using Equation 1, which includes lottery-strata and school fixed effects and LASSO-selected controls. Whiskers reflect 95% confidence intervals based on robust standard errors.

We find that the offer to enroll in the program improves student GPAs by 0.4 grade points or 0.49 standard deviations (SDs) ($p < 0.01$), moving from a "C+" to "B-" average. Inclusion of the full set of baseline controls slightly shrinks point estimates to 0.34 grade points and 0.43 SDs but these effect sizes are still highly statistically significant. Panel B confirms that these gains persist throughout high school by exploring cumulative effects in Boston Public School data that were collected seven or more years after the last cohort was expected to graduate high school.[17] On average, by the end of high school, initial lottery winners earned a cumulative GPA that was 0.48-0.58 SD higher than that of initial lottery losers and completed 1.32-1.55 additional courses.

We also find students are 17-18 percentage points more likely to graduate high school ($p < 0.02$). These graduation effects are substantial. The offer of admission increased high school graduation from 75 to 92 percent, moving students from below the state and national averages to 9 and 13 percentage points above them, respectively. The implied LATE estimates are 46 percent larger, suggesting near-universal graduation among compliers. These gains exceed most adolescent support interventions, which typically increase graduation by 5–7 percentage points (Rodriguez-Planas, 2012; Heller et al., 2017), and are larger than the

[17] As discussed in Section III, these data are only available for two-thirds of the sample, as one of the main high schools attended by applicants did not provide records for this research.

Table II: Effects on High School Educational Outcomes: GPA, Credits, Graduation

	Control Mean (weighted)	Intention to Treat	Intention to Treat
	(1)	(2)	(3)
Specification		*Basic*	*All Controls*
Panel A. Short-term Semesterly High School Records			
Overall GPA	2.430	0.401***	0.341**
		(0.146)	(0.134)
Standardized Overall GPA	0.000	0.487***	0.427***
		(0.171)	(0.160)
Panel B. Long-Term High School Outcomes			
Cumulative GPA	2.891	0.353**	0.430**
		(0.160)	(0.180)
Standardized Cumulative GPA	0.000	0.476**	0.580**
		(0.216)	(0.242)
Total Credits Earned	12.725	1.553**	1.318
		(0.696)	(0.836)
Graduated High School	0.749	0.172**	0.176**
		(0.066)	(0.072)

Notes: Column 1 shows mean high school educational outcomes for applicants who lost their first lottery, with observations weighted to reflect the distribution of lottery strata in the treatment group. Columns 2 and 3 report the ITT effect of an initial lottery win on these outcomes, estimated using Equation 1 with strata and school fixed effects. Panel A: Obs = 288, N = 96. Panel A: Column 2 regressions control for self-reported baseline GPA and an indicator for missing self-reported baseline GPA. Column 3 regressions include these controls along with controls for all self-reported baseline survey information and census-tract level characteristics reported in Table I. Panel B: Column 2 regressions control for applicants' standardized, weighted eighth grade GPA from BPS records and an indicator for missing eighth grade GPA. Column 3 regressions include these controls along with controls for race, having an IEP, ESL status, an indicator for economic disadvantage, math and ELA standardized test scores from eighth grade, and an indicator for missing standardized test scores in eighth grade. Panel B: N = 86. Controls differ between panels due to limitations on demographic data available in de-identified BPS records. Standard errors are clustered at the individual level in Panel A and robust standard errors for Panel B are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

closest benchmark, Oreopoulos et al. (2017), who find a 15 percentage point effect at the focal site and a 6 percentage point effect after expansion. The effects estimated here therefore lie at the upper end of the existing evidence.

B Post-secondary: college entry, graduation, and quality

We next explore how improvements in high school achievement carry forward into post-secondary enrollment, degree completion, and institutional quality. Panel A of Table III examines a standardized summary index that, with equal weight, aggregates post-secondary outcomes except for receipt of an associate's degree, for which the directional interpretation is ambiguous (Mountjoy, 2022).[18] The ITT effect on the summary index indicates substantial improvements in degree attainment and college quality. Initial lottery winners score 0.45–0.49 standard deviations higher on the index ($p < 0.02$).

Although all component outcomes move in the same direction, Panel B of Table III shows that the gains are largest for our other main outcome – bachelor's degree attainment. Initial lottery winners are 17–21 percentage points (31–37%) more likely to earn a bachelor's degree. As illustrated in Figure III, these gains appear to reflect improved persistence rather than increased entry into bachelor's programs. Nearly the entire increase in bachelor's degree attainment can be accounted for by fewer students failing to complete their bachelor's degree, a margin that fell by 15–21 percentage points (52–73%).

Associate's degree completion is unchanged, but treated students are also 15–16 percentage points less likely to enroll in an associate's program without completing either an associate's or bachelor's degree. Panel C shows that initial lottery winners are more likely to attend more selective post-secondary institutions. While point estimates are imprecise, their magnitudes are economically large. The treatment group is 9 percentage points (34–36%) more likely to attend a top-100 university or top-50 liberal arts college and 5–6 percentage points (132–162%) more likely to attend a top-30 university or top-5 liberal arts college. For academically prepared students from low-income households, attendance at more selective institutions may improve access to financial aid and lower the net cost of earning a bachelor's degree (Hoxby and Avery, 2013). Although we do not observe financial aid directly, the next section explores student loan balances and borrowing in credit bureau records when students are in their mid to late twenties.

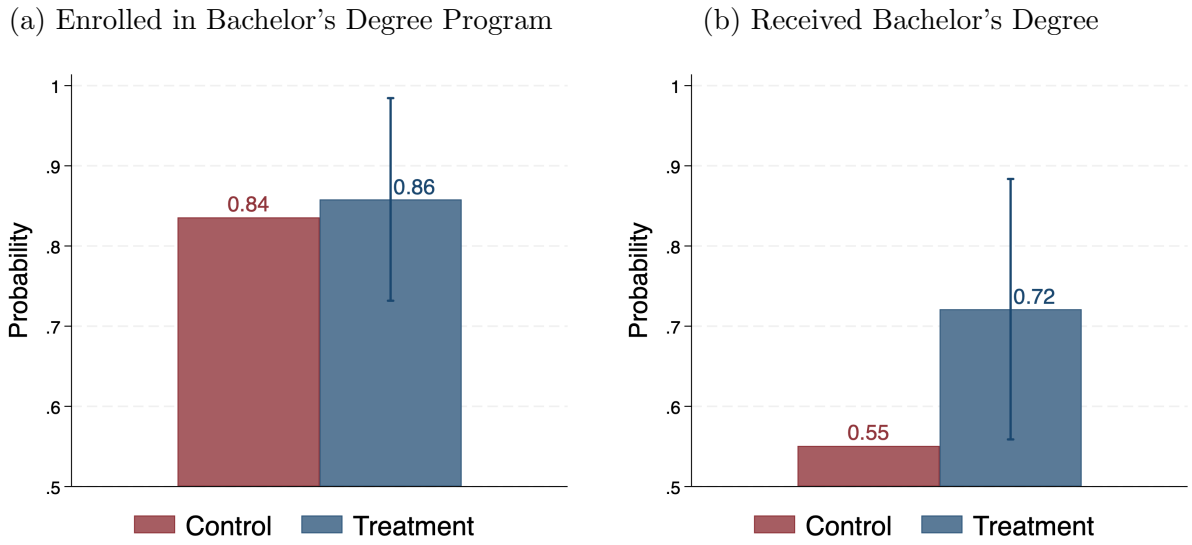
[18] We also omit broad enrollment measures from the index. Instead, we split enrollment into two mutually exclusive categories, such as bachelor's degree receipt and bachelor's enrollment without completion (which enters negatively). See Section III for more details.

Table III: Effects on Post-Secondary Schooling: College Entry, Graduation, Quality

	Control Mean (weighted)	Intention to Treat	Intention to Treat
	(1)	(2)	(3)
Specification		*Basic*	*All Controls*
Panel A. Standardized summary index			
Standardized index	0.000	0.451***	0.494***
		(0.166)	(0.178)
Panel B. Post-secondary enrollment and degree completion			
Enrolled in Bachelor's degree program	0.836	0.022	-0.003
		(0.063)	(0.073)
Received Bachelor's degree	0.551	0.170**	0.206**
		(0.082)	(0.091)
Enrolled in but did not graduate Bachelor's program	0.285	-0.148*	-0.209**
		(0.079)	(0.086)
Enrolled in Associate's program	0.301	-0.137*	-0.156*
		(0.074)	(0.083)
Received Associate's degree	0.080	-0.010	-0.011
		(0.042)	(0.042)
Enrolled in Associate's program but did not graduate (with an Associate's or Bachelor's degree)	0.201	-0.152***	-0.157**
		(0.056)	(0.063)
Panel C. Post-secondary institution quality			
Top 100 university or top 50 liberal arts	0.256	0.093	0.088
		(0.079)	(0.093)
Top 30 university or top 5 liberal arts	0.034	0.055	0.045
		(0.042)	(0.044)

Notes: Column 1 shows mean post-secondary educational outcomes for applicants who lost their first lottery, with observations weighted to reflect the distribution of lottery strata in the treatment group. Columns 2 and 3 report the effect of an initial lottery win on these outcomes, estimated using Equation 1 with strata and school fixed effects. Regressions in Column 2 control for self-reported baseline GPA and an indicator for missing self-reported baseline GPA. Regressions in Column 3 include these controls along with controls for race, grade level, the index of self-reported elements of academic comfort, an indicator for missing any components of this index, and the census-tract level characteristics reported in Table I. Robust standard errors reported in parentheses. N = 130. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure III: Effects on Post-secondary Outcomes



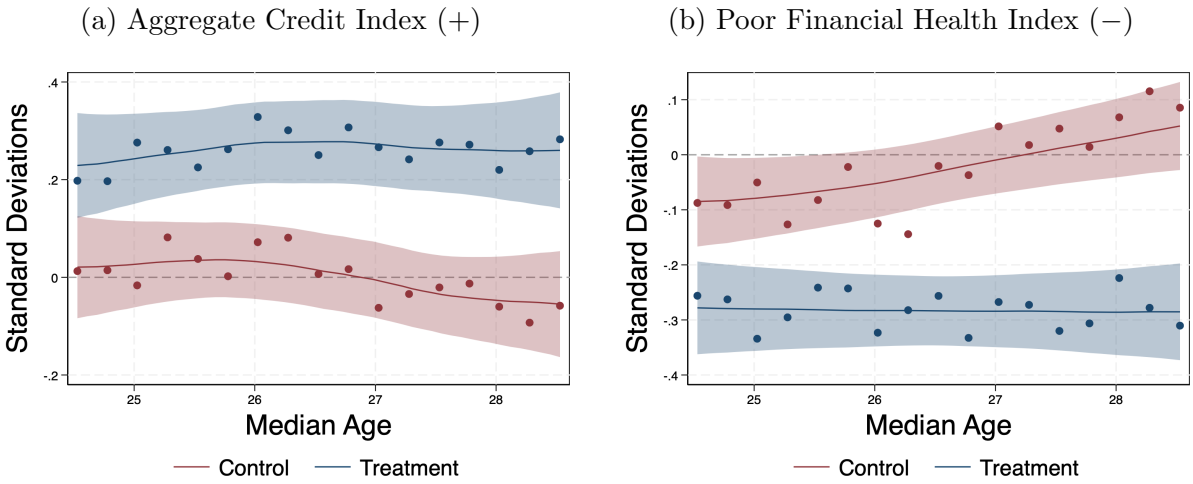
Notes: This figure plots ITT effects on post-secondary outcomes for those who won their first lottery in navy (Treatment), and lost their first lottery in maroon (Control). Panel A displays enrollment in bachelor's degree program and Panel B reports receipt of bachelor's degree. Control-group means are weighted to match the distribution of lottery strata among treatment-group applicants. Treatment-group outcomes are constructed by adding the estimated intention-to-treat effect of an initial lottery win to the weighted control-group mean. ITT effects are estimated using Equation 1, which includes lottery-strata and school fixed effects and controls for self-reported baseline GPA and an indicator for missing baseline GPA. Whiskers reflect 95% confidence intervals based on robust standard errors.

C Leading Labor Market Indicators in Adulthood

We first plot trends in our two standardized summary indices separately for treatment and control groups across the seventeen quarters of available credit-bureau records. The first combines all standardized credit outcomes, while the second captures adverse financial outcomes, such as having delinquent bills, a collections balance, poor credit, or student-loan debt. Over the full panel, students are observed between ages 22 and 31, with the median age ranging from 24.5 to 28.5. Figure IV shows that treatment-control differences emerge early and widen throughout applicants' twenties (Panel A), driven in part by a rising incidence of adverse financial outcomes among control-group students (Panel B).

Our main analyses, reported in Table IV, focus on the final eight quarters of available data, when applicants are in their late twenties and financial outcomes have begun to stabilize. These quarters additionally contain predicted earnings measures. We estimate gains of 0.37–0.38 standard deviations in the aggregate credit index ($p < 0.031$) and declines of 0.40–0.43 standard deviations in the poor financial health index ($p < 0.030$).

Figure IV: Raw Trends of Summary Indices by Study Group



Notes: This figure plots mean summary indices separately for initial lottery winners (treatment) and initial lottery losers (control) for each of the 17 quarters of available credit-bureau records. We overlay local-constant kernel regressions using an Epanechnikov kernel, corresponding to Stata's `lpolyci` command with `degree(0), kernel(epanechnikov)`. Shaded areas report pointwise 95% confidence intervals for the local mean smooth (Fan and Gijbels, 1996).

All individual outcomes move in the same direction, although multiple are not statistically significant on their own. The clearest improvements are an 11–13 percentage point (32–38%) reduction in the likelihood of having a poor credit score, a 14–15 percentage point (24–26%) reduction in the likelihood of having any bill 30 days delinquent, and a 13–14 percentage point (19–20%) reduction in the likelihood of having any student loan debt. Although Experian's predicted income variable is noisy, it provides an intuitive and easily interpretable metric to assess the program's effect on probable incomes in young adulthood. Point estimates imply gains in predicted annual earnings of \$1,900–\$4,740.

Table A.4 reproduces the credit-bureau results using all available quarters of data. We omit predicted earnings because these measures were only collected in the final eight quarters. The resulting point estimates closely mirror those in the main table, while the larger sample yields somewhat tighter confidence intervals. For example, coefficients on individual credit-score outcomes are similar in magnitude but are more precisely estimated, with p-values ranging from 0.01 to 0.09.

Table IV: Effects on Leading Labor Market Credit Indicators

	Control Mean	ITT	ITT
	(1)	(2)	(3)
Specification		*Basic*	*All Controls*
Panel A. Standardized Summary Indices			
Aggregate Credit Index	0.000	0.370**	0.377**
		(0.159)	(0.173)
Poor Financial Health Index	0.000	−0.430**	−0.396**
		(0.165)	(0.180)
Panel B. Credit Access			
Has an Experian record	0.901	0.070^{+}	0.036
		(0.046)	(0.043)
Panel C. Credit Score			
Vantage Credit Score	652.408	23.131^{+}	23.764^{+}
		(14.804)	(15.422)
log(Vantage Credit Score)	6.469	0.037^{+}	0.039^{+}
		(0.023)	(0.024)
Prime Credit (i.e., Score > 660)	0.517	0.098	0.108^{+}
		(0.073)	(0.076)
Poor Credit (i.e., Score < 600)	0.354	−0.115*	−0.134*
		(0.068)	(0.073)
Panel D. Predicted Annual Earnings			
Predicted earnings (in $1,000s)	$43.298	$1.904	$4.738
		(3.961)	(3.858)
log(Predicted earnings)	3.611	0.059	0.092
		(0.067)	(0.069)
Predicted earnings < $30,000	0.346	−0.095^{+}	−0.090
		(0.060)	(0.066)
Panel E. Assets			
Mortgage	0.100	0.014	0.036
		(0.044)	(0.041)
Auto loan	0.287	0.015	0.057
		(0.075)	(0.073)
Panel F. Financial Distress			
Collections Balance > 0	0.159	−0.048	−0.056
		(0.062)	(0.065)
Any bills delinquent for 30 days	0.567	−0.137*	−0.151*
		(0.082)	(0.083)
Any bills delinquent for 90 days	0.427	−0.107	−0.105
		(0.079)	(0.082)
Panel G. Student Loans			
Any student loan balance	0.705	−0.142*	−0.133*
		(0.079)	(0.080)
Student loan balance over past 3 months (in $1,000s)	$24.445	−$5.809	−$3.209
		(5.627)	(5.086)

Notes: Col 1: mean credit bureau outcomes for the control group re-weighted to match distribution of lottery strata in the treatment group. Columns 2 and 3 report the effect of an initial lottery win, estimated using Equation 1, including strata, school, and age-year fixed effects. Regressions in Col 2 control for self-reported baseline GPA and missing indicator. Regressions in Col 3 include the full set of baseline controls. Aggregate Credit Index includes positive standardized measures of *log*(predicted annual earnings), *log*(Vantage credit score), and having: an Experian record, prime credit, a mortgage, an auto loan, and negative standardized measures of student loans balance over past 3 months and having predicted annual earnings < $30,000, poor credit, any collections balance, delinquency in past 30 days, delinquency in past 90 days, any student-loan balance in the last 3 months. Poor Financial Health Index includes standardized measures of having no Experian record, income below $30,000, poor credit, any collections balance, delinquency in past 30 days, delinquency in past 90 days, and any student loans balance. SEs clustered at individual level reported in parentheses. Obs = 1,040, N = 130. $^{+}$ $p < 0.15$, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

D Robustness

Across all outcome categories, we report estimates from specifications that include either minimal controls selected using LASSO or the full set of available baseline controls. Given the balance between initial lottery winners and losers, the two specifications produce very similar estimates. Adding controls increases some estimates and attenuates others, leaving the overall pattern of results largely unchanged.

We further assess the robustness of our findings using randomization inference. Conventional confidence intervals rely on asymptotic approximations that may be less reliable in small samples, particularly in experimental settings with many outcomes or limited effective sample sizes (Young, 2019). To address this concern, we simulate the null distribution of treatment-effect t-statistics by re-running each lottery 1,000 times, preserving the original randomization strata and assignment probabilities, and re-estimating Equation 1 using placebo treatment assignments. Table A.2 reports conventional asymptotic p-values in Column 3 and randomization-inference p-values in Column 4. The two approaches yield both quantitatively and qualitatively similar results: across outcomes, the randomization-inference p-values differ from the conventional p-values by no more than 0.013.

Finally, we use randomization inference to assess the likelihood of observing a similarly strong pattern of effects across all of our main outcomes. Following Young (2019), our preferred omnibus statistic is the sum of squared treatment-effect t-statistics across outcomes. We omit cumulative high school GPA and high school graduation from this joint test because BPS records are de-identified and cannot be linked to the other datasets at the individual level.[19] Because this statistic does not impose a common sign on the treatment effects, it is conservative relative to a directional test that asks whether the effects are jointly significant and move in the expected direction. We estimate a joint randomization-inference p-value of 0.003, suggesting that it is highly unlikely that the full pattern of estimated effects would arise by chance under the null of no treatment effects.

VI Mechanisms and Welfare Analysis

A Additional mechanism results

Table V explores mechanisms underlying program effects on students' educational trajectories. Because Amphibious Achievement combined mentorship, athletics-based instruction, tutoring, and other program attributes (e.g., college campus exposure), we cannot separately

[19] If anything, Column 4 of Table A.2 suggests that including these outcomes would strengthen the evidence of joint significance, as their randomization-inference p-values are relatively small.

identify the contribution of each core component. Instead, we examine the pattern of impacts across outcomes that differ in their reliance on academic skills, persistence, and alignment with mentor backgrounds.

The evidence points to two complementary channels: improved academic achievement and strengthened non-cognitive skills, including goal-setting, confidence, and persistence. Panel A explores whether GPA effects differ across subject areas. Using course titles from transcript records, we classify courses into STEM and humanities/social science categories and construct standardized GPAs for each group. The program improves performance in both areas, but gains are larger in STEM: standardized STEM GPA rises by 0.52 standard deviations, compared with 0.38 standard deviations in humanities and social science. Although the difference is not statistically significant, the pattern is consistent with mentoring based at a technical university being especially well positioned to not only provide academic support, but also to model and motivate success in STEM-specific fields.

Panel B compares effects on course grades and standardized test scores. This analysis is limited to students who applied in ninth grade or the fall of tenth grade and are observed in BPS records, so estimates are suggestive. The comparison is useful because grades often capture elements of non-cognitive skills – for example, factoring in participation in class and perseverance over time – while standardized tests skew towards capturing cognitive skills.[20] We find large gains on both margins. Cumulative GPA rises by 0.66–0.81 standard deviations, and combined MCAS scores rise by 0.86 standard deviations. Point estimates are again larger in math than in English Language Arts, though the difference is not statistically significant. Effects on standardized tests suggest that the program improved measured academic skills, while large GPA effects are also consistent with improvements in both academic behaviors and non-cognitive skills.

Panel C examines whether stronger high school performance, particularly in STEM, translated into different post-secondary degree choices and success. We distinguish between receipt of Bachelor of Science (B.S.) and Bachelor of Arts degrees using National Student Clearinghouse data. This distinction is substantively important because STEM degrees are often associated with higher labor-market returns and may require both stronger preparation and greater confidence that such pathways are feasible. Lottery winners are roughly 13 percentage points more likely to earn a B.S. degree, compared with a 5–8 percentage point increase in B.A. attainment. Roughly two-thirds of the total effect on bachelor's degree completion is therefore driven by increased B.S. attainment. This pattern suggests that

[20]Standardized tests are not pure measures of cognitive skill: Brown et al. (2025) show that test performance also reflects non-cognitive factors such as cognitive fatigue and test-taking endurance. Grades, however, likely place greater weight on sustained academic behaviors, including homework completion, quiz performance, classroom engagement, and effort.

Table V: Mechanisms

	Control Mean (weighted)	Intention to Treat	Intention to Treat
	(1)	(2)	(3)
Specification		*Basic*	*All Controls*
Panel A. High School Records: GPA in STEM vs. Humanities and Social Science Courses			
Standardized STEM GPA	0.000	0.524***	0.501***
		(0.172)	(0.165)
Standardized HSS GPA	0.000	0.375*	0.279*
		(0.190)	(0.160)
Panel B. Standardized Test vs. Course Grades			
Standardized Cumulative GPA	0.000	0.657**	0.809**
		(0.285)	(0.372)
MCAS Combined	0.000	0.865**	0.863**
		(0.334)	(0.374)
MCAS Math	0.000	0.777*	0.980***
		(0.396)	(0.312)
MCAS ELA	0.000	0.589**	0.444
		(0.288)	(0.385)
Panel C. Bachelor's of Science vs. Bachelor's of Arts Degrees			
Bachelor of Science Degree	0.227	0.125^{+}	0.128^{+}
		(0.077)	(0.084)
Bachelor of Arts Degree	0.324	0.045	0.078
		(0.087)	(0.093)
Panel D. Post-secondary Records: Enrollment vs. Graduation			
Enrolled in Bachelor's degree program	0.836	0.022	−0.003
		(0.064)	(0.073)
Received Bachelor's degree	0.551	0.170**	0.206**
		(0.082)	(0.091)
Enrolled in but did not graduate Bachelor's program	0.285	−0.148*	−0.209**
		(0.079)	(0.086)
Enrolled in Associate's program	0.301	−0.137*	−0.156*
		(0.074)	(0.083)
Enrolled in Associate's program *and* Bachelor's program	0.227	−0.110^{+}	−0.127^{+}
		(0.067)	(0.077)
Enrolled in Associate's program *only*	0.075	−0.027	−0.029
		(0.038)	(0.042)
Received Associate's degree	0.080	−0.010	−0.011
		(0.042)	(0.042)
Enrolled in Associate's program but did not receive Associate's or Bachelor's degrees	0.201	−0.152***	−0.157**
		(0.056)	(0.063)

Notes: Column 1 shows mean outcomes for applicants who lost their first lottery, with observations weighted to reflect the distribution of lottery strata in the treatment group. Columns 2 and 3 report the effect of an initial lottery win on these outcomes, estimated using Equation 1 with strata and school fixed effects. See Table III for controls used in Panels A, C, and D. See Panel B of Table II for controls used in Panel B. Panel B restricts to the 50 students who applied in 9th grade or the fall of their 10th grade school year and outcomes are re-standardized to ease comparisons. Robust standard errors reported in parentheses. $^{+}$ $p < 0.15$, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

the program may have shifted not only academic performance, but also students' perceived opportunity sets and persistence in more quantitatively demanding fields.

Panel D explores the extent to which improvements in bachelor's degree receipt were driven by enrollment decisions, persistence, or both. The evidence points primarily to persistence rather than entry, although students' post-secondary enrollment paths also appear to have shifted. Initial lottery winners and losers are equally likely to enroll in bachelor's programs, but winners are 17–21 percentage points more likely to complete a bachelor's degree and 15–21 percentage points less likely to enroll without completing. Lottery winners were also 11–13 percentage points less likely to have enrolled in both associate's and bachelor's programs. This pattern is consistent with the program shifting students away from community-college-to-bachelor's pathways and toward direct enrollment in bachelor's programs after high school. Importantly, the program also reduced left-tail post-secondary outcomes: students were 15–16 percentage points less likely to enroll in an associate's program without ultimately earning either an associate's or bachelor's degree. Along with the reduction in bachelor's non-completion, this result suggests that the program reduced the share of students who incurred the costs of post-secondary enrollment without ultimately earning a credential.

Recent evidence on Big Brothers Big Sisters provides a useful comparison: Bell (2026) finds increased college attendance, but imprecise and negative earnings effects. One possible interpretation is that mentorship can raise college aspirations and entry, but that persistence may require more structured support for the skills needed to navigate college. Amphibious Achievement provided this structure through repeated goal-setting, personal records, and mentor feedback in both academic and athletic environments. Consistent with this distinction, our post-secondary effects operate almost entirely through persistence and completion rather than entry.

In summary, gains in MCAS scores and GPA indicate improvements in both measured achievement and broader school performance that reward non-cognitive skills. Moreover, stronger STEM trajectories in high school and increased B.S. attainment suggest that mentors may have shaped students' academic interests, skills, and sense of feasible education pathways. The post-secondary results also point to greater persistence after college entry, along with shifts in degree types and pathways. Overall, these patterns are consistent with complementarities between the cognitive and non-cognitive skills emphasized in the program, contributing to holistic human capital formation.

B Heterogeneity

B.1 Effects for the highest needs students

Table A.3 examines whether treatment effects differ across students with different baseline characteristics. Column 1 reports the main ITT estimates for the full sample across our primary outcomes. Columns 2 and 3 split the sample by students' predicted long-term outcome index, reporting effects separately for students below and above the median predicted value. Using only control-student data, this approach summarizes baseline characteristics in a single index that predicts long-run post-secondary and credit-market outcomes.[21] The resulting index is most strongly associated with baseline academic achievement and neighborhood opportunity.[22]

We find that program effects are concentrated among students with weaker predicted long-run trajectories. For students below the median of the predicted long-term outcome index, treatment increases high school GPA by 0.71 standard deviations ($p < 0.001$), the post-secondary schooling index by 0.82 standard deviations ($p = 0.001$), bachelor's degree receipt by 35 percentage points ($p = 0.005$), and the aggregate credit index by 0.63 standard deviations ($p = 0.002$). It also reduces the poor financial health index by 0.78 standard deviations ($p < 0.001$). For students above the median, estimated effects on human-capital outcomes are positive and effects on credit outcomes are slightly negative, although neither set of estimates is statistically significant. Figure V shows a similar pattern across the distribution of predicted long-run outcomes: program effects are largest for students with weaker predicted trajectories, who tend to have lower baseline GPAs and come from higher-poverty neighborhoods.

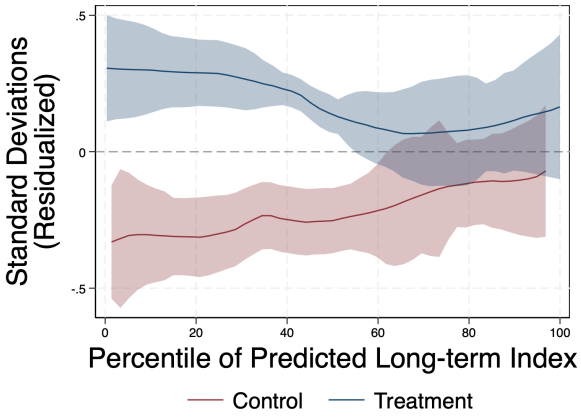
This pattern is consistent with the program being especially valuable for students with limited access to educational investments, college-going information, and academically successful peers. It also aligns with studies showing that university-student mentors improve educational choices, skills, and behaviors among disadvantaged adolescents (Falk et al., 2020; Resnjanskij et al., 2024), as well as work linking cross-SES economic connectedness to upward mobility for low-income children (Chetty et al., 2022, 2026). These results are consistent with theoretical predictions that suggest the marginal applicant least likely to apply to a program

[21] We construct the predicted long-term outcome index using the full set of baseline controls from students' initial applications. Among control students, we regress the average of the post-secondary schooling index and the most recent quarter of the aggregate credit index on these baseline covariates. For each control student, we generate a leave-one-out prediction to avoid mechanical overfitting. For treated students, predictions are generated using coefficients estimated from the full control group.

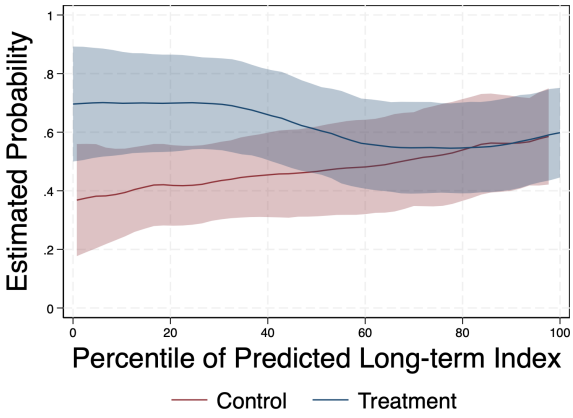
[22] Regressing the predicted long-term index on the baseline covariates shows that it is most strongly correlated with baseline GPA ($t = 25.4$) and Chetty et al.'s tract-level economic mobility measure ($t = 10.7$), followed by other tract-level characteristics, including the share with a bachelor's degree and the shares who are Black or Hispanic.

Figure V: ITT Effects by Predicted Long-run Trajectory at Baseline

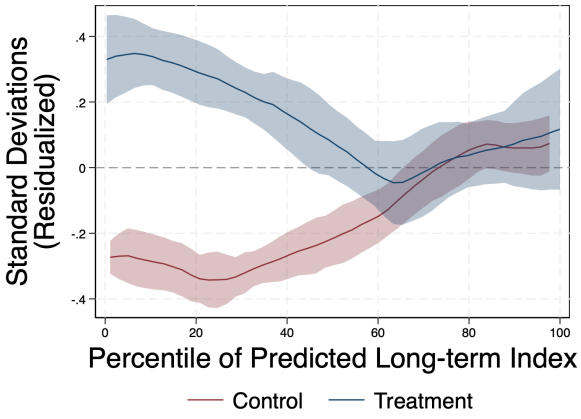
(a) High School GPA



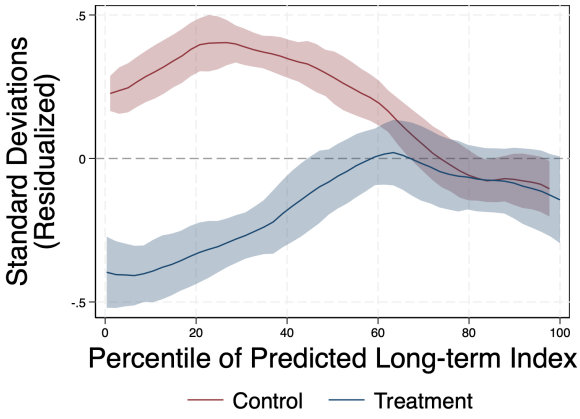
(b) Bachelor's Degree



(c) Aggregate Credit Index



(d) Poor Financial Health Index



Notes: This figure separately plots residualized outcomes, estimated using strata and school fixed effects and controlling for self-reported baseline GPA and an indicator for missing baseline GPA (as given in Equation 1), for lottery winners (treatment) and losers (control) by a student's predicted long-run outcome index at baseline. This index is constructed by regressing baseline covariates from a student's initial application on the average of a student's post-secondary schooling index and the most recent quarter of the aggregate credit index among control students. Panels A, B, C, and D respectively display the following residualized outcomes: semesterly high school GPA, estimated bachelor's degree receipt, aggregate credit index, and poor financial health index. For bachelor's degree receipt, we add residuals from the indicator outcome to the weighted control-group mean to preserve the scale of the binary outcome. For a given control-student prediction, we "leave out" that observation to prevent mechanical overfitting. Treatment-group predictions use coefficients estimated from the full control group. We plot local-constant kernel regressions using an Epanechnikov kernel and a bandwidth of 15 percentile points in the predicted long-run outcome ranking, corresponding to Stata's `lpolyci` command with `degree(0)`, `kernel(epanechnikov)`, and `bwidth(15)`. Shaded areas report pointwise 95% confidence intervals for the local mean smooth (Fan and Gijbels, 1996).

is also the most likely to benefit (Walters, 2018). Thus, leveraging athletics to motivate students who might not have otherwise applied to an enrichment program can help draw the highest return students. Because these subgroup comparisons are based on predicted baseline trajectories, however, we interpret them only as suggestive evidence that the program had the largest effects for students with more room to benefit.

B.2 Complementarities over time

In Table A.3 Columns 4 and 5, we split the sample by grade at application, comparing students who applied in ninth grade with those who applied in grades 10–12. If program effects operate through complementarities, then effects should be larger for students exposed earlier in high school. Results are consistent with this prediction. Among ninth-grade applicants, treatment increases high school GPA by 0.91 standard deviations ($p = 0.047$), the post-secondary schooling index by 0.66 standard deviations ($p = 0.032$), bachelor's degree receipt by 30 percentage points ($p = 0.054$), and the aggregate credit index by 0.72 standard deviations ($p = 0.005$), while reducing the poor financial health index by 0.85 standard deviations ($p = 0.002$). These point estimates are roughly 50–100 percent larger than the corresponding full-sample estimates. This pattern is consistent with earlier exposure giving students more time for holistic human capital to accumulate.

B.3 Equity

Finally, Table A.3 Columns 6 and 7 explore heterogeneity by student sex. Point estimates are consistently larger for female students than for male students. Among female students, treatment increases high school GPA by 0.57 standard deviations ($p = 0.009$), the post-secondary schooling index by 0.63 standard deviations ($p = 0.002$), bachelor's degree receipt by 26 percentage points ($p = 0.016$), and the aggregate credit index by 0.45 standard deviations ($p = 0.024$), while reducing the poor financial health index by 0.38 standard deviations ($p = 0.068$). Estimated effects for male students generally point toward improvements, particularly for high school GPA and the poor financial health index, but are smaller and not statistically significant. We interpret these gender differences cautiously because the male-female differences are not statistically significant, although the pattern is consistent with evidence of larger effects for female students in related college-going interventions (Angrist et al., 2009; Carrell and Sacerdote, 2017; Oreopoulos et al., 2017).

C Compliers and the Impact of Program Participation

We next estimate the effect of program participation for students whose take-up was induced by winning the initial lottery. As shown in Table A.1, 98.4 percent of initial lottery

winners attended at least one session, compared with 31.4 percent of initial lottery losers, many of whom later participated after reapplying in subsequent semesters. Thus, roughly two-thirds of applicants are *compliers* – students induced to participate by winning their initial lottery – while roughly one-third are *always-takers*. Never-takers are rare, accounting for less than 2 percent of lotteried applicants.

To characterize the complier population, we follow Heller et al. (2017) in estimating control complier means, weighting lottery strata by the treatment-group distribution.[23] Columns 1 and 2 of Table VI report weighted control means and control complier means, respectively. The complier means indicate that the marginal students induced to participate by winning the initial lottery were more disadvantaged than the average control applicant. Their short-term GPA was 0.31 standard deviations lower, and their cumulative high school GPA was 0.79 standard deviations lower. Their bachelor's degree completion rate was also much lower: only 30.0 percent of control compliers completed a bachelor's degree by adulthood, compared with 55.1 percent of the weighted control group and 40.2 percent of U.S. 25-to-29-year-olds in 2024 (U.S. Department of Education, National Center for Education Statistics, 2024). Compliers were also especially likely to reach but not complete the post-secondary margin, with 60.0 percent enrolling in a bachelor's program without completing a degree, compared with 28.5 percent of the weighted control group. These disadvantages extended into adulthood, where control compliers were more likely to experience poor financial outcomes.

Column 3 reports LATE estimates for this complier population. Because the regression-adjusted first stage is 68.3 percent, LATE estimates are approximately 46 percent larger than the corresponding ITT estimates. The implied effects are large relative to complier counterfactual outcomes. Program participation increases bachelor's degree completion by 25.0 percentage points, an 83 percent increase over the control complier mean. It also raises high school graduation by 26.4 percentage points, increases credits earned by 2.39 courses, and improves cumulative GPA by 0.74 standard deviations. Across the summary indices, program participation moves compliers from substantially below the control-group mean to roughly average or above-average outcomes in high school, post-secondary schooling, and

[23]Let Z_i indicate winning the initial lottery and D_i indicate ever attending the program. Within each lottery stratum s, we estimate the shares of always-takers, compliers, and never-takers as $\pi_s^A = \Pr(D_i = 1 \mid Z_i = 0, s)$, $\pi_s^C = \Pr(D_i = 1 \mid Z_i = 1, s) - \Pr(D_i = 1 \mid Z_i = 0, s)$, and $\pi_s^N = \Pr(D_i = 0 \mid Z_i = 1, s)$. Under monotonicity and the exclusion restriction, the control complier mean is recovered as

$$\bar{Y}_{C,s}(0) = \frac{\bar{Y}_{0s} - \pi_s^A \bar{Y}_{A,s}(0) - \pi_s^N \bar{Y}_{N,s}(0)}{\pi_s^C},$$

where $\bar{Y}_{0s}$ is the mean outcome among lottery losers in stratum s, $\bar{Y}_{A,s}(0)$ is the mean outcome among lottery-losing participants, and $\bar{Y}_{N,s}(0)$ is the mean outcome among lottery-winning nonparticipants. We then aggregate stratum-specific means using the distribution of lottery strata among treated applicants.

Table VI: Effect of Program Participation

	Control Mean (weighted)	Control Complier Mean	Local Average Treatment Effect
	(1)	(2)	(3)
Panel A. Short-term High School GPA			
Overall GPA	2.430	2.061	0.560***
			(0.209)
Standardized Overall GPA	0.000	-0.311	0.680***
			(0.244)
Panel B. Long-term High School Outcomes			
Cumulative GPA	2.891	2.307	0.545**
			(0.271)
Standardized Cumulative GPA	0.000	-0.788	0.736**
			(0.366)
Total Credits Earned	12.725	10.133	2.385**
			(1.099)
Graduated High School	0.749	0.733	0.264**
			(0.109)
Panel C. Post-Secondary Schooling Outcomes			
Post-secondary Index	0.000	-0.430	0.670***
			(0.244)
Received Bachelor's Degree	0.551	0.300	0.250**
			(0.118)
Enrolled in Bachelor's program but did not graduate	0.285	0.600	-0.217*
			(0.117)
Panel D. Standardized Credit Market Indices			
Aggregate Credit Index	0.000	-.095	0.544**
			(0.233)
Poor Financial Health Index	0.000	0.268	-0.631***
			(0.241)

Notes: Column 1 shows mean outcomes for applicants who lost their first lottery, weighted to reflect the treatment group's strata composition. Column 2 reports control complier means, also weighted to reflect the treatment group's strata composition. Column 3 reports 2SLS estimates of the average effect of program participation for compliers, estimated using equation 3, where initial lottery outcome instruments for ever attending a program session. Panels A and D cluster standard errors at the individual-level and Panels B and C report robust standard errors, which are given in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

adult credit-market outcomes.

LATE estimates are particularly relevant when comparing programs across settings, as the variation in effects in education is often driven by implementation fidelity and take-up (Angrist and Meager, 2023). Moreover, the negative selection we observe on compliers reinforces one of the potential mechanisms at play: athletics select for lower performing students who might not otherwise enroll in academic enrichment programs, and these are the students who are likely to benefit most. This result is consistent with theoretical predictions suggesting that program expansions are likely to be most effective for those least likely to apply in the first place, but when they apply, they benefit the most (Walters, 2018).

D Cost Effectiveness and Welfare

We first assess welfare by calculating the program's marginal value of public funds (MVPF). Following Hendren and Sprung-Keyser (2020), the MVPF is participants' willingness to pay (WTP) divided by net government cost, where net government cost equals direct program costs net of induced fiscal externalities. We decompose the calculation into three components: participants' WTP, direct program costs, and the fiscal externality generated by higher tax revenue. We use two approaches to estimate the present value of participants' earnings gains. In both approaches, after-tax earnings gains provide an estimate of participants' WTP, while tax revenue generated by those additional earnings provides an estimate of the program's fiscal externality.

The first approach draws on our direct estimate of effects on predicted annual wage income in credit data. Using the lifecycle projection approach in Hendren and Sprung-Keyser (2020), we input the average LATE estimate across our two specifications, a \$5,080 increase in predicted wage income at age 28, into a lifecycle model that assumes 1.2% annual real wage growth and employment through age 64.[24] With a 3% real discount rate, this yields a present value of gross future earnings gains of approximately \$139,000 per participant. Applying a net tax-and-transfer rate of 18.7%, this approach implies WTP of approximately \$113,000 per participant and a positive fiscal externality of approximately \$26,000 per participant.[25]

Our second approach estimates the present value of increases in educational attainment by adding the expected labor-market returns from a 26.3 percentage point increase in high school graduation and a 28.5 percentage point increase in bachelor's degree receipt, averaging coefficients across both LATE specifications.[26] Using lifetime earnings estimates from

[24] We calculate \$5,080 as the average of two LATE estimates, \$2,817 and \$7,343. The larger LATE estimate reflects that age-year fixed effects and full controls slightly reduce the regression-adjusted first stage.

[25] Predicted wage income in credit data suggests that the average complier falls within 300–349% of the federal poverty line, which implies a tax rate net of transfers of 18.7% (Hendren and Sprung-Keyser, 2020).

[26] The LATE estimates are 26.356 and 26.296 percentage points for high school graduation and 24.954

Belfield et al. (2011) and Tamborini et al. (2015) that compare high school graduates to high school dropouts, we value high school graduation at approximately $454,000 in 2025 dollars relative to dropping out. This implies a benefit of approximately $119,000 per participant.[27] For the bachelor's-degree component, we use Mountjoy (2026), who estimates causal returns for marginal four-year college applicants near an admissions cutoff, a population similar to compliers in our setting. Mountjoy finds that target-university enrollment increases annual earnings by $3,339 and four-year bachelor's degree receipt by 11.9 percentage points. Scaling this earnings effect by the ratio of our bachelor's-degree effect to Mountjoy's first stage implies an annual earnings gain of $7,997 in applicants' late twenties, or a present-value earnings gain of approximately $219,000.[28] Combining the high school and bachelor's-degree components yields a total present-value earnings gain of approximately $338,000 per participant, with after-tax private benefits of $275,000 and public tax revenue gains of $63,000.

The program operated at very low cost. In its first year, annual operating costs were roughly $10,000, or about $200 per enrolled student. As the program secured donations from local restaurants and university discounts, subsequent annual budgets fell to roughly $2,000. Its main expenses were public transportation passes, athletic facility costs, and lunch. Averaging across the two budgets yields an annual cost of roughly $6,000, or $121 per enrolled student in 2012 dollars, equivalent to $170 in 2025 dollars.[29] These low costs reflect the program's operating model: college volunteers supplied mentoring and instruction, offsetting the largest cost driver of comparable programs. We do not include volunteer undergraduate labor in program costs. Because mentors voluntarily supply their time, their willingness to pay to participate must be weakly positive. We therefore conservatively assign zero value to mentor surplus in participants' WTP, rather than treating volunteer time as an additional program cost.[30]

Under these assumptions, the MVPF is infinite: the program pays for itself fiscally before accounting for private gains to participants. Estimated returns are also large relative to

and 32.090 percentage points for bachelor's degree receipt.

[27]Belfield et al. (2011) estimate a lifetime present-value earnings gain of $303,430 for high school graduates relative to dropouts, which we convert to approximately $434,000 in 2025 dollars using CPI-U. The corresponding estimate from Tamborini et al. (2015) is approximately $315,000 in 2009 dollars, which we convert to approximately $474,000 in 2025 dollars using CPI-W, following their adjustment of earnings to 2009 dollars using CPI-W. We use the midpoint of these estimates. These estimates are observational and may partially reflect selection.

[28]The scaling factor is $28.5/11.9 = 2.395$, so the implied annual earnings effect is $\$3{,}339 \times 2.395 = \$7{,}997$. We assume this gain persists from ages 28 to 64, with 1.2% real wage growth and a 3% real discount rate.

[29]We convert 2012 dollars to 2025 dollars using CPI-U.

[30]If instead we price mentor labor at $20 per hour, a typical wage for undergraduate work-study compensation, total annual program costs would increase to approximately $200,000 per year, or $4,000 per participant. Even under this alternative costing assumption, fiscal externalities still far exceed program costs, so the program continues to pay for itself.

costs: across the two approaches, per-participant private benefits range from approximately $113,000 to $275,000, while per-participant fiscal externalities range from approximately $26,000 to $63,000, far exceeding program costs of only $170 per participant-year.[31] These estimates also imply large break-even margins: program costs could rise to roughly $26,000 per participant under our more conservative estimate before induced tax revenue would no longer cover costs.

We visualize the distribution of MVPF values for over 200 high-profile public policies in the United States reviewed by Hendren and Sprung-Keyser (2020) in Figure A.1. Amphibious Achievement ranks at the 80th percentile of the MVPF distribution because roughly 20 percent of reviewed policies also have infinite values. However, this comparison undervalues Amphibious Achievement's relative cost-effectiveness since once policies "pay for themselves" the MVPF calculation has a ceiling effect and can no longer differentiate policies at the top end of the distribution. To compare relative cost-effectiveness among policies that pay for themselves, standard benefit-cost ratios provide a useful benchmark.

We compute a standard benefit-cost ratio (BCR) of 819 to 1,988.[32] Even the lower-end estimate ranks in the highest percentile of high-return education investments, exceeding most low-cost interventions in the global education literature (Angrist et al., 2023). For comparison, providing parents with information on the returns to education—classified as a "best buy" by the Global Education Evidence Advisory Panel (GEEAP)—has an estimated BCR of 732 (Akyeampong et al., 2023). Even under the very conservative assumption that earnings effects last only one year, the program has a BCR of roughly 30, comparable to leading global education interventions such as targeted instruction and structured pedagogy. For a prominent benchmark from the early-childhood literature, Perry Preschool has a BCR of 7–12 and an MVPF of approximately 44 (Hendren and Sprung-Keyser, 2020).

Comparisons with high-dosage tutoring underscore the cost-effectiveness of Amphibious Achievement. Guryan et al. (2023) report that Saga's Chicago tutoring model cost $3,200 to $4,800 per student per year during the study period, with later costs falling to about $1,800 after AmeriCorps subsidies and adopting a blended-learning model. Amphibious Achievement was more than an order of magnitude less expensive than this subsidized model, while delivering larger estimated long-run benefits.

[31] The average treatment-control difference in program attendance is 14 sessions. Because one program year consists of 20 offered sessions and participants attended roughly 70% of sessions, this attendance difference corresponds to one participant-year. The $170 cost is therefore measured per participant-year, aligning the cost unit with the additional program attendance induced by winning the lottery.

[32] This range divides estimated present-value lifetime earnings gains of $139,287–338,000 per participant by direct program costs of $170 per participant.

VII Conclusion

This paper evaluates a low-cost enrichment program for disadvantaged high school students that combines college-student mentorship, athletics-based instruction, and academic tutoring. Using random assignment from oversubscribed admissions lotteries, we find that the offer to participate had a large causal effect on students' educational outcomes. We estimate ITT effect sizes of about 0.48 SD on GPA – approximately five times higher than the median effective education intervention that generates an effect of about 0.1 SD (Kraft, 2020; Evans and Yuan, 2022). Most strikingly, we find persistent effects on long-run educational trajectories, such as high school graduation and bachelor's degree receipt, and leading labor market credit outcomes when students are in their late twenties. Effects operate both through cognitive and non-cognitive skill channels, in line with the program's focus on building holistic human capital.

The program also addresses two common barriers to effective and scalable educational interventions for adolescents: sustaining engagement among low-income students facing multifaceted disadvantage and keeping labor costs low. The combination of college-student mentorship and athletics helped attract and retain more disadvantaged students, who are often less likely to participate in such programs but whom we find benefited the most. At the same time, operating on a college campus and leveraging student volunteers kept costs extremely low, increasing the program's scalability.

The combination of low costs and persistent impacts implies unusually high private and public returns. Direct costs were only about $170 per participant-year, while benchmark present-value estimates of lifetime earnings gains imply expected private benefits of $113,000 to $275,000 per participant and fiscal externalities of $26,000 to $63,000 per participant through higher tax revenue. The program therefore more than pays for itself and ranks among the highest-return youth interventions by benefit-cost ratio.

Several limitations point to natural directions for future work. Because the intervention bundled several components, we cannot separately identify the contribution of mentorship, athletics, tutoring, campus exposure, or their interaction. Future work could test which components are most essential for sustaining long-run effects and whether similar gains can be produced using other sports or volunteer networks. In addition, our evidence comes from a unique sample in Boston with unusually rich long-run administrative records. Future work could test whether the model can be replicated in other settings and at larger scale.

References

Akyeampong, K., T. Andrabi, A. Banerjee, R. Banerji, S. Dynarski, R. Glennerster, S. Grantham-McGregor, K. Muralidharan, B. Piper, S. Ruto, et al. (2023). Cost-effective approaches to improve global learning-what does recent evidence tell us are "smart buys" for improving learning in low-and middle-income countries. *London, Washington DC, New York*.

Alan, S., T. Boneva, and S. Ertac (2019). Ever failed, try again, succeed better: Results from a randomized educational intervention on grit. *The Quarterly Journal of Economics 134*(3), 1121–1162.

Angrist, J., D. Lang, and P. Oreopoulos (2009). Incentives and services for college achievement: Evidence from a randomized trial. *American Economic Journal: Applied Economics 1*(1), 136–163.

Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996). Identification of causal effects using instrumental variables. *Journal of the American statistical Association 91*(434), 444–455.

Angrist, N., E. Aurino, H. A. Patrinos, G. Psacharopoulos, E. Vegas, R. Nordjo, and B. Wong (2023). Improving learning in low-and lower-middle-income countries. *Journal of Benefit-Cost Analysis 14*(S1), 55–80.

Angrist, N., P. Bergman, and M. Matsheng (2022). Experimental evidence on learning using low-tech when school is out. *Nature human behaviour 6*(7), 941–950.

Angrist, N. and R. Meager (2023). Implementation matters: Generalizing treatment effects in education. *Available at SSRN 4487496*.

Bakker, T. J., S. DeLuca, E. A. English, J. S. Fogel, N. Hendren, and D. Herbst (2025). Credit access in the united states. Technical report, National Bureau of Economic Research.

Belfield, C., F. Hollands, and H. Levin (2011, October). Providing comprehensive educational opportunity to low income students: What are the social and economic returns? Technical report, Campaign for Educational Equity, Teachers College, Columbia University, New York, NY.

Bell, A. (2026). Long-run effects of mentors: Evidence from a randomized evaluation linked to tax records. Technical report, Working paper.

Bernard, T., S. Dercon, K. Orkin, G. Schinaia, and A. S. Taffesse (2026). The future in mind: Aspirations and long-term outcomes in rural ethiopia. *The Quarterly Journal of Economics*, qjag002.

Boston Public Schools (2015). Cohort 2014 Four-Year Graduation Rate Report. `https://www.bostonpublicschools.org/fs/resource-manager/view/c92d57ec-098e-469e-8fb1-4cb837f21f1f`. Accessed May 22, 2026.

Brown, C., S. Kaur, G. Kingdon, and H. Schofield (2025). Cognitive endurance as human capital. *The Quarterly Journal of Economics 140*(2), 943–1002.

Carlana, M. and E. La Ferrara (2025). Apart but connected: Online tutoring, cognitive outcomes, and soft skills. *American Economic Review 115*(10), 3487–3513.

Carrell, S. and B. Sacerdote (2017). Why do college-going interventions work? *American Economic Journal: Applied Economics 9*(3), 124–151.

Chetty, R., R. Diamond, T. B. Foster, L. F. Katz, S. R. Porter, M. Staiger, and L. Tach (2026, January). Creating high-opportunity neighborhoods: Evidence from the hope vi program. Technical report, Opportunity Insights.

Chetty, R., J. Friedman, N. Hendren, M. R. Jones, and S. R. Porter (2022). Replication Data for: The Opportunity Atlas: Mapping the Childhood Roots of Social Mobility.

Chetty, R. and N. Hendren (2018). The impacts of neighborhoods on intergenerational mobility i: Childhood exposure effects. *The quarterly journal of economics 133*(3), 1107–1162.

Chetty, R., M. O. Jackson, T. Kuchler, J. Stroebel, N. Hendren, R. B. Fluegge, S. Gong, F. Gonzalez, A. Grondin, M. Jacob, et al. (2022). Social capital i: measurement and associations with economic mobility. *Nature 608*(7921), 108–121.

Cohen, E. D., H. Huo, K. Guyot, and C. Gaffney (2024, April). High school longitudinal study of 2009 (hsls:09): A first look at the 2021 postsecondary enrollment, completion, and financial aid outcomes of fall 2009 ninth-graders. Technical Report NCES 2024-022, National Center for Education Statistics, U.S. Department of Education, Washington, DC.

Cohodes, S. R. and A. C. Pineda (2026). Different paths to college success: The impact of massachusetts's charter schools on college trajectories. *American Economic Review: Insights 8*(2), 160–178.

Cook, P. J., K. Dodge, G. Farkas, R. G. Fryer, J. Guryan, J. Ludwig, S. Mayer, H. Pollack, L. Steinberg, et al. (2014). The (surprising) efficacy of academic and behavioral intervention with disadvantaged youth: Results from a randomized experiment in chicago. Technical report, National Bureau of Economic Research.

Cunha, F. and J. J. Heckman (2007). The technology of skill formation. *American Economic Review 97*(2), 31–47.

Cunha, F., J. J. Heckman, and S. M. Schennach (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica 78*(3), 883–931.

Davis, J. M., J. Guryan, K. Hallberg, and J. Ludwig (2017). The economics of scale-up. Technical report, National Bureau of Economic Research.

Dobbie, W. and R. G. Fryer (2020). Charter schools and labor market outcomes. *Journal of Labor Economics 38*(4), 915–957.

Duckworth, A. L., C. Peterson, M. D. Matthews, and D. R. Kelly (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology 92*(6), 1087–1101.

Dynarski, S. M., S. W. Hemelt, and J. M. Hyman (2015). The missing manual: Using national student clearinghouse data to track postsecondary outcomes. *Educational Evaluation and Policy Analysis 37*(1_suppl), $53S - -79S$.

Evans, D. K. and F. Yuan (2022). How big are effect sizes in international education studies? *Educational evaluation and policy analysis 44*(3), 532–540.

Experian (2019). When do deferred student loans show up on a credit report? `https://www.experian.com/blogs/ask-experian/student-loans-may-appear-in-credit-report-while-in-deferment/`. Accessed May 15, 2026.

Experian (2026). Income insight and income insight wage. `https://www.experian.com/business/products/income-insight`. Accessed May 15, 2026.

Falk, A., F. Kosse, and P. Pinger (2020). Mentoring and schooling decisions: Causal evidence. Technical report, IZA Discussion Papers.

Fan, J. and I. Gijbels (1996). *Local Polynomial Modelling and Its Applications*, Volume 66 of *Monographs on Statistics and Applied Probability*. London: Chapman and Hall.

Fryer Jr, R. G. (2011). Financial incentives and student achievement: Evidence from randomized trials. *The Quarterly journal of economics 126*(4), 1755–1798.

Gertler, P., J. Heckman, R. Pinto, A. Zanolini, C. Vermeersch, S. Walker, S. M. Chang, and S. Grantham-McGregor (2014). Labor market returns to an early childhood stimulation intervention in jamaica. *Science 344*(6187), 998–1001.

Gray-Lobe, G., P. A. Pathak, and C. R. Walters (2023). The long-term effects of universal preschool in boston. *The Quarterly Journal of Economics 138*(1), 363–411.

Guryan, J., J. Ludwig, M. P. Bhatt, P. J. Cook, J. M. Davis, K. Dodge, G. Farkas, R. G. Fryer Jr, S. Mayer, H. Pollack, et al. (2023). Not too late: Improving academic outcomes among adolescents. *American Economic Review 113*(3), 738–765.

Hanushek, E. A., P. E. Peterson, L. M. Talpey, and L. Woessmann (2019). The achievement gap fails to close. *Education Next 19*(3), 8–17.

Heckman, J. J., C. P. Loughlin, and H. Tian (2025). The benefits of scholastic athletics. Technical report, National Bureau of Economic Research.

Heller, S. B., A. K. Shah, J. Guryan, J. Ludwig, S. Mullainathan, and H. A. Pollack (2017). Thinking, fast and slow? some field experiments to reduce crime and dropout in chicago. *The Quarterly Journal of Economics 132*(1), 1–54.

Hendren, N. and B. Sprung-Keyser (2020). A unified welfare analysis of government policies. *The Quarterly journal of economics 135*(3), 1209–1318.

Hoxby, C. M. and C. Avery (2013). The missing "one-offs": The hidden supply of high-achieving, low-income students. *Brookings Papers on Economic Activity 2013*(1), 1–65.

Jackson, C. K. (2018). What do test scores miss? The importance of teacher effects on non–test score outcomes. *Journal of Political Economy 126*(5), 2072–2107.

Kautz, T., J. J. Heckman, R. Diris, B. ter Weel, and L. Borghans (2014, December). Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success. NBER Working Paper 20749, National Bureau of Economic Research.

Kling, J. R., J. B. Liebman, and L. F. Katz (2007). Experimental analysis of neighborhood effects. *Econometrica 75*(1), 83–119.

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational researcher 49*(4), 241–253.

Kraft, M. A., J. A. List, J. A. Livingston, and S. Sadoff (2022). Online tutoring by college volunteers: Experimental evidence from a pilot program. In *AEA Papers and Proceedings*, Volume 112, pp. 614–618. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

La Ferrara, E. (2016). Mass media and social change: Can we use television to fight poverty? *Journal of the European Economic Association 14*(4), 791–827.

List, J. A. (2022). *The voltage effect: How to make good ideas great and great ideas scale.* Crown Currency.

Massachusetts Department of Elementary and Secondary Education (2014). 2014 Report Card: Boston (00350000). `https://profiles.doe.mass.edu/reportcard/rc.aspx?fycode=2014&linkid=38&orgcode=00350000&orgtypecode=5`. Accessed May 22, 2026.

Mountjoy, J. (2022). Community colleges and upward mobility. *American Economic Review 112*(8), 2580–2630.

Mountjoy, J. (2026). Marginal returns to public universities. *Quarterly Journal of Eco-*

nomics 141(1), 429–497.

National Center for Education Statistics (2020). Table 219.46: Public high school 4-year adjusted cohort graduation rate (ACGR), by selected student characteristics and state: 2010–11 through 2017–18. https://nces.ed.gov/programs/digest/d19/tables/dt19_219.46.asp. Digest of Education Statistics 2019; accessed May 22, 2026.

National Center for Education Statistics (2026). Table 1. number and percentage distribution of students enrolled at title iv institutions, by control of institution, student level, level of institution, enrollment status, and other selected characteristics: United states, fall 2024. IPEDS Data Explorer, U.S. Department of Education. Provisional data from the 2024–25 collection; accessed April 3, 2026.

National Student Clearinghouse (2026). Studenttracker. https://www.studentclearinghouse.org/solutions/ed-insights/studenttracker/. Accessed May 15, 2026.

Nickow, A., P. Oreopoulos, and V. Quan (2020). The impressive effects of tutoring on prek-12 learning: A systematic review and meta-analysis of the experimental evidence.

Oreopoulos, P., R. S. Brown, and A. M. Lavecchia (2017). Pathways to education: An integrated approach to helping at-risk high school students. *Journal of political economy 125*(4), 947–984.

Resnjanskij, S., J. Ruhose, S. Wiederhold, L. Woessmann, and K. Wedel (2024). Can mentoring alleviate family disadvantage in adolescence? a field experiment to improve labor market prospects. *Journal of Political Economy 132*(3), 1013–1062.

Rodriguez-Planas, N. (2012). Longer-term impacts of mentoring, educational services, and learning incentives: Evidence from a randomized trial in the united states. *American Economic Journal: Applied Economics 4*(4), 121–139.

Schroeder, J., D. Van Riper, S. Manson, K. Knowles, T. Kugler, F. Roberts, and S. Ruggles (2025). Ipums national historical geographic information system. [dataset].

Stevenson, B. (2010). Beyond the classroom: Using title ix to measure the return to high school sports. *The Review of Economics and Statistics 92*(2), 284–301.

Tamborini, C. R., C. Kim, and A. Sakamoto (2015). Education and lifetime earnings in the united states. *Demography 52*(4), 1383–1407.

Tebes, J. (2014). Sink or Swim? An Evaluation of an Athletics Enrichment Program for Urban High School Students. *The MIT Undergraduate Journal of Economics 8*, 87–115.

U.S. Census Bureau (2010). State-based census tracts gazetteer files: Massachusetts. Gazetteer file, U.S. Census Bureau. Accessed March 16, 2026.

U.S. Census Bureau (2024). Current population survey, september 2023: Civic engagement and volunteering supplement technical documentation. Technical Report CPS-23, U.S. Census Bureau, Washington, DC. Accessed April 3, 2026.

U.S. Department of Education, National Center for Education Statistics (2024). Digest of Education Statistics 2024, Table 104.20: Percentage of Persons 25 to 29 Years Old with Selected Levels of Educational Attainment, by Race/Ethnicity and Sex: Selected Years, 1920 through 2024. Reports educational attainment among U.S. 25-to-29-year-olds in 2024. Data from the Current Population Survey Annual Social and Economic Supplement. Table prepared December 2024.

Walters, C. R. (2018). The demand for effective charter schools. *Journal of Political Economy 126*(6), 2179–2223.

Young, A. (2019). Channeling fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results. *The Quarterly Journal of Economics 134*(2), 557–598.

A Appendix

Table A.1: Program Participation Rates by Initial Lottery Assignment

	Control Mean (weighted)	Treatment Mean	First Stage	P-Value
	(1)	(2)	(3)	(4)
Ever attended program	0.314	0.984	0.683***	0.000
			(0.059)	
Total sessions attended	7.146	21.032	13.886***	0.000
			(2.811)	
Attended 5+ sessions	0.314	0.873	0.576***	0.000
			(0.072)	
Attended 10+ sessions	0.221	0.683	0.469***	0.000
			(0.078)	
Attended 15+ sessions	0.200	0.476	0.278***	0.001
			(0.082)	
Attended 20+ sessions	0.184	0.413	0.233***	0.005
			(0.081)	

Notes: Columns 1 and 2 show mean program participation rates for applicants who lost and won their first lottery, respectively. Control group means are weighted to reflect the distribution of lottery strata in the treatment group. In Column 3, the first-stage effect of an initial lottery win on program participation is estimated using Equation 1, with strata and school fixed effects and controls for self-reported baseline GPA and an indicator for missing self-reported baseline GPA. Column 4 contains p-values from the test that the first-stage effect is 0. Robust standard errors reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.2: Statistical Significance using Randomization Inference

	Control Mean (weighted)	Intention to Treat	P-value	
			Standard	Random. Inf.
	(1)	(2)	(3)	(4)
Panel A. High School Outcomes				
Standardized Semester-level GPA	0.000	0.487***	0.005	0.007
		(0.171)		
Standardized Cumulative GPA^{+}	0.000	0.476**	0.031	0.031
		(0.216)		
Graduated High School^{+}	0.749	0.172***	0.012	0.008
		(0.066)		
Panel B. Post-secondary Schooling Outcomes				
Post-Secondary Index	0.000	0.451**	0.008	0.010
		(0.166)		
Received Bachelor's Degree	0.551	0.170**	0.040	0.048
		(0.082)		
Panel C. Standardized Credit Market Indices				
Aggregate Credit Index	0.000	0.370**	0.021	0.034
		(0.159)		
Poor Financial Health Index	0.000	-0.430**	0.010	0.020
		(0.165)		
Joint Test Across Outcomes (excluding $^{+}$ outcomes)		***		0.003

Notes: Column 1 reports the control group mean weighted to reflect the distribution of lottery strata in the treatment group. Column 2 reports intention-to-treat coefficients using Equation 1 with standard errors given below in parentheses. Column 3 reports standard p-values while Column 4 reports randomization-inference p-values. We simulate the null distribution of t-statistics by re-running each lottery with appropriate strata-specific weights and running our main regression 1,000 times. Randomization p-values are estimated as follows: $p_{rand} = \frac{1+\sum_{r=1}^{R} |t_r| \geq |t_{true}|}{1+R}$ with $R = 1,000$. The bottom row reports an omnibus test that compares the true sum of squared t-statistics across all available outcomes to the simulated distribution of that same test statistic: $p_{\text{rand omnibus}} = \frac{1+\sum_{r=1}^{R} \mathbf{1}\left\{\sum_{j=1}^{J}\left(t_{jr}^{\text{sim}}\right)^2 \geq \sum_{j=1}^{J}\left(t_{j}^{\text{true}}\right)^2\right\}}{R+1}$ where j includes all outcomes in the table except for BPS outcomes. We omit BPS outcomes, denoted by $^{+}$, because the de-identified nature of BPS data prevents us from linking individuals to the other administrative datasets. Stars denote statistical significance using randomization p-values: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.3: Heterogeneous Treatment Effects

	All	Predicted Long-term Index		Grade at Application		Student Sex	
		Bottom	Top	9	10 - 12	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A. Short-term Standardized High School GPA							
Control Mean (wtd)	0.000	-0.407	0.431	0.196	-0.167	-0.170	0.147
Intention to Treat	0.487***	0.711***	0.231	0.907**	0.167	0.174	0.574***
(Standard Error)	(0.171)	(0.138)	(0.312)	(0.446)	(0.204)	(0.268)	(0.211)
{p-value}	{0.005}	{0.000}	{0.462}	{0.047}	{0.418}	{0.521}	{0.009}
Panel B. Post-secondary Schooling Index							
Control Mean (wtd)	0.000	-0.426	0.378	-0.127	0.150	-0.002	0.002
Intention to Treat	0.451***	0.820***	0.111	0.663**	0.075	0.190	0.628***
(Standard Error)	(0.166)	(0.228)	(0.318)	(0.301)	(0.271)	(0.282)	(0.191)
{p-value}	{0.008}	{0.001}	{0.728}	{0.032}	{0.784}	{0.506}	{0.002}
Panel C. Received Bachelor's Degree							
Control Mean (wtd)	0.551	0.400	0.685	0.484	0.630	0.490	0.599
Intention to Treat	0.170**	0.352***	0.049	0.301*	0.066	-0.039	0.263**
(Standard Error)	(0.082)	(0.119)	(0.137)	(0.153)	(0.129)	(0.151)	(0.106)
{p-value}	{0.040}	{0.005}	{0.721}	{0.054}	{0.610}	{0.799}	{0.016}
Panel D. Aggregate Credit Index							
Control Mean (wtd)	0.000	-0.375	0.333	-0.036	0.042	0.055	-0.044
Intention to Treat	0.370**	0.634***	-0.199	0.719***	0.215	-0.006	0.446**
(Standard Error)	(0.159)	(0.198)	(0.227)	(0.248)	(0.198)	(0.258)	(0.194)
{p-value}	{0.021}	{0.002}	{0.385}	{0.005}	{0.282}	{0.983}	{0.024}
Panel E. Poor Financial Health Index							
Control Mean (wtd)	0.000	0.422	-0.374	-0.006	0.007	0.003	-0.003
Intention to Treat	-0.430**	-0.783***	0.137	-0.850***	-0.201	-0.263	-0.383*
(Standard Error)	(0.165)	(0.212)	(0.242)	(0.258)	(0.223)	(0.268)	(0.208)
{p-value}	{0.010}	{0.000}	{0.574}	{0.002}	{0.371}	{0.330}	{0.068}

Notes: Column 1 reports full-sample estimates. Columns 2 and 3 split the sample by whether students are below or above the median of the predicted long-term outcome index. This index is constructed by regressing baseline covariates from a student's initial application on the average of the post-secondary schooling index and the most recent quarter of the aggregate credit index among control students. For control students, predictions are generated using leave-one-out fitted values to avoid mechanical overfitting; for treated students, predictions use coefficients estimated from the full control group. Columns 4 and 5 split the sample by grade at application, comparing students who applied in ninth grade to those who applied in grades 10–12. Columns 6 and 7 split the sample by student sex. Control means are weighted to match the treatment group's distribution of lottery strata. Intention-to-treat estimates are obtained from Equation 1, estimated separately for each subgroup, including lottery-strata and school fixed effects and controls for self-reported baseline GPA and an indicator for missing baseline GPA. Panels D and E additionally control for student age in years. Standard errors are reported in parentheses and p-values in braces. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.4: Effects on Leading Labor Market Credit Indicators: All Available Quarters

	Control Mean	ITT	ITT
	(1)	(2)	(3)
Specification		*Basic*	*All Controls*
Panel A. Standardized Summary Indices			
Aggregate Credit Index	0.000	0.338**	0.354**
		(0.157)	(0.173)
Poor Financial Health Index	0.000	−0.338**	−0.326*
		(0.158)	(0.174)
Panel B. Credit Access			
Has an Experian record	0.890	0.059	0.013
		(0.046)	(0.043)
Panel C. Credit Score			
Vantage Credit Score	647.988	22.344*	25.004*
		(12.903)	(13.024)
log(Vantage Credit Score)	6.463	0.036*	0.041**
		(0.020)	(0.021)
Prime Credit (i.e., Score > 660)	0.500	0.135**	0.161**
		(0.067)	(0.068)
Poor Credit (i.e., Score < 600)	0.347	−0.112*	−0.146**
		(0.062)	(0.067)
Panel D. Assets			
Mortgage	0.082	0.008	0.023
		(0.039)	(0.036)
Auto loan	0.288	0.002	0.032
		(0.070)	(0.068)
Panel E. Financial Distress			
Collections Balance > $0	0.165	−0.047	−0.067
		(0.054)	(0.057)
Any bills delinquent for 30 days	0.537	-0.111^{+}	−0.145*
		(0.076)	(0.077)
Any bills delinquent for 90 days	0.389	−0.091	−0.111
		(0.076)	(0.078)
Panel F. Student Loans			
Any student loan balance	0.691	−0.104	−0.083
		(0.076)	(0.076)
Student loan balance over past 3 months (in $1,000s)	$23.632	−$4.532	−$2.140
		(4.610)	(4.281)

Notes: This table includes all available credit bureau data spanning quarters Q3 2021 through Q3 2025. Median sample age is 24.5 years in the first observed quarter and 28.5 years in the last quarter. Predicted earnings data are only available for the final 8 quarters and thus are omitted from this table. Col 1: mean credit bureau outcomes for the control group re-weighted to match distribution of lottery strata in the treatment group. Columns 2 and 3 report the effect of an initial lottery win, estimated using Equation 1, including strata, school, and age-year fixed effects. Regressions in Col 2 control for self-reported baseline GPA and missing indicator. Regressions in Col 3 include the full set of baseline controls. Aggregate Credit Index includes positive standardized measures of *log*(Vantage credit score), and having: an Experian record, prime credit, a mortgage, an auto loan, and negative standardized measures of student loans balance over past 3 months and having poor credit, any collections balance, delinquency in past 30 days, delinquency in past 90 days, any student-loan balance in the last 3 months. Poor Financial Health Index includes standardized measures of having no Experian record, poor credit, any collections balance, delinquency in past 30 days, delinquency in past 90 days, and any student loans balance. SEs clustered at individual level reported in parentheses. Obs = 2,210, N = 130. $^{+}$ $p < 0.15$, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.5: This Study in Context: Comparison of Leading Mentorship Programs in Adolescence

Study	Program & setting	Design	Components	Horizon	Key result(s)	Cost per participant
This paper (Angrist & Tebes, 2026)	Amphibious Achievement; disadvantaged Boston public-HS students; MIT campus	RCT (admissions lotteries)	Athletics (swim/row) + tutoring + near-peer college-volunteer mentorship	Long run (10+ yrs, to late 20s)	GPA +0.49 SD; HS graduation +17 pp; bachelor's degree +17–21 pp; labor market and credit index +0.37 SD	˜$170/yr
Oreopoulos, Brown & Lavecchia (2017), *JPE*	Pathways to Education; Toronto public-housing youth	Quasi-experimental (comparison community)	Tutoring + mentoring + financial incentives	Medium/long run	HS graduation +15 pp (focal site), +6 pp (at scale); higher PS enrollment	High (comprehensive; cash transfers)
Rodriguez-Planas (2012), *AEJ: Applied*	Quantum Opportunity Program; U.S. urban high schools	RCT	Mentoring + educational services + incentives	Long run	Small short-run gains; *no* long-run gains in bachelor's degree or labor market	High (multi-year, stipends)
Falk, Kosse & Pinger (2020), *IZA DP*	Germany; low-SES children/adolescents	RCT	1:1 volunteer mentoring	Medium run	Closes ∼1/3 of the SES gap in high-track entry	Low (volunteer mentors)
Resnjanskij et al. (2024), *JPE*	Germany; disadvantaged adolescents	RCT	University-student mentoring	Short run	Improves math, patience/social skills, and labor-market orientation (esp. for disadvantaged)	Low (university mentors)

Notes: HS = high school; PS = post-secondary; pp = percentage points; SD = standard deviations; RCT = randomized controlled trial. Comparison entries summarize headline findings as characterized in this paper and the original studies. Pathways to Education was evaluated with a quasi-experimental comparison-community design; the remaining comparison studies are randomized. Costs are approximate, per enrolled participant per year where reported (Amphibious Achievement in 2025 dollars).

Table A.6: Program Features Across Leading Mentorship Studies in Adolescence

	RCT	Athl.	Tutor	Ment.	Long run	Effects Persist	Labor/ fin.	Low cost
This paper (Angrist & Tebes, 2026)	✓	✓	✓	✓	✓	✓	✓	✓
Oreopoulos et al. (2017), Pathways	×	×	✓	✓	✓	✓	×	×
Rodriguez-Planas (2012), QOP	✓	×	✓	✓	✓	×	✓	×
Falk, Kosse & Pinger (2020)	✓	×	×	✓	×	?	×	✓
Resnjanskij et al. (2024)	✓	×	×	✓	×	?	×	✓

Notes: ✓ = feature present, evidence collected, or effect demonstrated; it does not denote effect size or sign. Column keys: RCT = randomized design; Athl. = athletics; Tutor = academic tutoring; Ment. = mentoring; Long run = post-secondary or adult outcomes tracked; Persist. = program effects shown to persist rather than fade; Labor/fin. = adult earnings, credit, or financial outcomes examined (positive here; tracked but null in Rodriguez-Planas, 2012); Low cost is approximate (Amphibious Achievement is lowest, about $170/participant/yr). Long run and Persist. are distinct: Rodriguez-Planas (2012) tracked long-run outcomes but found that short-run gains faded. Pathways to Education was evaluated quasi-experimentally; all other comparison studies are randomized. Amphibious Achievement is the only program combining athletics with academic support.

Figure A.1: Marginal Value of Public Funds (MVPF) Welfare Comparison



Notes: This figure plots the MVPF of Amphibious Achievement relative to a set of over 200 high-profile U.S. policies analyzed (Hendren and Sprung-Keyser, 2020). As illustrated, the program ranks in the 80th percentile of welfare gains, tied with the 20% of the sample with infinite MVPF.

Figure A.2: Sample Semester of Personal Records in Athletics



Notes: This figure plots a sample semester with Personal Record milestone markers. The y-axis is a set of students and the x-axis is a given session. A session is highlighted if a personal record was achieved per student and session.

Figure A.3: Sample Semester Schedule and Personal Records in Academics

AmphibiousAchievement

success in and out of the water

Program Schedule — Summary

The academic program runs as **ten consecutive Sunday sessions**. Each Sunday follows the same arc — athletics (row or swim), a hearty lunch, then academics. Diagnostic exams bookend the reading and math tracks — the checkpoints where Achievers set and break their personal records — while take-home OTAs and team competition recur throughout.

Wk	Critical Reading	Math	Grammar & Writing
1	**Intro to Critical Reading** Class models reading an article; groups read & answer SAT questions; present summaries	**Kickoff + fun math video** Set up the geometry focus; build excitement for the curriculum	**Grammar Jeopardy!** Intro game; assess which grammar topics need work
2	**Intro to SPEAK** Acronym + good/bad examples; present summaries with SPEAK OTA	Diagnostic – Personal Records **Math Diagnostic**	**Intro to Thesis Statements** Definition, strong vs. weak; turn a thesis into an essay outline
3	Diagnostic – Personal Records **Reading Diagnostic**	**Area & Perimeter of Polygons** Puzzler introduced + solved in tutoring groups OTA	**Grammar Review + Jeopardy!** Pronouns & verb agreement
4	**Vocabulary Skits begin** Daily lunch vocab (3 words/day); present with SPEAK OTA	**Polygons — Practice** SAT questions; solutions presented at the board	**Thesis in Groups** SAT prompt; groups present + class critiques
5	**Review reading techniques** Group summaries delivered with SPEAK	**Triangles** Puzzler introduced + solved in tutoring groups OTA	**Grammar Review + Jeopardy!** Parallelism, perfect tense, voice, subjunctive
6	**Public Speaking Workshop** Mentor-led session + follow-up activities OTA	**Triangles — Practice** SAT questions; solutions presented at the board	**Thesis in Groups** SAT prompt; groups present + class critiques
7	**Reading practice** Group summaries delivered with SPEAK	**Circles** Puzzler introduced + solved in tutoring groups OTA	**Grammar Review + Jeopardy!** Semicolons, apostrophes, colons
8	Diagnostic – Personal Records **Reading Diagnostic** Beat your earlier score	**Circles — Practice** SAT questions; solutions at the board OTA	**Thesis from an Article** Group conviction → individual → master thesis
9	**Reading practice** Group summaries delivered with SPEAK OTA	Diagnostic – Personal Records **Math Diagnostic** Beat your earlier score	**Grammar Review + Jeopardy!** Misplaced & dangling modifiers
10 ★	**Class Debate (extended)** PRO/CON teams; SPEAK + specific article evidence	**Volume** Puzzler introduced + solved in tutoring groups (practice follows)	**Final & Fun Review** A surprise!

Critical Reading Math Grammar & Writing Diagnostic – Personal Records two per track, bookending

OTA take-home "Opportunity to Achieve" ★ capstone session

The weekly rhythm

Critical Reading: model the skill, then read & present every week with SPEAK; diagnostics at Wk 3 & 8; builds to a Week-10 debate.

Math: paired concept + practice weeks (Polygons · Triangles · Circles); diagnostics at Wk 2 & 9; closes on Volume.

Grammar & Writing: alternates grammar-review + Jeopardy! (odd weeks) with thesis-building (even weeks), ending with an article-based thesis and a surprise finale.

Math content continues into a Volume Practice session beyond Week 10; vocabulary skits recur daily at lunch from Week 4 onward.

Figure A.4: Sample Semester Schedule in Athletics

AmphibiousAchievement

success in and out of the water

Athletics — Rowing (Crew)

A typical Sunday on the water

- 9:00 **Set up ergs**
- 9:10 **Warm-up** — catch · finish · pick drill · power 10
- 9:20 **Team meeting** / motivation
- 9:30 **5-minute piece** — personal record
- 9:35 **Stations** — tanks · sliders · boats
- 10:20 **Relays / race piece**
- 10:35 **End of day reflection**

Wk	Rowing focus & key work	Setting & milestone
1	**Meet the stroke** What rowing is & why we row; catch & finish positions, pick drill, Lean With It–Rock With It, butt-off-the-seat, stand-up; 300m segmented relay (arms→body→legs).	Ergs
2	**Refine the stroke** Nutrition & reading split numbers; partner pick drill, "cut the cake"; relays (pause at ¾, 200m); watch a rowing clinic.	Ergs
3	**Connection & timing** US Rowing & what makes a team; domino pick drill in unison; connect-2 on sliders, 1-2 drive, spoon rowing; 500m connected race.	Sliders
4	**Continuous rowing** "A rower on & off the water"; 3×3-min pieces (technique, leg connection); cut the cake; relays.	Sliders
5	**Power & teamwork** Motivation + Power 10; pick drill + power 15s; sliders → tanks; relays.	Tanks Captain vote
6	**Racing strategy** Power 15 (easy→hard→easy); partner slider work; tanks; 3×3-min pieces with 1-on-1 coaching.	Tanks
7	**Practice & focus** "Practice + meditate"; power 15s; missing a goal is the point of pushing yourself; free individual rowing; relays.	Tanks 2 wks to water
8	**Boats begin** Captain talk + voting; warm-up & 5-min piece; tanks ↔ boat-training rotation; boat carrying.	Boats Swim test
9	**On the water** Final captain vote; warm-up & 5-min piece; tanks ↔ boats; carry & launch.	Water
10 ★	**Race day** Launch; 5-min piece; two shifts (tanks + water); water drills — tapps, set the boat, pick by pairs, lean/rock, full rowing by pairs & fours.	Water Race day

Signature touches

A student-made "Teach Me How to Row" video (a riff on the academics' "Teach Me How to Factor"), on-erg music drills (Lean Wit It–Rock Wit It, Crank It Up, Va Va Voom), a 5-minute PR piece every session, captain elections, and an end-of-term banner signing with photos on the dock.



Academic Curriculum

Academic Mission Statement

The mission of Amphibious Achievement, an athletic and academic mentorship program, is to promote academic and athletic excellence. Amphibious Achievement's academic portion will encourage constructive study habits and expand the Achiever's mastery of critical reading, math, and grammar through a combination of mini-lectures and individualized tutoring. In particular, our academic goal is to help our Achievers obtain important skills that are important to but also transcend standardized tests, such how to become expert problem-solvers, critical thinkers, and succeed under time-pressure.

The academic portion is split into three parts: critical reading, math, and writing. In our reading portion, our mentors guide our Achievers through an active reading of *The Economist*. Using *The Economist*, an advanced and challenging magazine, prepares the Achievers to tackle any passage of text they may encounter in different aspects of their life, from SAT passages to reading a chapter in their World History textbook to a job description in the Sunday newspaper. The math portion utilizes YouTube clips, and challenging puzzles. The math portion also uses custom-made practice questions, which often incorporate our Achiever's experiences, such as splitting up Au Bon Pain pastries to learn fractions. In this way, our math portion engages our Achievers, teaches basic math skills, and teaches Achievers how to problem-solve. The writing portion involves a competitive Jeopardy!-style game where students compete in teams to revise incorrect sentences on the board and then explain their corrections for full credit, in the shortest amount of time. At the end of the program, the team which earns the most Jeopardy! points wins the opportunity to choose a local charity to donate an open-ended $250 grant on behalf of all of Amphibious Achievement.

All three components of the academic curriculum reflect our five core ideals, **friendly competition, tangible incentives, advanced resources, high expectations, and fun**.

Critical Reading

Schedule

WEEK ONE: ***Introduction to Critical Reading Techniques***

- Entire class learns how to critically read an article together with an example paragraph on the board
- Achievers read articles and answer original SAT questions in tutoring groups
- Each group presents a concise article summary and a few relevant details to the rest of the class

WEEK TWO: ***Introduction to SPEAK***

- SPEAK acronym introduced (good and bad presentation examples)
- Achievers read articles and answer original SAT questions in tutoring groups
- Each group presents a concise article summary to the rest of the class while practicing SPEAK
- *OTA distributed*

WEEK THREE: ***Critical Reading Diagnostic Exam 1***

- Achievers work through diagnostic exam

WEEK FOUR: ***Introduction of Vocabulary Skits***

- During lunch mentors present vocab skit to teach three new vocab words (will happen each day during lunch from this day onward)
- Achievers read articles and answer original SAT questions in tutoring groups
- Each group presents a concise article summary to the rest of the class while practicing SPEAK
- *OTA distributed*

WEEK FIVE: ***Review Critical Reading Techniques***

- Review important techniques again with entire class
- Achievers read articles and answer original SAT questions in tutoring groups
- Each group presents a concise article summary to the rest of the class while practicing SPEAK

WEEK SIX: ***Public Speaking Workshop***

- One mentor presents a short workshop on being an effective public speaker
- Follow up activities from workshop presentation
- *OTA distributed*

WEEK SEVEN: ***Critical Reading Practice***

- Achievers read articles and answer original SAT questions in tutoring groups
- Each group presents a concise article summary to the rest of the class while practicing SPEAK

WEEK EIGHT: ***Critical Reading Diagnostic Exam 2***

- Achievers work through diagnostic exam

WEEK NINE: ***Critical Reading Practice***

- Achievers read articles and answer original SAT questions in tutoring groups
- Each group presents a concise article summary to the rest of the class while practicing SPEAK
- *OTA distributed*

WEEK TEN: ***Class Debate*** *(extended session)*

- Achievers read articles and answer original SAT questions in tutoring groups
- Based on PRO/CON assignments, tutoring groups then have a debate on the topic of the day
- During debate all Achievers should practice their SPEAK skills and use specific examples from assigned articles to strength their arguments

How To Actively Read *The Economist*

1. **Analyze the title: What do you think the article will be about?**
   -Notice any interesting clues, puns, cultural references, etc
   -Use supplementary pictures or graphs as hints

2. **Ask questions as you read: What is the author trying to say?**
   -Pay attention to tone and mood
   -Think about information provided and why it is included
   -Question the meaning behind the words (read "between the lines")

3. **Visualize what the text is describing: Who? What? Where?**
   -Use senses that help you connect to the text
   -Paint a mental picture or literally draw what is being described

4. **Pick out interesting vocabulary words: How can I figure out the definition?**
   -Dissect sentence structure surrounding the word
   -Find context clues that suggest the definition

5. **Note interesting punctuation usage: What is its function?**
   -Figure out why the author is using a particular punctuation mark
   -Evaluate the effectiveness of this usage

6. **Synthesize information: What is the main focus?**
   -Re-read all of the topic and concluding sentences of each paragraph
   -Outline the supporting topics/points presented in the article
   -Create your own thesis based on the author's purpose

7. **Make connections to the outside world: Where else is this relevant for me?**
   -Reflect on what you learned
   -Look for associations between what you already knew and this new information

How to Present Your Ideas

Remember to always...

Stand up

Project

Eye-contact

Articulate

Keep it cool

Math

Schedule

WEEK ONE: ***Fun Math Video and Introduction of Math Curriculum Structure***

- Watch YouTube video as a class and tie into Geometry focus for the next three semesters
- Get Achievers excited for the new math curriculum structure

WEEK TWO: ***Math Diagnostic Exam 1***

- Achievers work through diagnostic exam

WEEK THREE: ***Area and Perimeter of Polygons***

- Quick review of topic with entire class
- Activity/puzzler introduced with the entire class
- Achievers work on activity/puzzler with tutoring group
- Activity solution and purpose discussed with entire class
- *OTA distributed*

WEEK FOUR: ***Area and Perimeter of Polygons Practice***

- Quick review of topic and material covered the previous week
- Achievers work on original SAT questions in tutoring group
- Achievers present solutions to the SAT questions on the board to entire class

WEEK FIVE: ***Triangles***

- Quick review of topic with entire class
- Activity/puzzler introduced with the entire class
- Achievers work on activity/puzzler with tutoring group
- Activity solution and purpose discussed with entire class
- *OTA distributed*

WEEK SIX: ***Triangles Practice***

- Quick review of topic and material covered the previous week
- Achievers work on original SAT questions in tutoring group
- Achievers present solutions to the SAT questions on the board to entire class

WEEK SEVEN: ***Circles***

- Quick review of topic with entire class
- Activity/puzzler introduced with the entire class
- Achievers work on activity/puzzler with tutoring group
- Activity solution and purpose discussed with entire class
- *OTA distributed*

WEEK EIGHT: ***Circles Practice***

- Quick review of topic and material covered the previous week
- Achievers work on original SAT questions in tutoring group
- Achievers present solutions to the SAT questions on the board to entire class
- *OTA distributed*

WEEK NINE: ***Math Diagnostic Exam 2***

- Achievers work through diagnostic exam

WEEK TEN: ***Volume***

- Quick review of topic with entire class
- Activity/puzzler introduced with the entire class
- Achievers work on activity/puzzler with tutoring group
- Activity solution and purpose discussed with entire class

Grammar and Writing

Schedule

WEEK ONE: ***Grammar Jeopardy!***

- Introduce Grammar Jeopardy! to Achievers
- Achievers engage in a game of Grammar Jeopardy!
- Assess which grammar topics Achievers are comfortable with and need more help in

WEEK TWO: ***Introduction to Writing a Thesis Statement***

- Go over definition of a thesis statement and explain its importance
- Discuss what makes a thesis statement strong/weak
- Go over a sample introduction paragraph and have Achievers identify the thesis statement
- Provide Achievers with a sample thesis statement and show how to transform it into the outline of an essay

WEEK THREE: ***Grammar Review and Grammar Jeopardy!***

- Review pronouns and verb agreement
- Achievers engage in a game of Grammar Jeopardy!

WEEK FOUR: ***Improving Thesis Statements in Groups***

- Achievers are given an SAT writing prompt and come up with a thesis statement in their groups
- Groups present thesis statements and class discusses the statements' strengths/weaknesses

WEEK FIVE: ***Grammar Review and Grammar Jeopardy!***

- Review parallelism, prefect tense, voice, and subjunctive mood
- Achievers engage in a game of Grammar Jeopardy!

WEEK SIX: ***Improving Thesis Statements in Groups***

- Achievers are given an SAT writing prompt and come up with a thesis statement in their groups
- Groups present thesis statements and class discusses the statements' strengths/weaknesses

WEEK SEVEN: ***Grammar Review and Grammar Jeopardy!***

- Review semicolons, apostrophes, and colons
- Achievers engage in a game of Grammar Jeopardy!

WEEK EIGHT: ***Creating a Thesis Statement Based on an Article***

- Groups discuss main themes of the article from the Critical Reading portion and come up with a conviction about the article
- Achievers individually create their own thesis statements based on the discussed convictions
- Achievers come back together and share thesis statements before forming a master thesis statement
- Groups present their thesis statements and class discusses strengths/weaknesses of each thesis statement

WEEK NINE: ***Grammar Review and Grammar Jeopardy!***

- Review misplaced and dangling modifiers
- Achievers engage in a game of Grammar Jeopardy!

WEEK TEN: ***Final and Fun Review Activity***

- A surprise!