Who Benefits from a Smaller Honors Track? - ONLINE APPENDIX*

Zachary Szlendak
IDA

Richard Mansfield
University of Colorado-Boulder

January 23, 2024

*Zachary Szlendak is an economist at the Institute for Defense Analyses. Richard Mansfield is an assistant professor at the University of Colorado-Boulder. We thank Terra McKinnish, Brian Cadena, Francisca Antman, Allison Atteberry, Taylor Jaworski, Corey Woodruff, and Brachel Champion as well as seminar participants at the Institute for Defense Analyses and CNA, the editor and two insightful but anonymous referees for helpful comments and discussions. This research uses confidential data from the North Carolina Education Research Data Center (NCERDC) at Duke University. We acknowledge the North Carolina Department of Public Instruction and the NCERDC for collecting and providing this information. The data can be requested via application through the NCERDC. The authors have no conflicts of interest to disclose. Contact: zachszlendak@gmail.com, richard.mansfield@colorado.edu (corresponding author). JEL codes: I21, I24, J24.

Appendix

A Theoretical Model

In this appendix we introduce a simple education production function and classroom sorting equilibrium in order to formally derive a methodology for estimating the subgroup-specific functions mapping honors shares f into expected achievement, $E[\Delta\overline{Y}_q(f)]$, which are inputs required for the administrator's problem (2). In the process, we also highlight the assumptions our approach requires.

A.1 Test Score Production

Let Y_{istj} capture the standardized test score of student i in course j taken at school s during year t. We model the educational production function as follows:

$$Y_{istj} = d^h_{istj} h(q_{istj}, \epsilon_{istj} | \vec{q}_h, \vec{\epsilon}_h) + d^r_{istj} r(q_{istj}, \epsilon_{istj} | \vec{q}_r, \vec{\epsilon}_r) + X^O_{istj}\beta^O + X^U_{istj}\beta^U + \mu_{istj}. \quad \text{(A.1)}$$

The student's choice of track is represented by the indicator variables d^h_{istj} and d^r_{istj}, with values of 1 signifying enrollment in honors and regular tracks, respectively. Schools that do not offer separate tracks in a given course feature both $d^h_{istj} = 0$ and $d^r_{istj} = 0$. The functions $h(q_{istj}, \epsilon_{istj} | \vec{q}_h, \vec{\epsilon}_h)$ and $r(q_{istj}, \epsilon_{istj} | \vec{q}_r, \vec{\epsilon}_r)$ capture shifts in achievement from taking the honors and regular tracks, respectively. These shifts are functions of the student's own inputs, which are partly predictable based on the student's observable subgroup q_{istj} but also depend on an unobservable idiosyncratic component ϵ_{istj}. ϵ_{istj} captures deviations in expected performance due to, for example, accumulated skills or effort unaccounted for by subgroup. Such deviations vary not just across students but within students across school-course-year combinations. Importantly, the impact of the track choice on achievement also depends on the peer environment within the chosen track, which is reflected in the dependence of the functions $h(\cdot)$ and $r(\cdot)$ on the vectors $(\vec{q}_h, \vec{\epsilon}_h)$ and $(\vec{q}_r, \vec{\epsilon}_r)$ capturing the subgroups and idiosyncratic contributions of other members of the honors and regular tracks. This flexible formulation of track effects acknowledges that students' production in the classroom will be affected by how the material matches with their ability and how the peer environment interacts with their own ability and effort. Track-specific teacher inputs and course rigor are assumed to be functions of the kinds of students selecting into the track in a given school-course-year, and thus are implicitly captured by the functions $h(\cdot)$ and $r(\cdot)$.

X^O_{istj} and X^U_{istj} capture other student, school, and course inputs that affect i's learning that observed and unobserved, respectively, while μ_{istj} captures measurement error that causes the test score to fail to perfectly reflect the student's learning in the chosen course. Importantly, by imposing that these inputs are additively separable from the inputs that enter the track-specific functions $h(\cdot)$ and $r(\cdot)$, we have assumed they have the same impact on test scores regardless of track. This implicitly requires that the standardized tests used to assess knowledge in each course do not depend on the track chosen, which is true in the

North Carolina context we consider.[1] While somewhat restrictive, the additive separability assumption implies that these inputs are irrelevant to the administrator's tracking problem. Thus, we can rewrite achievement in terms of the difference between performance in the chosen track and performance in a pooled version of the course with no tracks:

$$\Delta Y_{istj} = d^h_{istj} h(q_{istj}, \epsilon_{istj} | \vec{q}_h, \vec{\epsilon}_h) + d^r_{istj} r(q_{istj}, \epsilon_{istj} | \vec{q}_r, \vec{\epsilon}_r). \tag{A.2}$$

Recasting achievement production this way facilitates a focus on the interactions between the student and peer characteristics that are likely to be of primary importance. Note that this formulation is nonetheless less restrictive than many linear specifications in the literature, since it allows the impact of observed and unobserved student ability components q and ϵ to depend on each other and on the choice of track.

A.2 A Simple Model of Student Track Choice

Now consider the student's choice of honors vs. regular track in a course that features only these two tracks. Suppose that each student chooses the track that maximizes his or her test score net of track-specific effort costs, scheduling opportunity costs, and GPA boosts. Let c_{istj} capture student i's idiosyncratic composite cost (measured in test-score utility equivalents) of joining the honors track h relative to the regular track r at school s at time t in course j. Next, let α_{stj} capture a component of the composite cost of the honors track that is common to all students in (s, t, j). Importantly, assume that the administrator has the ability to shift α_{stj} by any arbitrary amount by adjusting the relative GPA boost or homework load in the honors track.

The student's track choice can thus be written as:

$$d^h_{istj} = \begin{cases} 1, & \text{if } \underbrace{h(q_{istj}, \epsilon_{istj} | \vec{q}_h, \vec{\epsilon}_h) - r(q_{istj}, \epsilon_{istj} | \vec{q}_r, \vec{\epsilon}_r)}_{\text{Difference in academic gains}} \underbrace{-c_{istj} - \alpha_{stj}}_{\substack{\text{Effort, convenience,} \\ \text{and grade cost}}} > 0 \\ 0, & \text{otherwise} \end{cases}$$

Note that because we allow a student's unobserved ability to differentially affect their academic performance in the honors versus the regular track, we are accommodating the possibility that students may select into courses based on unobserved ability to benefit. Along with peer effects, such "selection on gains" generally complicate efforts to extrapolate from track impacts on marginal students to broader average treatment effects of interest. We show here that peer effects and selection on unobserved gains need not undermine the ability to estimate the key inputs to the administrator's tracking problem.

Next, let $g_{stj}(\epsilon, c|q)$ denote the cohort's joint conditional distribution of students' unobserved ability components and idiosyncratic effort/scheduling costs for any given subgroup q. To simplify notation, we assume that the school cohorts in consideration are large enough and the ability groups are few enough to approximate $g_{stj}(\epsilon, c|q)$ for each q with a continuous

[1] Furthermore, administrator, parent, and student preferences for high scores help ensure that the curricula for the two tracks do not diverge too far from one another.

joint density. Then we can define $\alpha^*_{stj}(f)$ as the threshold common cost component α^*_{stj} that causes a fraction f of students in the chosen school-year-course to choose the honors track. Specifically, $\alpha^*_{stj}(f)$ is implicitly defined as the solution to the following equation:[2]

$$\sum_q W_{stq} \iint d^h_{istj}(\alpha_{stj}, q, \vec{q}_h(\alpha_{stj}), \vec{q}_r(\alpha_{stj}), \epsilon, c) g_{stj}(\epsilon, c|q) d\epsilon dc = f. \tag{A.3}$$

Next, we assume that the composition of students across schools, years, and courses is very similar among a large subset $\mathcal{S}$ of school-year-course combinations:

Assumption 1. $g_{stj}(\epsilon, c|q) \approx g(\epsilon, c|q)\ \forall q\ \forall (s,t,j) \in \mathcal{S}$ *and* $W_{stjq} \approx W_q\ \forall\ (s,t,j) \in \mathcal{S}$

Under Assumption 1, as courses become large the threshold cost function $\alpha^*_{stj}(f)$ becomes common among sufficiently similar schools and course-year combinations within schools: $\alpha^*_{stj}(f) \approx \alpha^*(f)$ for all $(s,t,j) \in \mathcal{S}$. Furthermore, because the conditional distribution $g(\epsilon, c|d^h, q)$ also becomes common, the vectors of track-specific peers $(\vec{q_r}, \vec{\epsilon_r})$ and $(\vec{q_h}, \vec{\epsilon_h})$ also depend only on f (through $\alpha^*(f)$) rather than separately on s, t, or j. This in turn implies that $h(q_{istj}, \epsilon_{istj}|\vec{q}_h, \vec{\epsilon}_h) \approx h(q_{istj}, \epsilon_{istj}|f_{stj})$ and $r(q_{istj}, \epsilon_{istj}|\vec{q}_r, \vec{\epsilon}_r) \approx r(q_{istj}, \epsilon_{istj}|f_{stj})$. It also implies that the subgroup-specific probability of choosing honors depends only on f:

$$P(d^h = 1|q_{istj} = q, f) = \iint d^h(\alpha^*(f), \epsilon, c, q) g(\epsilon, c|q_{istj} = q) d\epsilon dc \tag{A.4}$$

Thus, the implicit choice of f by the administrator (through $\alpha^*(f)$) can serve as a sufficient statistic for the peer composition of both the honors and regular tracks in all school-year-course combinations where this common joint distribution of ability and costs represents a sufficiently close approximation. Essentially, this assumption rules out heterogeneous treatments across schools or courses for the same honors fraction, so that differences in achievement distributions across schools or courses featuring different honors fractions can be interpreted as (possibly heterogeneous) treatment effects.

In our empirical work, we attempt to make this approximation plausible by 1) removing schools from our sample whose students exhibit a distribution of past performance on state exams that is too far from the state norm, and 2) controlling for the shares and mean test scores of students in the chosen school-year-course with predicted test scores in each decile of the statewide predicted distribution, interacted with the student's own quintile of predicted performance.

Note that Assumption 1 is sufficient but not necessary for the fraction in honors to serve as a sufficient condition for peer environment. Suppose, for example, that each student's relative performance across tracks depends only on the tracks' relative peer quality rather than separately on the absolute peer quality in each track. Then estimates of tracking effects may be unbiased even if comparisons are made between schools or school-year-courses with

[2] Note that since d^h_{istj} depends on α_{stj} both directly and indirectly through the peer vectors $\vec{q}_h(\alpha_{stj}), \vec{\epsilon}_h(\alpha_{stj})$, we must assume that the track-specific achievement functions $h(\cdot)$ and $r(\cdot)$ are sufficiently insensitive to small changes in peer composition that the fraction choosing honors is monotonically and smoothly decreasing in α_{stj} for each q and spans a large range of fractions for feasible administrator choices of α_{stj}. This ensures that there exist unique solutions to equation A.3 for a wide range of f values.

different baseline student quality distributions, as long as these units would feature the same peer quality differences across tracks if they chose the same honors fraction f and controls for the direct achievement effects of cohortwide quality distributions are included (as they are in our empirical work).

Even if the distributions $g(\epsilon, c|q)$ are roughly common among schools, however, they may not be known by any school administrator, since both ϵ and c are unobserved for each student. Thus, any given principal will have a difficult time inferring both $g(\epsilon, c|q)$ and the track-specific achievement functions $h(q, \epsilon|f)$ and $f(q, \epsilon|f)$ from data on student performance.

Note, though, that the administrator's problem (1) only requires as inputs $E[\Delta\overline{Y}_q(f)]$, the subgroup-specific mean test score performance gains relative to a trackless course as functions of the honors fraction f. Thus, we can exploit the fact that $E[\Delta\overline{Y}_q(f)]$ can be written as a simple weighted average of the expected track-specific performance of the subsets of group q that sort into the honors and regular tracks, respectively:

$$E[\Delta\overline{Y}_q(f)] = P(d^h = 1|q, f)E[h(q, \epsilon|f)|d^h = 1] + P(d^r = 1|q, f)E[r(q, \epsilon|f)|d^r = 1] \quad \text{(A.5)}$$

Since the conditional expectation functions $E[h(q, \epsilon|f)|d^h = 1]$ and $E[r(q, \epsilon|f)|d^r = 1]$ in (A.5) depend only on $g(\epsilon, c|q)$, $d^h(\alpha^*(f), \epsilon, c, q)$, $h(q, \epsilon|f)$, and $r(q, \epsilon|f)$, which are themselves determined by f through $\alpha^*(f)$, $E[\Delta\overline{Y}_q(f)]$ only depends on the school, course, and year through the administrator's choice of f.[3] Since the objects $E[h(q, \epsilon|f)|d^h = 1]$ and $E[r(q, \epsilon|f)|d^r = 1]$ are means of performance among selected samples of students sorting into each track (partly on the basis of unobserved ability ϵ), they are not objects of interest in their own right, and they do not allow the recovery of the full structural functions $h(q, \epsilon|f)$ $r(q, \epsilon|f)$ without much stronger assumptions on either $h(*)$ and $r(*)$ or $g(\epsilon, c|q)$. However, the above progression makes clear that as long as $g(\epsilon, c|q)$ and W_q are roughly stable for each q across courses and time, identification of the structural functions is unnecessary to solve the administrator's problem.

Essentially, one can simply aggregate over the student-level choice of track, utilizing the fact that every student must choose some track, and compare mean outcomes of students in the same subgroup across schools, cohorts, or courses featuring different administrator choices of f to identify the conditional expectation functions $E[\overline{Y}_q(f)]$ for each subgroup q. Importantly, these functions capture not only the achievement gains or losses from students who have their track choice altered through changes to α^*_f but also how changing f alters the peer effects and level of instruction experienced by other members of the subpopulation.

[3] $E[h(q, \epsilon|f)|d^h = 1]$ and $E[r(q, \epsilon|f)|d^r = 1]$ are defined by:

$$E[h(q, \epsilon|f)|d^h = 1] = \frac{\iint d^h(\alpha^*(f), \epsilon, c, q)h(q, \epsilon|f)g(\epsilon, c|q)d\epsilon dc}{\iint d^h(\alpha^*(f), \epsilon, c, q)g(\epsilon, c|q)d\epsilon dc} \text{ and} \quad \text{(A.6)}$$

$$E[r(q, \epsilon|f)|d^r = 1] = \frac{\iint d^r(\alpha^*(f), \epsilon, c, q)r(q, \epsilon|f)g(\epsilon, c|q)d\epsilon dc}{\iint d^r(\alpha^*(f), \epsilon, c, q)g(\epsilon, c|q)d\epsilon dc} \quad \text{(A.7)}$$

Appendix Tables

Table A1: Estimates of the Parameters $\{\gamma\}$ Governing the Quintile-Specific Treatment Effect Functions of the Honors Enrollment Share $E[\Delta\overline{Y}_q(f)]$ for the Baseline and Alternative Specifications

	(1)	(2)	(3)	(4)
Quintile 5-Linear Coefficient	0.528	0.408	0.742	0.343
	(0.125)	(0.117)	(0.204)	(0.126)
Quintile 4-Linear Coefficient	0.279	0.130	0.503	0.062
	(0.107)	(0.098)	(0.190)	(0.109)
Quintile 3-Linear Coefficient	0.248	0.096	0.479	0.076
	(0.104)	(0.097)	(0.185)	(0.108)
Quintile 2-Linear Coefficient	0.228	0.071	0.458	0.053
	(0.100)	(0.093)	(0.185)	(0.104)
Quintile 1-Linear Coefficient	0.273	0.138	0.526	0.080
	(0.106)	(0.103)	(0.201)	(0.110)
Quintile 5-Squared Coefficient	-1.169	-0.921	-1.708	-0.760
	(0.342)	(0.319)	(0.601)	(0.357)
Quintile 4-Squared Coefficient	-0.556	-0.215	-1.030	-0.029
	(0.299)	(0.269)	(0.578)	(0.309)
Quintile 3-Squared Coefficient	-0.495	-0.141	-1.179	-0.104
	(0.286)	(0.258)	(0.563)	(0.302)
Quintile 2-Squared Coefficient	-0.618	-0.234	-1.231	-0.227
	(0.277)	(0.252)	(0.560)	(0.292)
Quintile 1-Squared Coefficient	-1.011	-0.645	-1.882	-0.541
	(0.309)	(0.285)	(0.623)	(0.311)
Quintile 5-Cubic Coefficient	0.611	0.503	1.007	0.391
	(0.241)	(0.230)	(0.440)	(0.258)
Quintile 4-Cubic Coefficient	0.266	0.087	0.543	-0.043
	(0.214)	(0.199)	(0.428)	(0.229)
Quintile 3-Cubic Coefficient	0.269	0.076	0.747	0.058
	(0.205)	(0.188)	(0.413)	(0.220)
Quintile 2-Cubic Coefficient	0.408	0.188	0.791	0.196
	(0.198)	(0.185)	(0.410)	(0.214)
Quintile 1-Cubic Coefficient	0.769	0.544	1.396	0.483
	(0.233)	(0.213)	(0.462)	(0.233)
Observations	170757	170757	139345	169419
School FEs	NO	YES	NO	YES
Lagged IV	NO	NO	YES	NO
Class Share IV	NO	NO	NO	YES

Notes: Robust standard errors clustered at the school level are in parentheses. 'School FEs' : A full set of school fixed effects is included. 'Lagged IV' uses honors enrollment share two years prior (and its square and cube) as instruments for its contemporary counterparts in the chosen school-year-course. 'Class Share IV': instruments for the current course's honors enrollment share (and its square and cube) using its honors classroom share (and its square and cube).

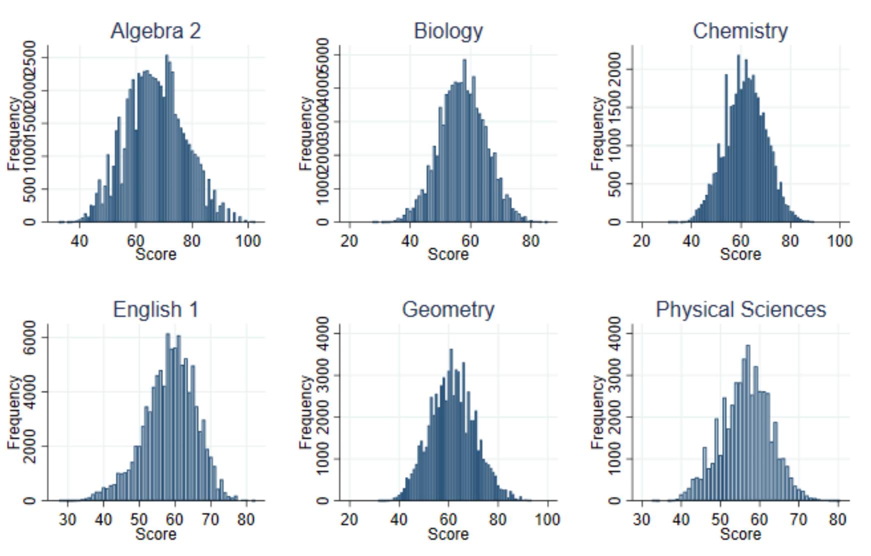
Table A2: Estimates of the Parameters $\{\gamma\}$ Governing the Quintile-Specific Treatment Effect Functions of the Honors Enrollment Share $E[\Delta\overline{Y}_q(f)]$ for Several Specifications Testing Robustness to Functional Form and Endogeneity Assumptions

	(1)	(2)	(3)	(4)	(5)	(6)
Quintile 5-Linear Coefficient	0.454	0.819	0.390	0.341	0.522	0.399
	(0.103)	(0.219)	(0.121)	(0.124)	(0.137)	(0.116)
Quintile 4-Linear Coefficient	0.140	0.342	0.115	0.048	0.160	0.121
	(0.090)	(0.206)	(0.105)	(0.107)	(0.121)	(0.098)
Quintile 3-Linear Coefficient	0.050	0.223	0.062	0.068	0.109	0.092
	(0.089)	(0.192)	(0.105)	(0.107)	(0.112)	(0.096)
Quintile 2-Linear Coefficient	0.037	0.181	0.031	0.067	0.136	0.064
	(0.085)	(0.183)	(0.101)	(0.108)	(0.110)	(0.094)
Quintile 1-Linear Coefficient	0.081	0.329	0.072	0.061	0.185	0.119
	(0.094)	(0.210)	(0.108)	(0.117)	(0.117)	(0.102)
Quintile 5-Squared Coefficient	-1.010	-1.675	-0.895	-0.813	-1.197	-0.882
	(0.279)	(0.607)	(0.334)	(0.343)	(0.365)	(0.315)
Quintile 4-Squared Coefficient	-0.215	-0.368	-0.183	-0.017	-0.325	-0.182
	(0.243)	(0.579)	(0.291)	(0.292)	(0.322)	(0.269)
Quintile 3-Squared Coefficient	-0.000	-0.272	-0.084	-0.093	-0.226	-0.123
	(0.236)	(0.519)	(0.281)	(0.290)	(0.296)	(0.257)
Quintile 2-Squared Coefficient	-0.133	-0.300	-0.166	-0.298	-0.397	-0.205
	(0.224)	(0.508)	(0.276)	(0.288)	(0.296)	(0.252)
Quintile 1-Squared Coefficient	-0.491	-1.046	-0.515	-0.518	-0.824	-0.605
	(0.254)	(0.594)	(0.293)	(0.328)	(0.327)	(0.283)
Quintile 5-Cubic Coefficient	0.556	0.856	0.493	0.461	0.680	0.473
	(0.187)	(0.410)	(0.243)	(0.244)	(0.261)	(0.227)
Quintile 4-Cubic Coefficient	0.075	0.026	0.063	-0.024	0.179	0.062
	(0.164)	(0.392)	(0.217)	(0.208)	(0.234)	(0.199)
Quintile 3-Cubic Coefficient	-0.050	0.049	0.047	0.068	0.150	0.063
	(0.158)	(0.351)	(0.204)	(0.207)	(0.216)	(0.187)
Quintile 2-Cubic Coefficient	0.096	0.119	0.157	0.270	0.298	0.165
	(0.149)	(0.346)	(0.203)	(0.206)	(0.219)	(0.185)
Quintile 1-Cubic Coefficient	0.410	0.717	0.477	0.481	0.712	0.517
	(0.171)	(0.409)	(0.213)	(0.246)	(0.249)	(0.209)
Observations	170757	170773	170863	169419	128005	170757
School FEs	YES	NO	NO	NO	YES	YES
Constrained Coefficients	YES	YES	NO	NO	NO	NO
Other Course IV	NO	YES	NO	NO	NO	NO
School-Year FEs	NO	NO	YES	NO	NO	NO
School-Course FEs	NO	NO	NO	YES	NO	NO
Class Share IV	NO	NO	NO	YES	NO	NO
Restricted School Set	NO	NO	NO	NO	YES	NO
Augmented Controls	NO	NO	NO	NO	NO	YES

Notes: Robust standard errors clustered at the school level are in parentheses. 'Constrained Coefficients': Cubic coefficients are restricted so that the treatment effect function equals zero at an honors enrollment share of 1 as well as 0. 'Other Course IV' uses the contemporaneous honors enrollment share (and its square and cube) in the other tested courses in the same school-year as instruments for the share and its square and cube in the chosen course. 'Class Share IV': instruments for the current course's honors enrollment share (and its square and cube) using its honors classroom share (and its square and cube). 'Restricted School Set': the sample is restricted to schools featuring a distribution of preparedness quintiles such that at most .33 quintile shifts per student are required on average to match the statewide uniform distribution. 'Augmented Controls': includes additional controls for free/reduced price lunch eligibility, sets of indicators for various learning disabilities and teacher education categories, and school-average math scores.

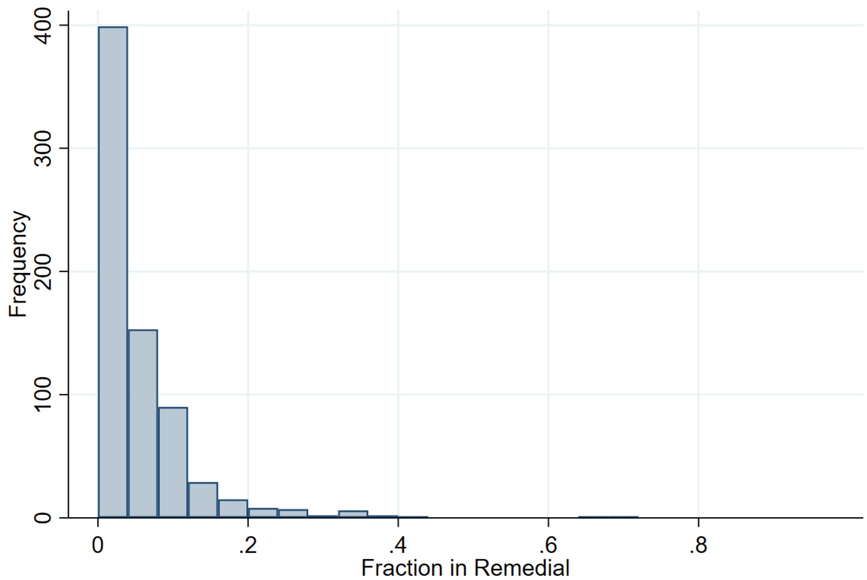
Appendix Figures

Figure A1: Confirming the Absence of Floor and Ceiling Effects - The 2006 Distribution of Pre-Standardized Scale Scores for the Sample Courses



Notes: Each histogram depicts the distribution of pre-standardized student scale scores for a separate course included in the final sample for the year 2006. The histograms confirm the absence of bunching near the ceiling or floor of the test score range. Other years display extremely similar distributions.

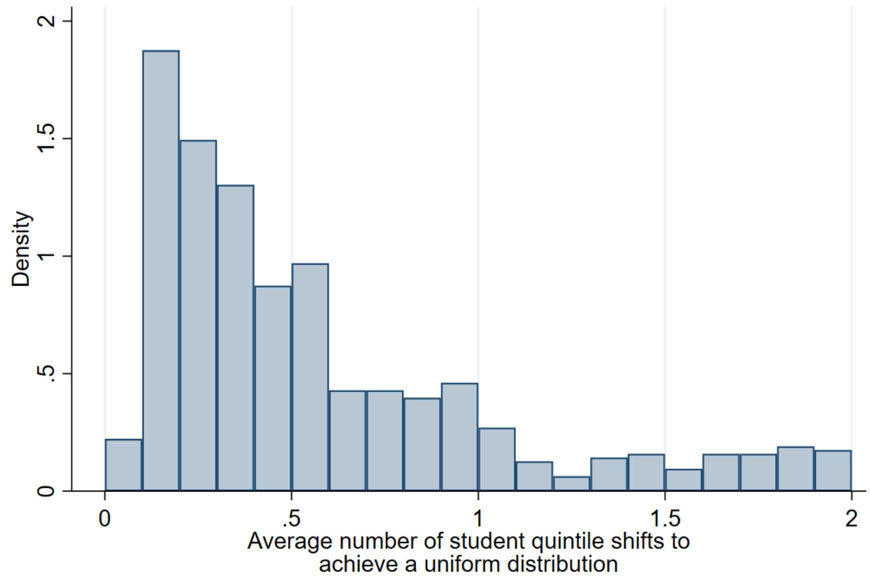
Figure A2: The Distribution of Enrollment Shares for the Remedial Track



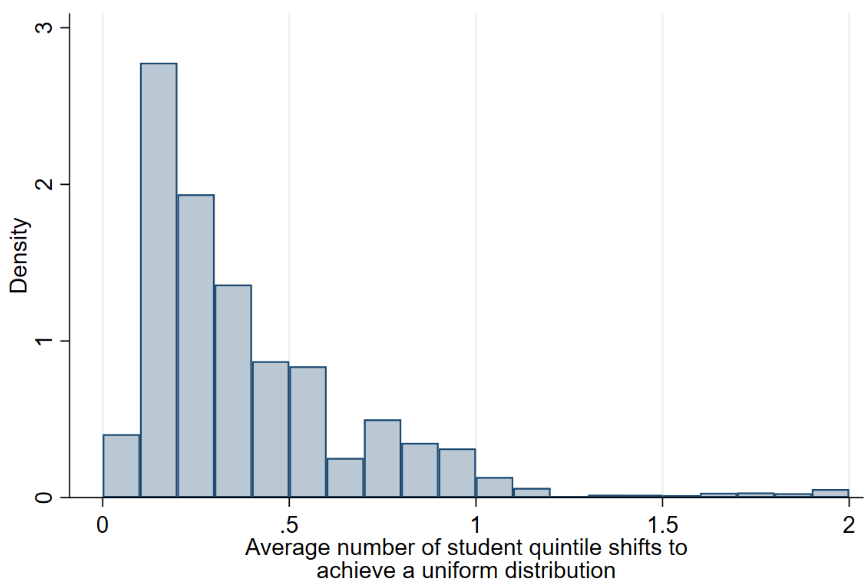
Notes: This figure depicts the fraction of students in the remedial track for school-year-courses from the baseline sample in which a remedial track exists. Fewer than 4% of school-year-courses in the sample contain a remedial track.

Figure A3: Assessing the Validity of Assumption 1 - The Distribution of School-Specific Deviations from the Statewide Composition of Student Predicted Performance

(a) School-Weighted Distribution



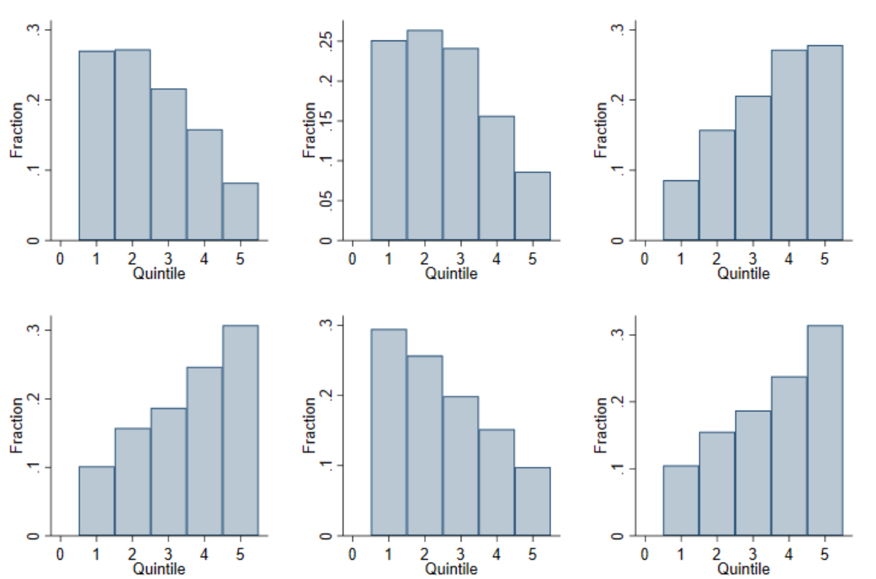
(b) Student-Weighted Distribution



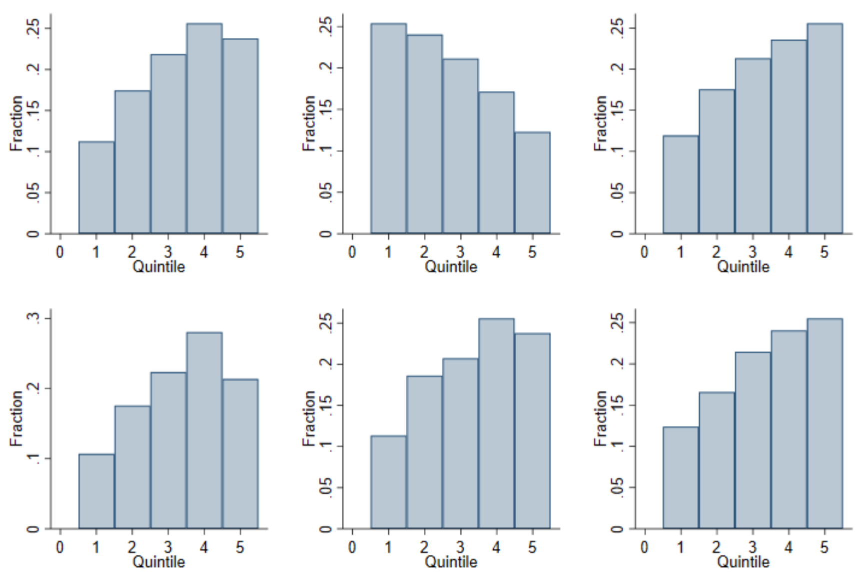
Notes: This figure displays the school-weighted (Panel A) and student-weighted (Panel B) distributions among high schools of the average number of quintiles of an index of predicted test score performance by which the school's students would need to be shifted to match the statewide (uniform) distribution of student predicted performance quintiles. Larger values indicate that the school's student population is more atypical.

Figure A4: The Distribution of Student Predicted Performance Quintiles for the Schools on the Margin of Sample Inclusion

(a) 0.5 Quintile Shifts/Student

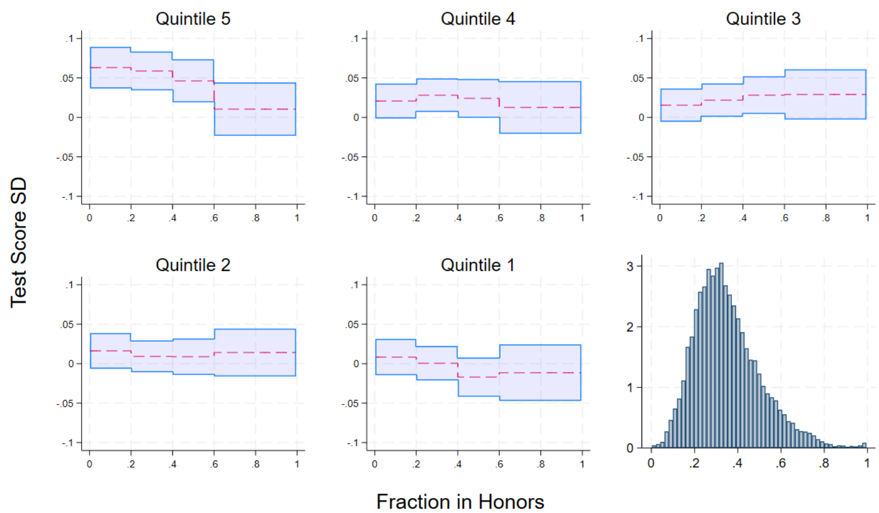


(b) 0.33 Quintile Shifts/Student



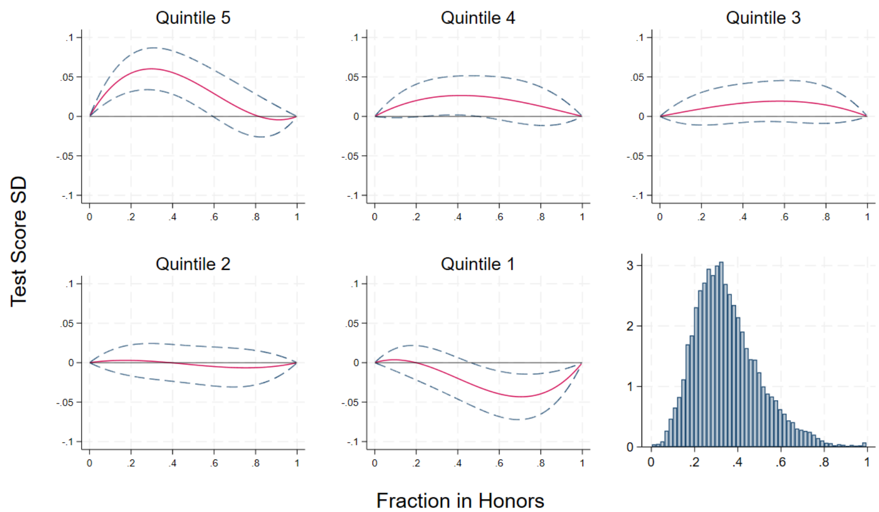
Notes: Figure (a) displays the distribution of students classified by statewide quintile of a regression index of predicted test scores for the six schools with the highest deviations from the statewide (uniform) distribution of quintiles that still qualified for the baseline sample (0.5 required quintile shifts per student on average to reach the uniform distribution). Figure (b) plots the distributions for the six marginal schools when the standard is lowered to one-third quintile shifts per student.

Figure A5: Testing Robustness to Alternative Functional Forms for the Treatment Effect Functions $E[\Delta\overline{Y}_q(f)]$ - Estimating Quintile-Specific Treatment Effects Separately by 20% Interval of Honors Enrollment Share



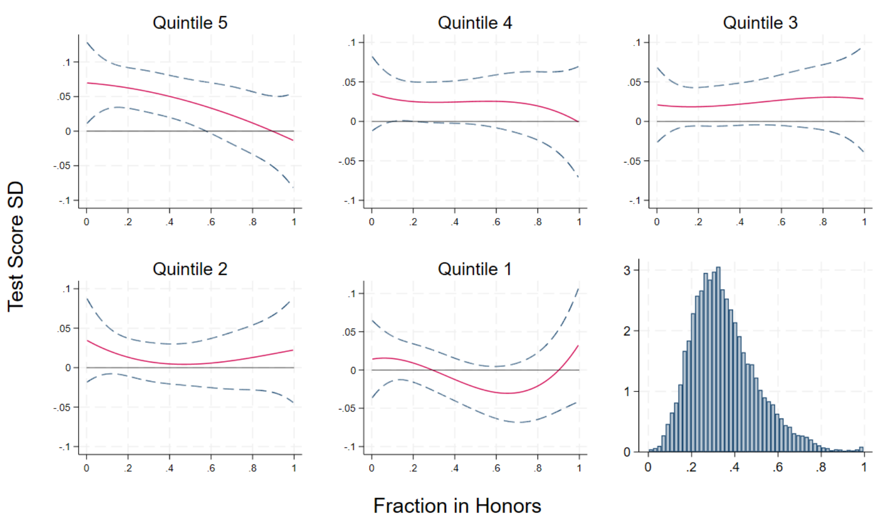
Notes: The first five graphs plot estimates of the function $E[\Delta\overline{Y}_q(f)]$ that maps the coursewide honors enrollment share into expected standardized test performance by quintile of predicted performance for a school FE specification that replaces the baseline cubic polynomial with interactions between indicators for student preparedness quintile and indicators for whether the current course' honors share falls in a particular interval of width 0.2 (with the last two intervals combined due to minimal support). The bottom right graph in each panel displays the density of honors enrollment shares for the chosen sample. 90% pointwise confidence intervals computed using the delta method are displayed with solid lines. The figures rely on the baseline sample of school-year-course-quintile observations (See Section 3.1 for details).

Figure A6: Testing Robustness to Alternative Functional Forms for the Treatment Effect Functions $E[\Delta\overline{Y}_q(f)]$ - Restricted Cubic Specification



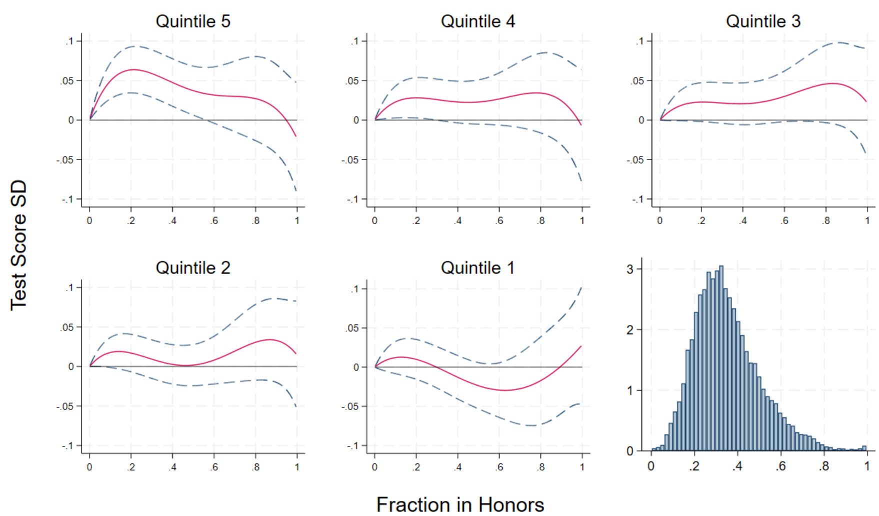
Notes: This figure plots estimates of the function $E[\Delta\overline{Y}_q(f)]$ that maps the coursewide honors enrollment share into expected standardized test performance by quintile of predicted performance for a school FE specification that restricts the value of the treatment effect to be zero at the right end of the unit interval in addition to the left end in order to capture the idea that 100% of students in the honors track also represents an absence of meaningful tracking. 95% pointwise confidence intervals computed using the delta method are displayed with dashes. The bottom right graph in each panel displays the density of honors enrollment shares for the chosen sample. The figures rely on the baseline sample of school-year-course-quintile observations (See Section 3.1 for details).

Figure A7: Testing Robustness to Alternative Functional Forms for the Treatment Effect Functions $E[\Delta\overline{Y}_q(f)]$ - Discontinuity Permitted at a Zero Honors Enrollment Share



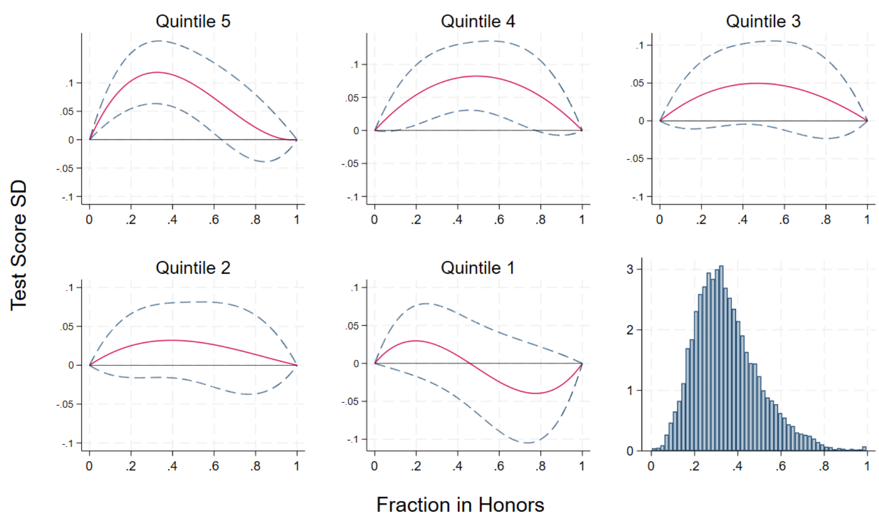
Notes: This figure plots estimates of the function $E[\Delta\overline{Y}_q(f)]$ that maps the coursewide honors enrollment share into expected standardized test performance by quintile of predicted performance for a school FE specification that also includes a separate indicator for whether the course features any tracking. This ensures that predicted values at low enrollment shares are not affected by performance in untracked schools or courses. 95% pointwise confidence intervals computed using the delta method are displayed with dashes. The bottom right graph in each panel displays the density of honors enrollment shares for the chosen sample. The figures rely on the baseline sample of school-year-course-quintile observations (See Section 3.1 for details).

Figure A8: Testing Robustness to Alternative Functional Forms for the Treatment Effect Functions $E[\Delta\overline{Y}_q(f)]$ - Quartic Specification



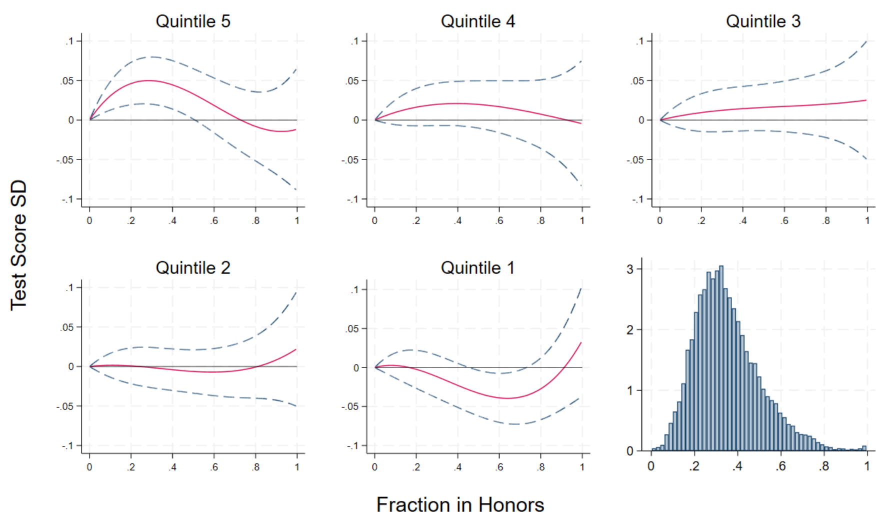
Notes: This figure plots estimates of the function $E[\Delta\overline{Y}_q(f)]$ that maps the coursewide honors enrollment share into expected standardized test performance by quintile of predicted performance for a school FE specification that parameterizes the treatment effect function as a quartic rather than cubic polynomial. 95% pointwise confidence intervals computed using the delta method are displayed with dashes. The bottom right graph in each panel displays the density of honors enrollment shares for the chosen sample. The figures rely on the baseline sample of school-year-course-quintile observations (See Section 3.1 for details).

Figure A9: Testing Robustness to Alternative Functional Forms for the Treatment Effect Functions $E[\Delta\overline{Y}_q(f)]$ - Using the Mean Share of Honors Classrooms in Other Courses in the Same School-Year as an Instrument for the Honors Enrollment Share



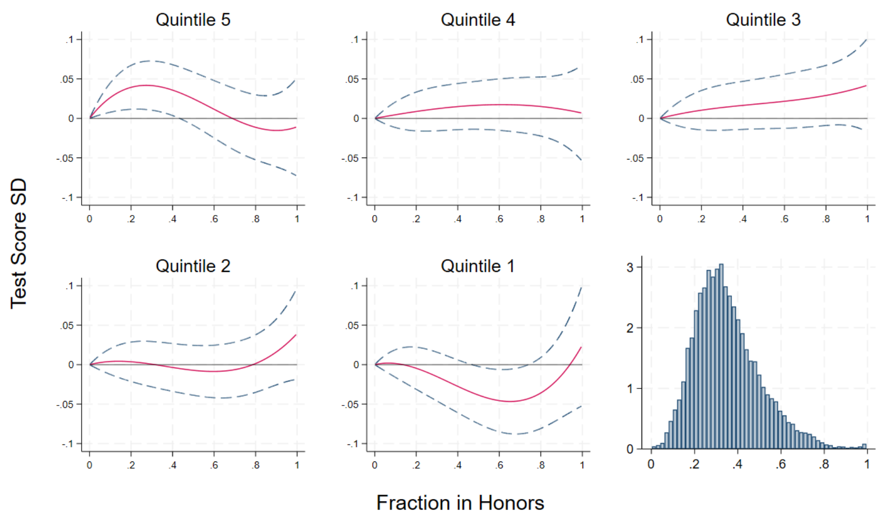
Notes: This figure plots estimates of the function $E[\Delta\overline{Y}_q(f)]$ that maps the coursewide honors enrollment share into expected standardized test performance by quintile of predicted performance for a specification in which the current course's honors enrollment share is instrumented with the mean share among other courses in the same school-year combination. 95% pointwise confidence intervals computed using the delta method are displayed with dashes. The bottom right graph in each panel displays the density of honors enrollment shares for the chosen sample. The figures rely on the baseline sample of school-year-course-quintile observations (See Section 3.1 for details).

Figure A10: Testing Robustness to Alternative Functional Forms for the Treatment Effect Functions $E[\Delta\overline{Y}_q(f)]$ - School-Year Fixed Effects Specification



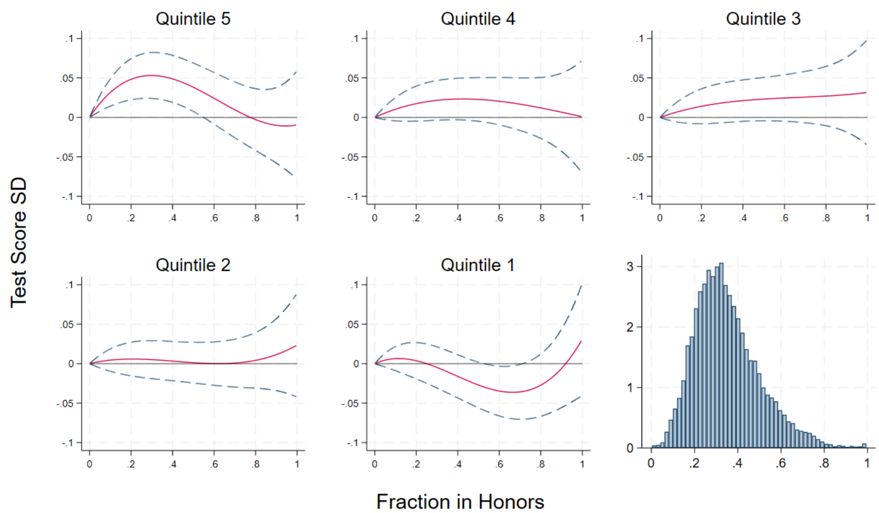
Notes: This figure plots estimates of the function $E[\Delta\overline{Y}_q(f)]$ that maps the coursewide honors enrollment share into expected standardized test performance by quintile of predicted performance for a specification that augments the baseline specification by including a set of school-year fixed effects. 95% pointwise confidence intervals computed using the delta method are displayed with dashes. The bottom right graph in each panel displays the density of honors enrollment shares for the chosen sample. The figures rely on the baseline sample of school-year-course-quintile observations (See Section 3.1 for details).

Figure A11: Testing Robustness to Alternative Functional Forms for the Treatment Effect Functions $E[\Delta\overline{Y}_q(f)]$ - Combining the Class Share IV with School-Course Fixed Effects



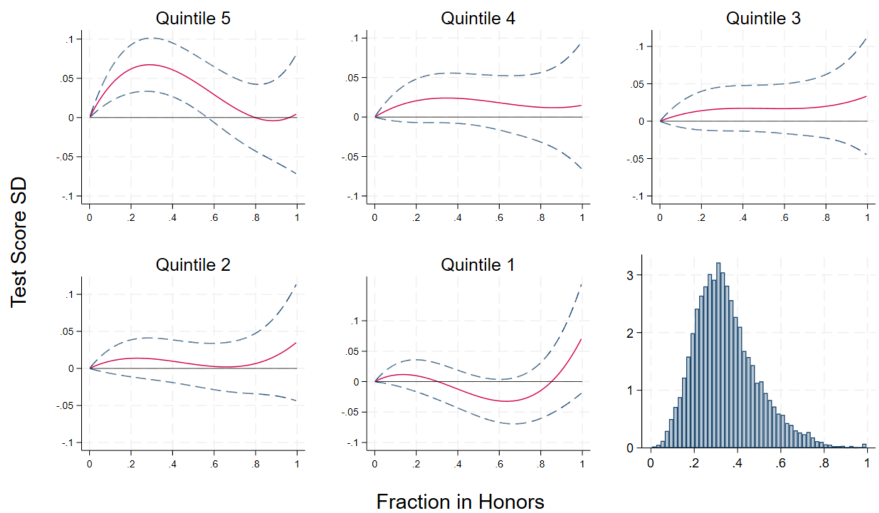
Notes: This figure plots estimates of the function $E[\Delta\overline{Y}_q(f)]$ that maps the coursewide honors enrollment share into expected standardized test performance by quintile of predicted performance for a specification in which the current course's share of enrollment in the honors track (and its square and cube) are instrumented with the course's share of honors classrooms (and its square and cube) and a full set of school-course fixed effects are included. 95% pointwise confidence intervals computed using the delta method are displayed with dashes. The bottom right graph in each panel displays the density of honors enrollment shares for the chosen sample. The figures rely on the baseline sample of school-year-course-quintile observations (See Section 3.1 for details).

Figure A12: Testing Robustness of the Treatment Effect Functions $E[\Delta\overline{Y}_q(f)]$ to Additional Controls - Low Income and Learning Disability Indicators, and Teacher Education Category Shares



Notes: This figure plots estimates of the function $E[\Delta\overline{Y}_q(f)]$ that maps the coursewide honors enrollment share into expected standardized test performance by quintile of predicted performance for a school fixed-effects specification that uses an augmented set of controls that includes a common but coarse administrative indicator for low parental income, indicators for various forms of learning disabilities, and shares of the teachers in the chosen school-course year who received bachelor's, master's, professional, and PhD degrees. 95% pointwise confidence intervals computed using the delta method are displayed with dashes. The bottom right graph in each panel displays the density of honors enrollment shares for the chosen sample. The figures rely on the baseline sample of school-year-course-quintile observations (See Section 3.1 for details).

Figure A13: Testing Robustness to Violations of Assumption 1 - Specification Featuring a Restricted Sample of Schools with More Typical Distributions of Predicted Student Performance Based on Middle School Performance



Notes: This figure plots estimates of the function $E[\Delta\overline{Y}_q(f)]$ that maps the coursewide honors enrollment share into expected standardized test performance by quintile of predicted performance for the school fixed-effects specification but using an alternative sample that restricts the set of schools to those where the average student would need to shift their quintile of the preparedness index by less than 1/3 in order for the school to match the statewide (uniform) distribution of quintiles. 95% pointwise confidence intervals computed using the delta method are displayed with dashes. The bottom right graph in each panel displays the density of honors enrollment shares for the chosen sample.