Bright lines, big effects: Unintended consequences of student proficiency thresholds

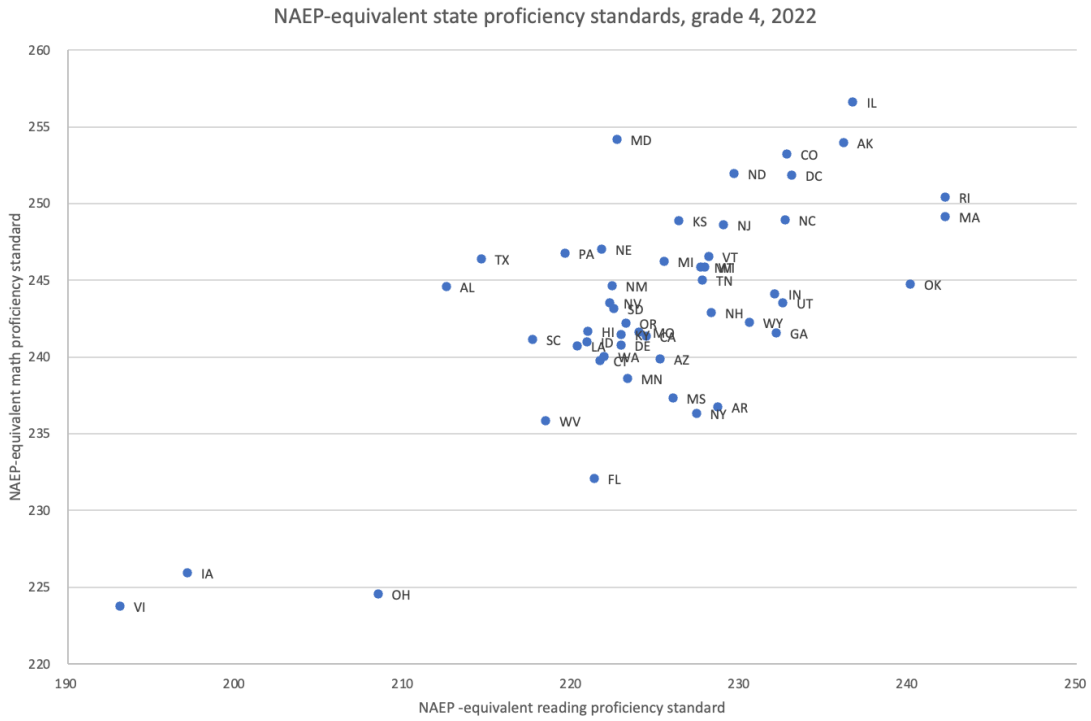
David Figlio, University of Rochester and NBER
(visiting Hoover Institution, Stanford University)
Umut Özek, RAND Corporation

February 2025

Cutpoints are everywhere

- Many sectors use discrete classifications of continuous variables to allocate resources, determine services, assign to treatment
  - Birth weight, date of birth, family income, test scores, etc., etc., etc.
  - There must be 10,000+ papers using RD
- The use of proficiency cutpoints in testing
  - Every state in the USA uses proficiency cutpoints in standardized testing
  - Where states draw the line varies dramatically – see, e.g., study of NAEP-equivalent measures of state proficiency standards in 2022 (more on this soon)
  - States/districts use proficiency cutpoints for grade retention, remediation, course assignment, school accountability, and so on
  - Lots of arguments about whether standards are "too low" or "too high"
- Where we draw the line is highly consequential – but while we know a lot about consequences of cutpoints for people on either side of the line, relative to one another, we don't know the effect of drawing a line in a specific place.
  - Are kids below the line "overtreated" and those above the line "undertreated", say?

Cutpoints vary dramatically across states



What this paper is about

- This paper is not about the treatment effects that people typically study – but rather about the use/location of the bright lines themselves
- We exploit a situation in which there was a year in which (still valid) test scores weren't available for assessment due to an implementation glitch – where the researcher sees the test score that WOULD have been used had educators had access to it
- This allows us to observe the counterfactual – what would have happened to kids with precisely the same test scores in testing regimes where information was known to schools vs. unknown to schools
- Three states of the world:
  - Test scores observed, state mandate to use cutpoints for high-stakes purposes
  - Test scores NOT observed (but the researcher sees them ex post)
  - Test scores observed, no state mandate for cutpoint usage by educators

What this paper is about

- Three states of the world:
  - Test scores observed, state mandate to use cutpoints for high-stakes purposes
  - Test scores NOT observed (but the researcher sees them ex post)
  - Test scores observed, no state mandate for cutpoint usage by educators
- (At least) two potential effects of the "bright lines":
  - Differential treatments below vs. above threshold – potential for "overtreatment" below cutpoint and "undertreatment" above cutpoint (if mandate → treatment for those below the threshold)
  - Signal sent to educators/students/families leading to focus on kids at margin
- This unusual natural experiment allows us to directly study effects of "bright lines" on the treatment of students and proximate and long-run student outcomes both
  - When "bright lines" are high-stakes; and
  - When "bright lines" are low-stakes

The policy environment – before 2015

- Florida has administered statewide standardized tests since 1998
  - Florida Comprehensive Assessment Test (FCAT) aligned to Sunshine State Standards: 1998 - 2011
  - Florida Comprehensive Assessment Test 2.0 (FCAT 2.0) aligned to Next Generation Sunshine State Standards: 2011 (transition year) - 2014
  - Florida Standards Assessments (FSAs) aligned to Florida Standards: 2015 - 2022
- FCAT and FCAT 2.0 were high-stakes for students – schools were required to remediate students (including grade repetition in some situations) scoring below proficiency

The policy environment – high-stakes policy

- Middle/high school remediation in Florida, enacted 2004, expanded 2006
- Requires students who score below proficiency in ELA to take a remedial ELA course in the subsequent year, in addition to the regular course
- Provides additional instructional resources
  - Additional instructional time in remediation subject
  - Requires remedial course teachers to have Reading Endorsement or Certification in Reading, and evidence of success (as determined by district)
  - Requires remedial courses to have adequate infrastructure (class size, materials)
- Some exceptions (e.g., if student has greater needs in math)
  - 24 percent of high school students in our setting took a remedial ELA course

The policy environment – 2015 and later

- Florida dropped the "high-stakes" requirement effective spring 2015
- The transition from FCAT 2.0 to the FSAs was bumpy
  - The roll-out of the FSAs in 2015, delivered by a third-party contractor, was very glitchy
    - Middle and high schoolers took the new FSA assessments online
    - But there were rampant log-in issues, including delays and error messages, which dramatically delayed roll-out of the new assessments
    - The new FSAs were field tested in Utah, a much smaller environment than Florida
  - The issue was not the test itself: An independent study commissioned by the Florida Department of Education determined that the FSAs were valid
  - As a consequence, test administration and scoring were sufficiently delayed that the outcomes of the assessments were not known to educators and school districts until it was too late to use them for the 2015-16 academic year
- FSA administration in 2016 and thereafter was smooth
- Proficiency standards didn't change: NAEP equivalent proficiency in 8th grade reading was 262 in 2013, 261 in 2015, 264 in 2017

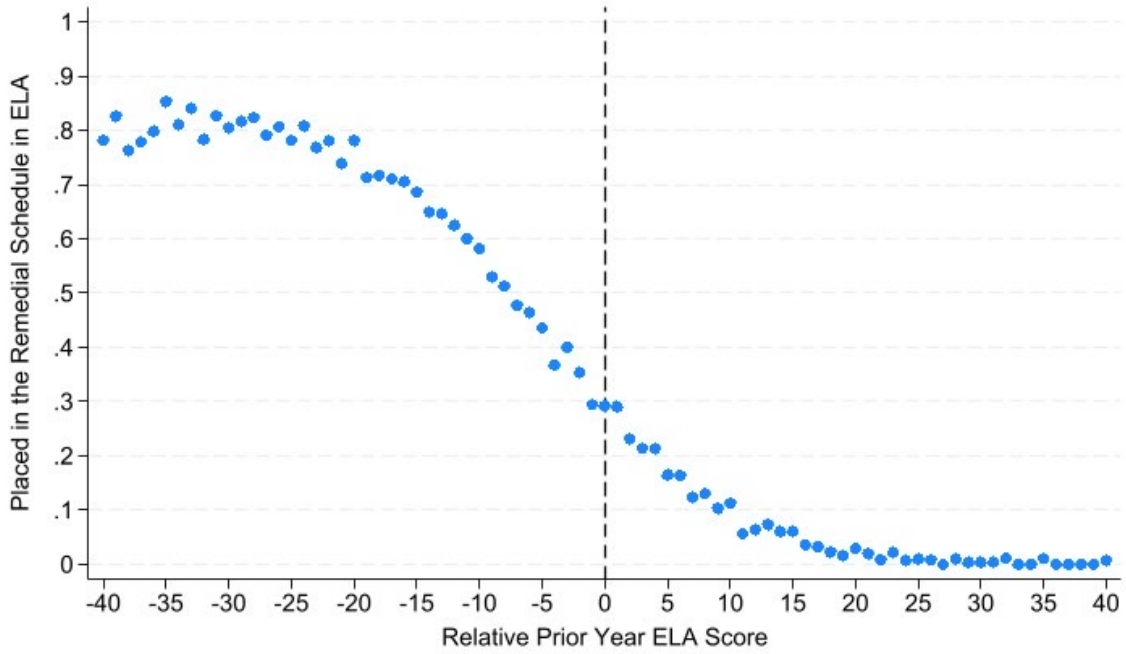
Setting and data

- Large Urban School District (LUSD) – one of 10 largest in the USA
  - More than 200 schools; over 200,000 students
  - Diverse student body: Over 30% Hispanic, over 25% Black, over 10% foreign born, over 30% non-native speaker of English, over 60% receiving subsidized meals
- Longitudinal student-level records from 2005-06 through 2021-22 including test scores, disciplinary incidents, attendance, demographic information (e.g., race, gender, ethnicity, country of birth, language spoken at home, student age, schools attended), detailed course enrollment records/teacher data, matched to National Student Clearinghouse (NSC) data (postsecondary enrollment, college selectivity, completion) for graduates from 2013-2023 until July 2024

Population for today's analysis

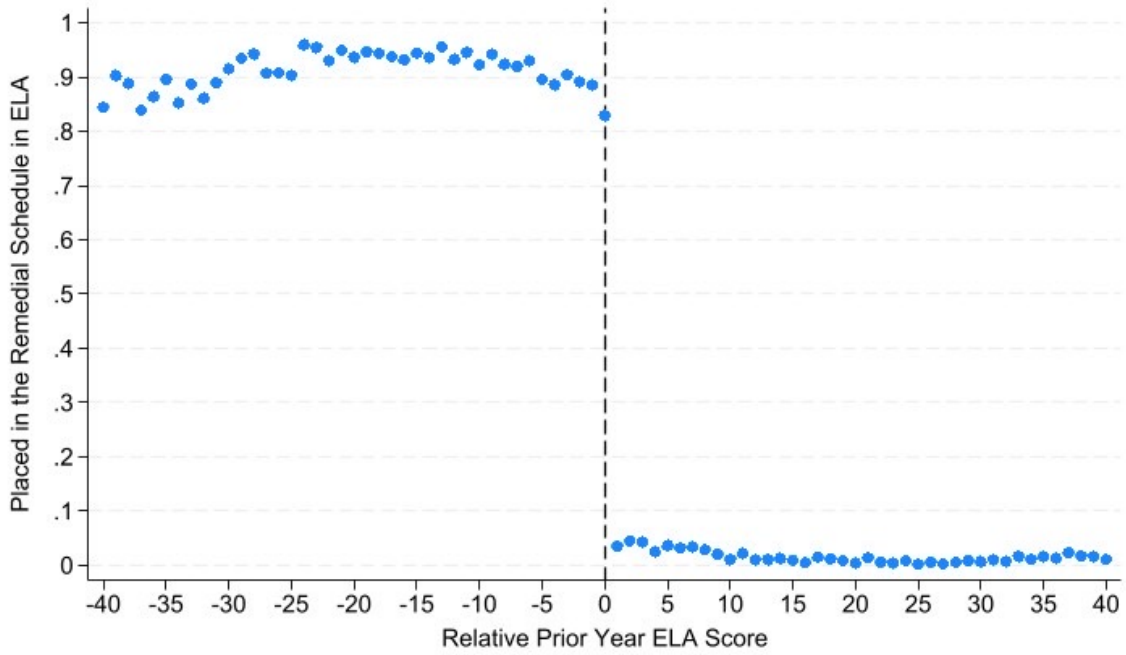
- Students in grades 8-10 between 2012-13 and 2017-18 (exam takers in 2012-2017)
  - Young enough to take either FCAT 2.0 or FSAs but old enough to have been observable in the NSC records
- Six years of data:
  - Three "high-stakes" regime years: tests administered 2012, 2013, 2014
  - One "testing pause" regime year: tests administered 2015
  - Two "low-stakes" regime years: tests administered 2016, 2017

Likelihood of being placed in remedial ELA schedule, 2015-16 (testing pause year)

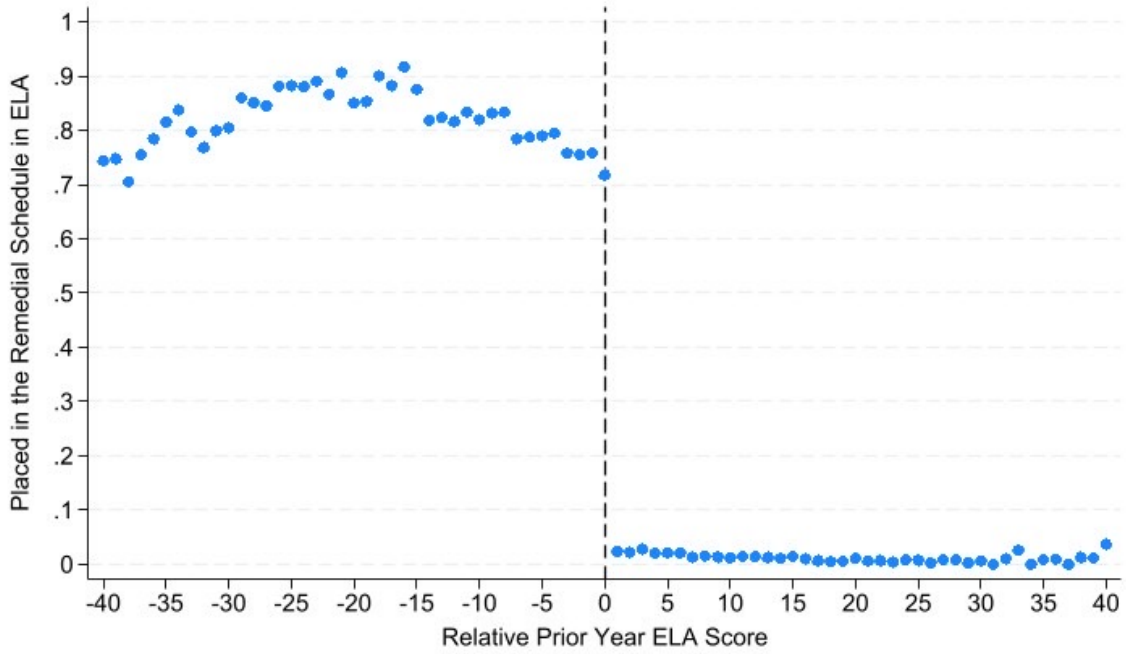


"Placed in remedial schedule" = took remedial ELA course in addition to regular ELA course

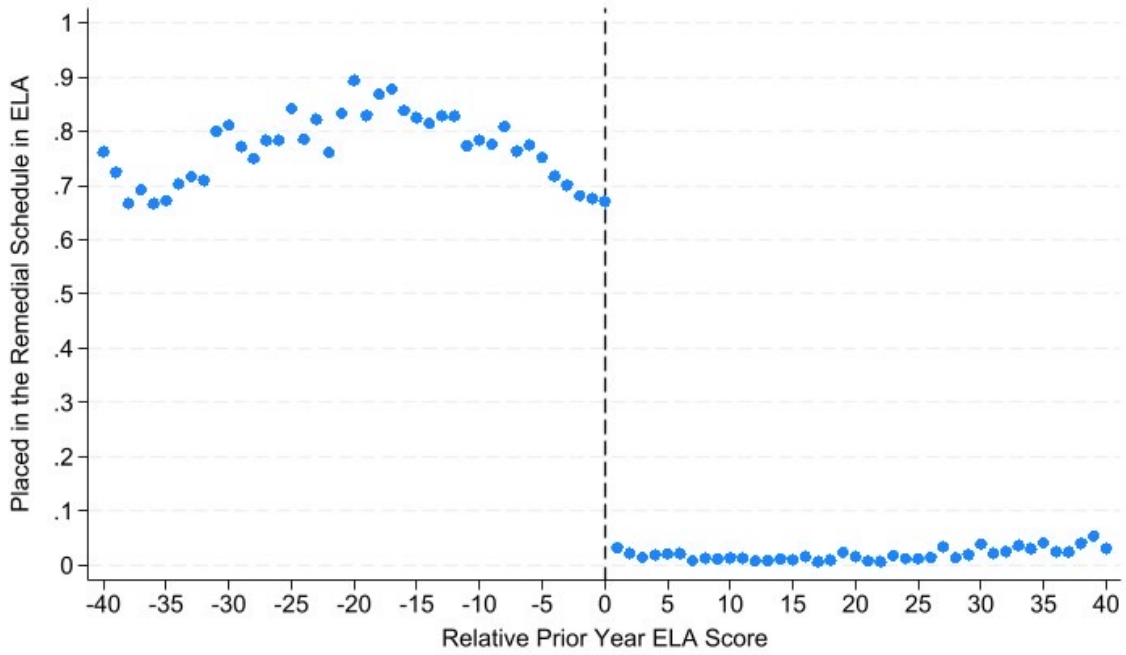
Likelihood of being placed in remedial ELA schedule, 2014-15 (last high stakes year)



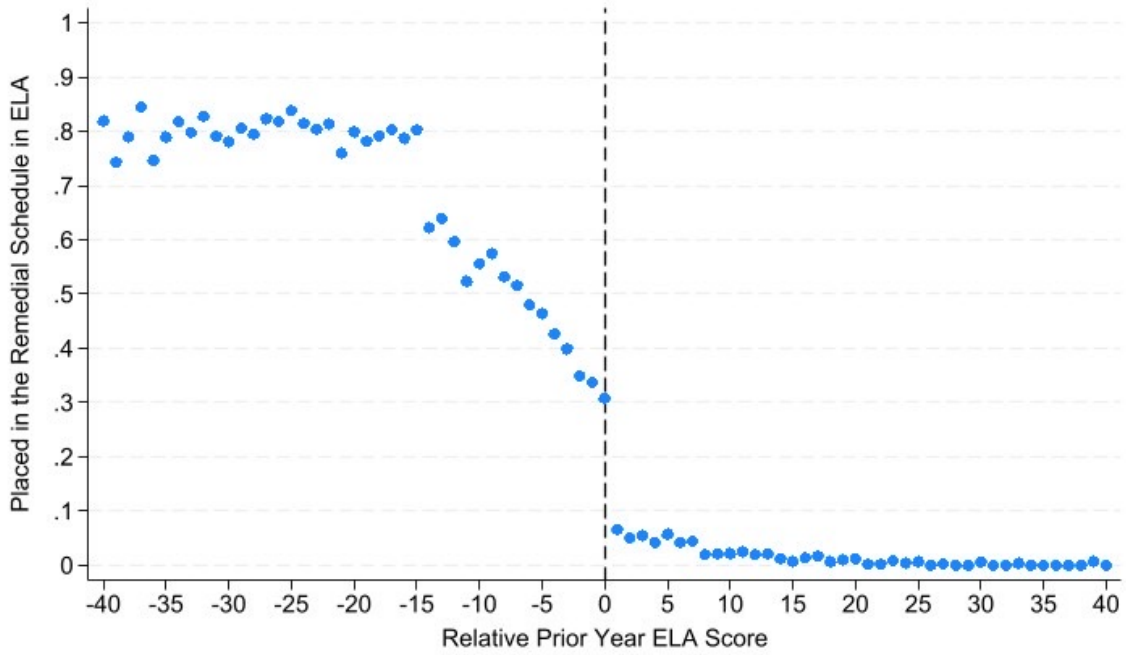
Likelihood of being placed in remedial ELA schedule, 2013-14 (second to last high stakes year)



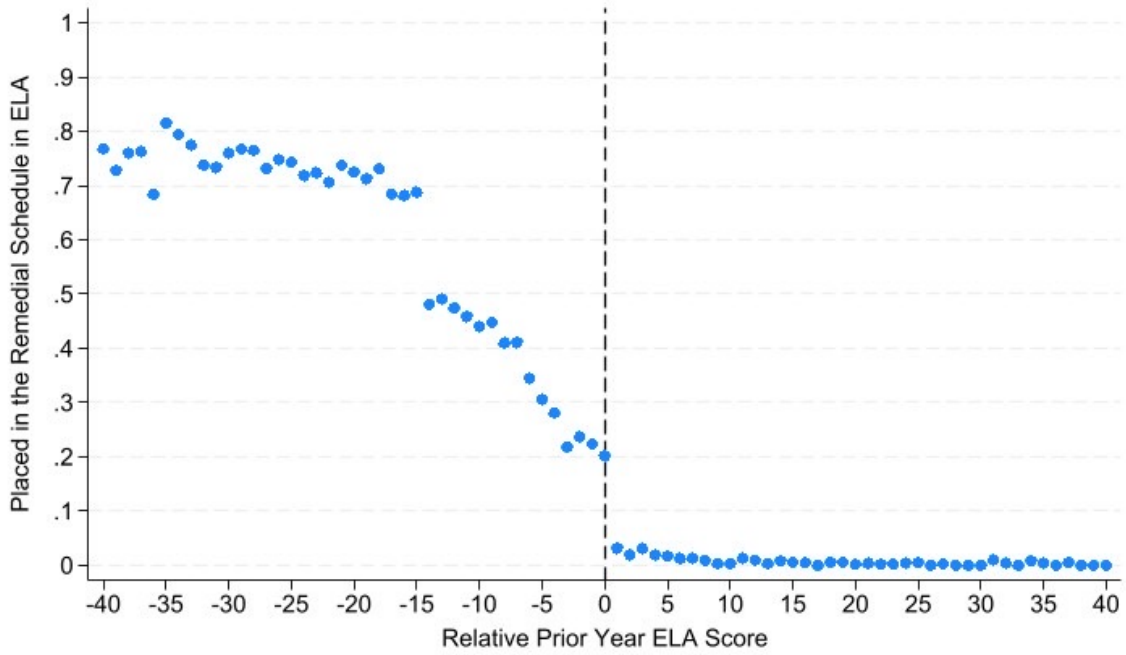
Likelihood of being placed in remedial ELA schedule, 2012-13 (third to last high stakes year)



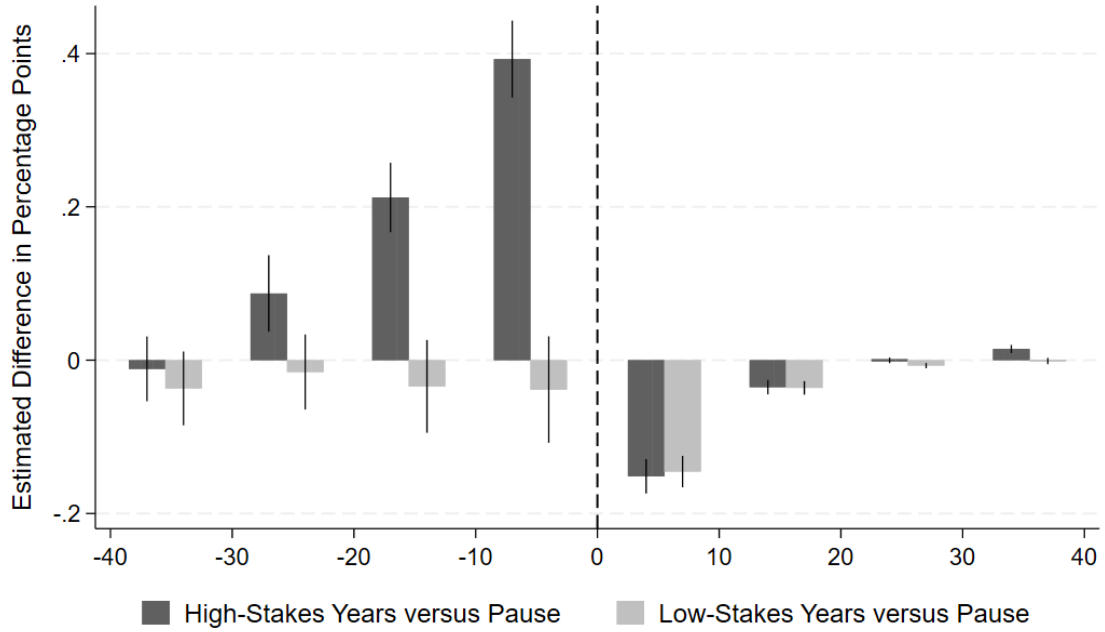
Likelihood of being placed in remedial ELA schedule, 2016-17 (first low-stakes year)



Likelihood of being placed in remedial ELA schedule, 2017-18 (second low-stakes year)

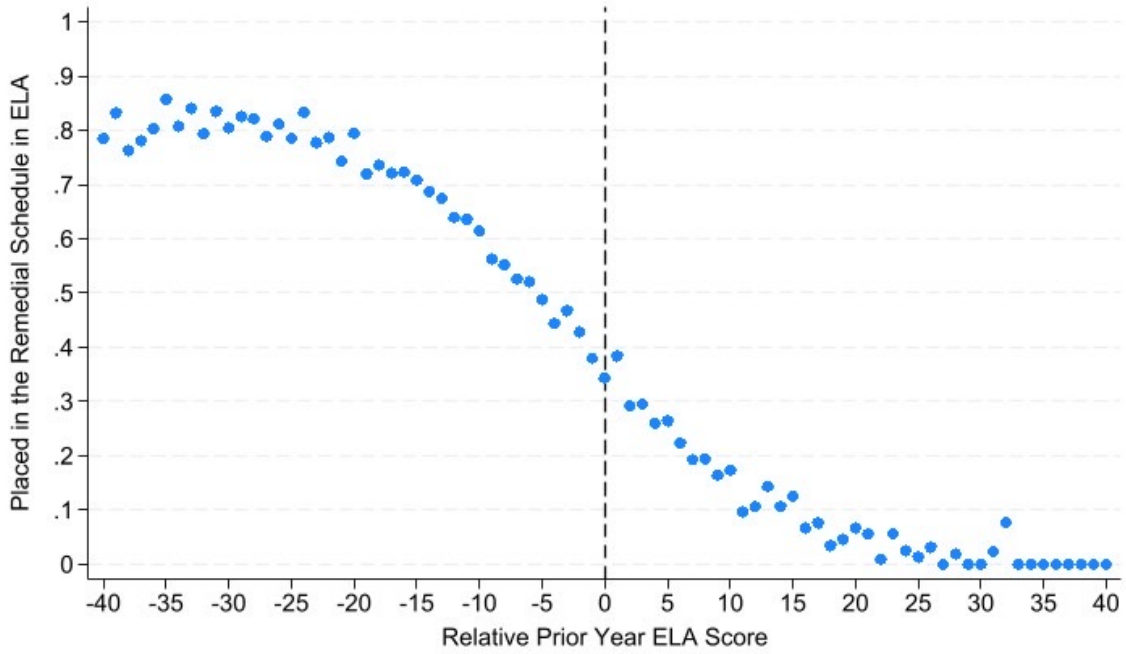


Students most affected by the shifts tend to be right around the cutpoint

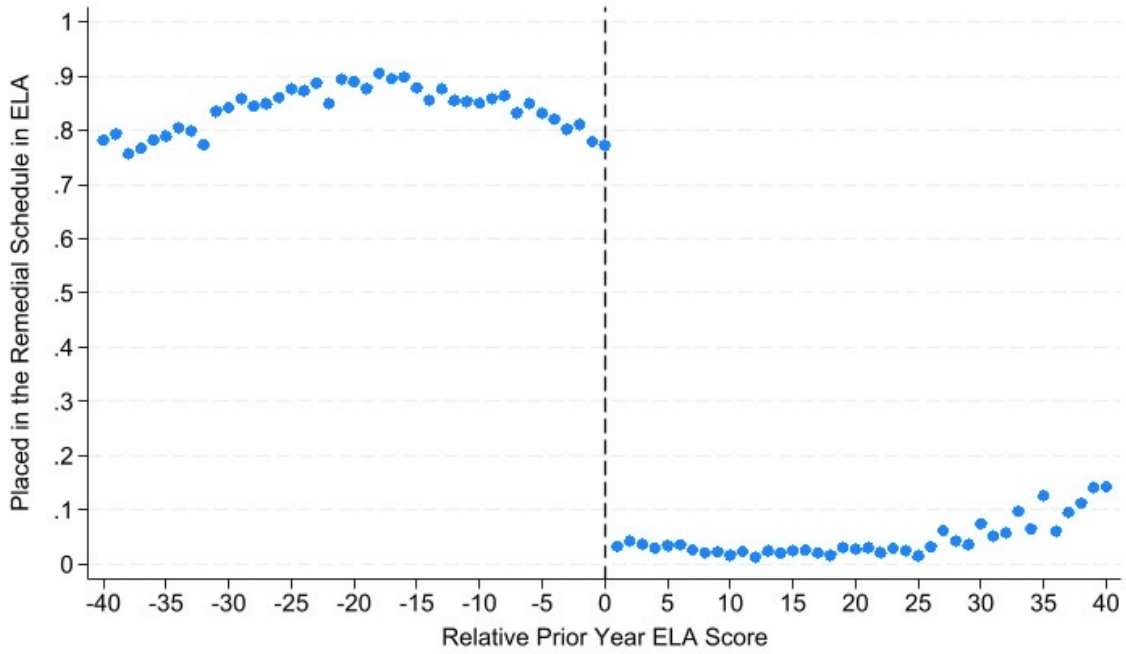


Of course, students away from the cutoff can be affected by regime changes due to classroom composition, etc.

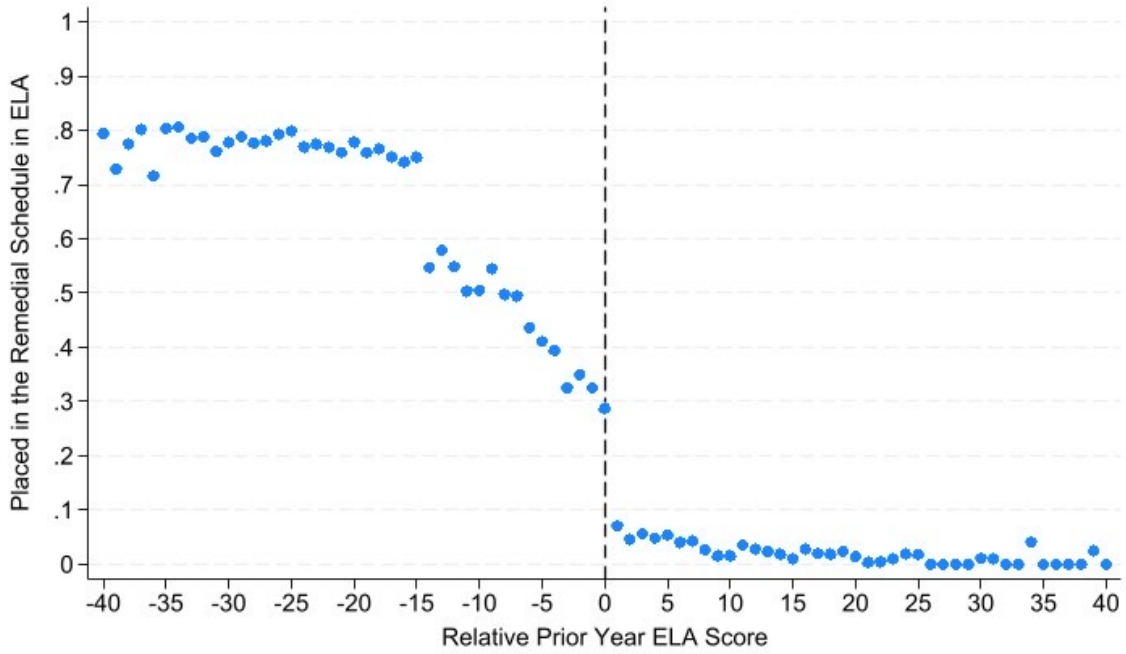
Likelihood of being placed in remedial ELA schedule, 2015-16 (testing pause year) – historically low-performers in math



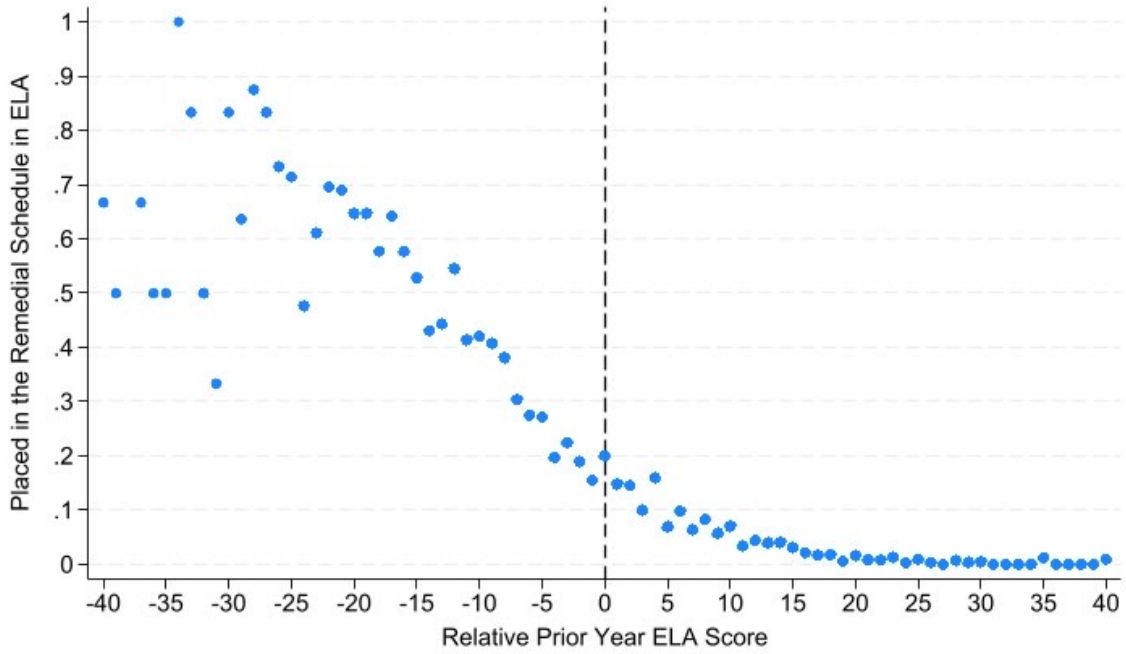
Likelihood of being placed in remedial ELA schedule, high-stakes years – historically low-performers in math



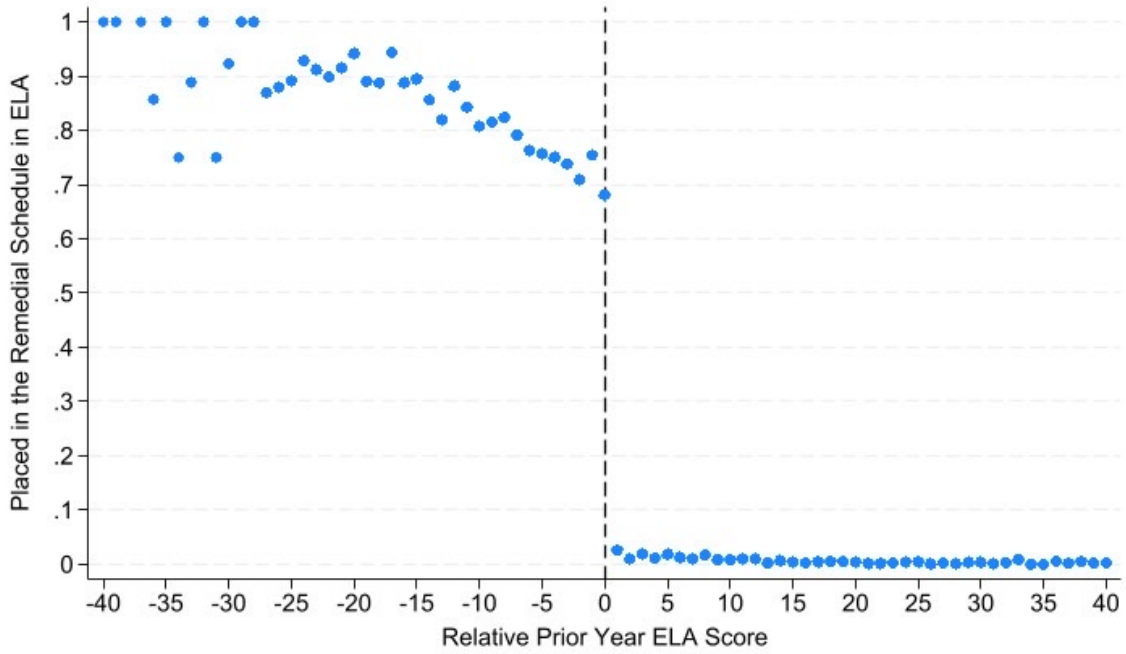
Likelihood of being placed in remedial ELA schedule, low-stakes years – historically low-performers in math



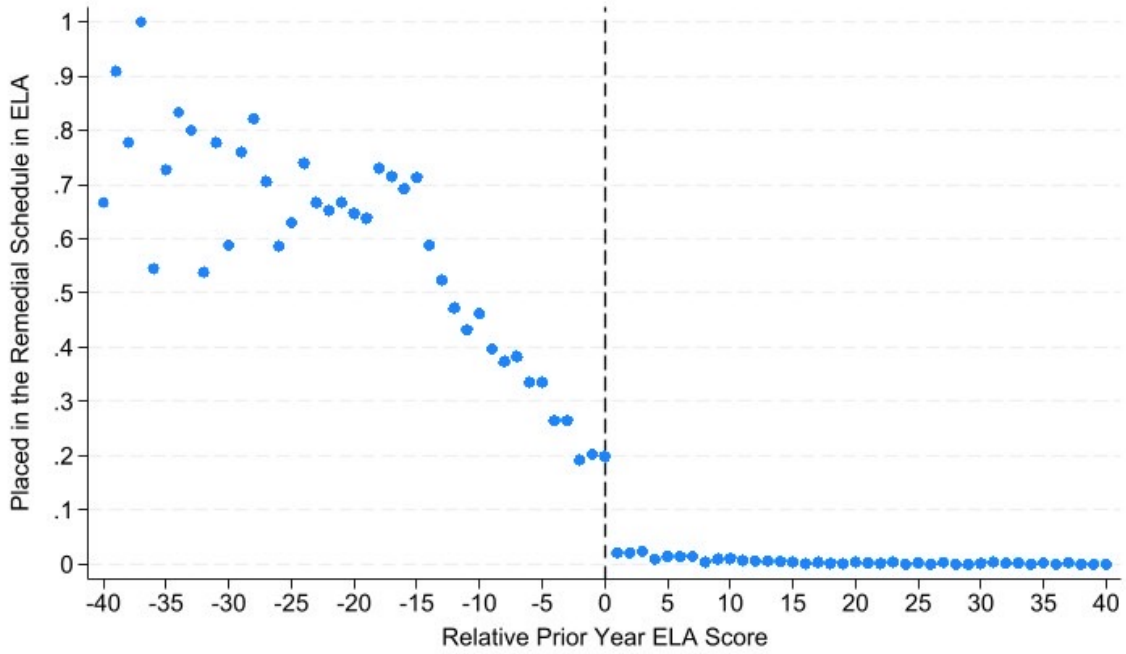
Likelihood of being placed in remedial ELA schedule, 2015-16 (testing pause year) – historically high-performers in math



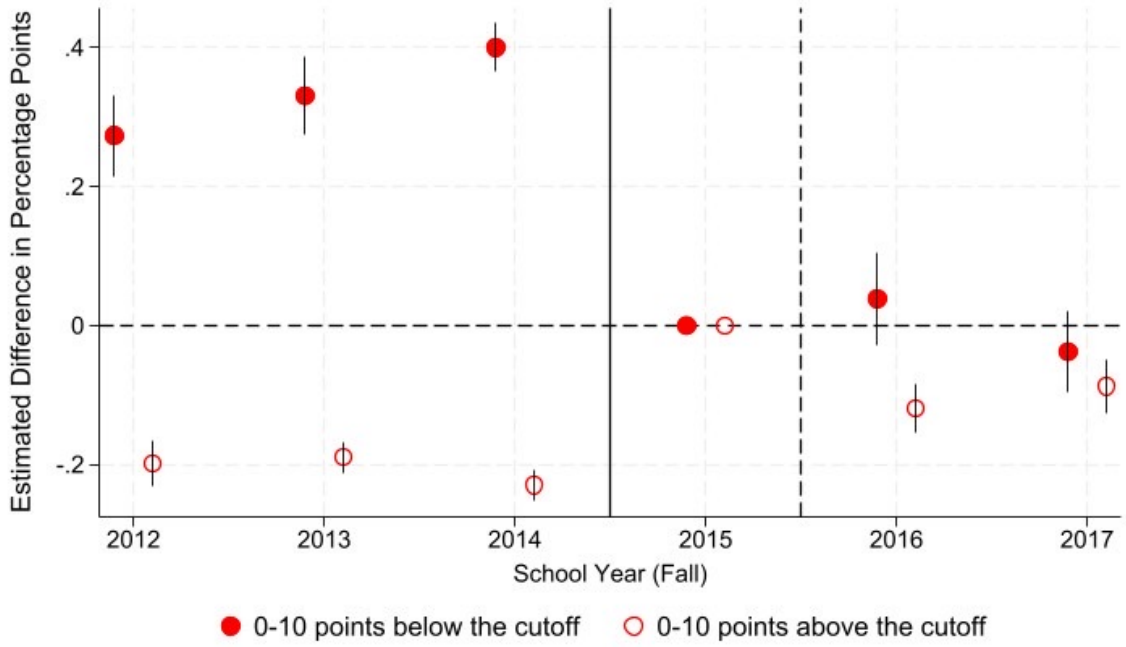
Likelihood of being placed in remedial ELA schedule, high-stakes years – historically high-performers in math



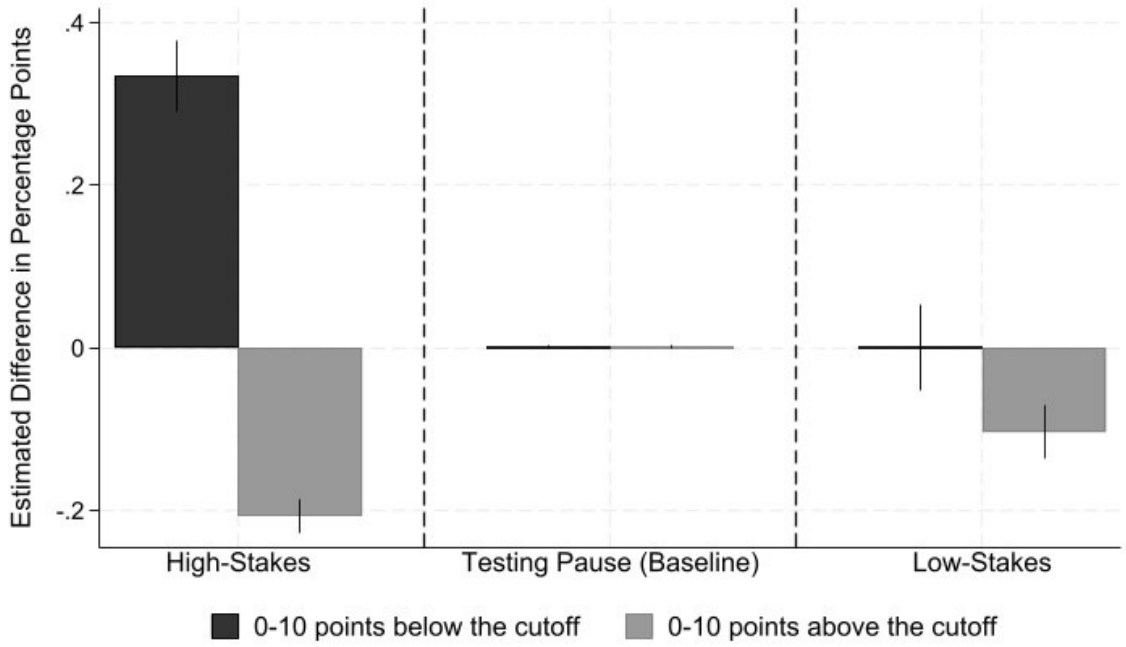
Likelihood of being placed in remedial ELA schedule, low-stakes years – historically high-performers in math



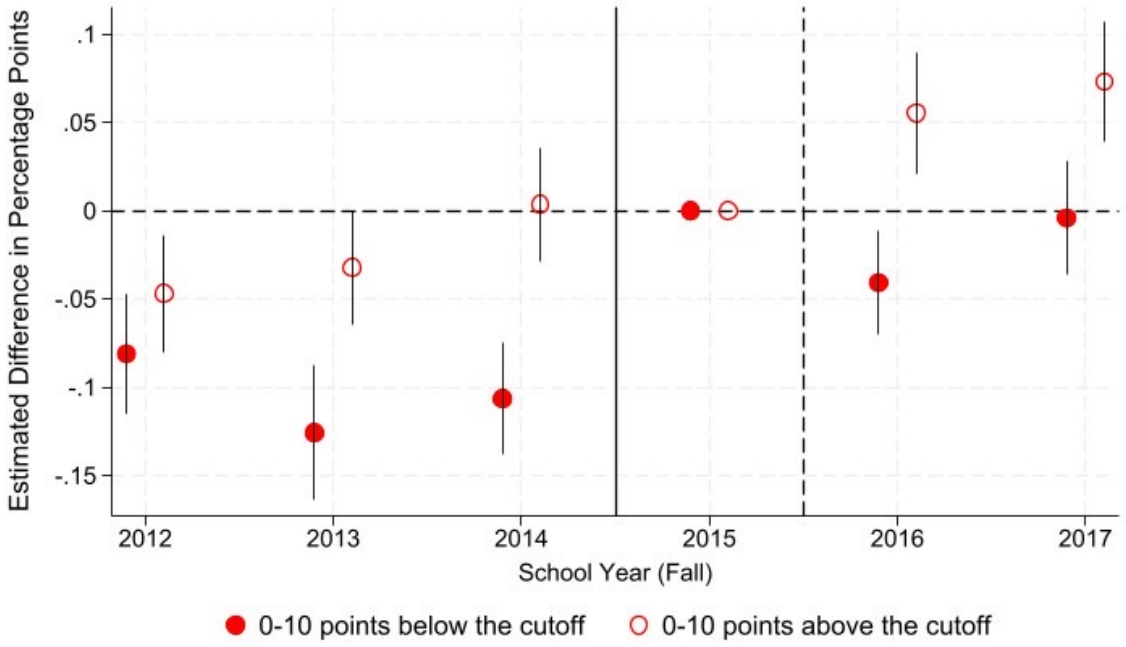
Event study estimates – remedial ELA class



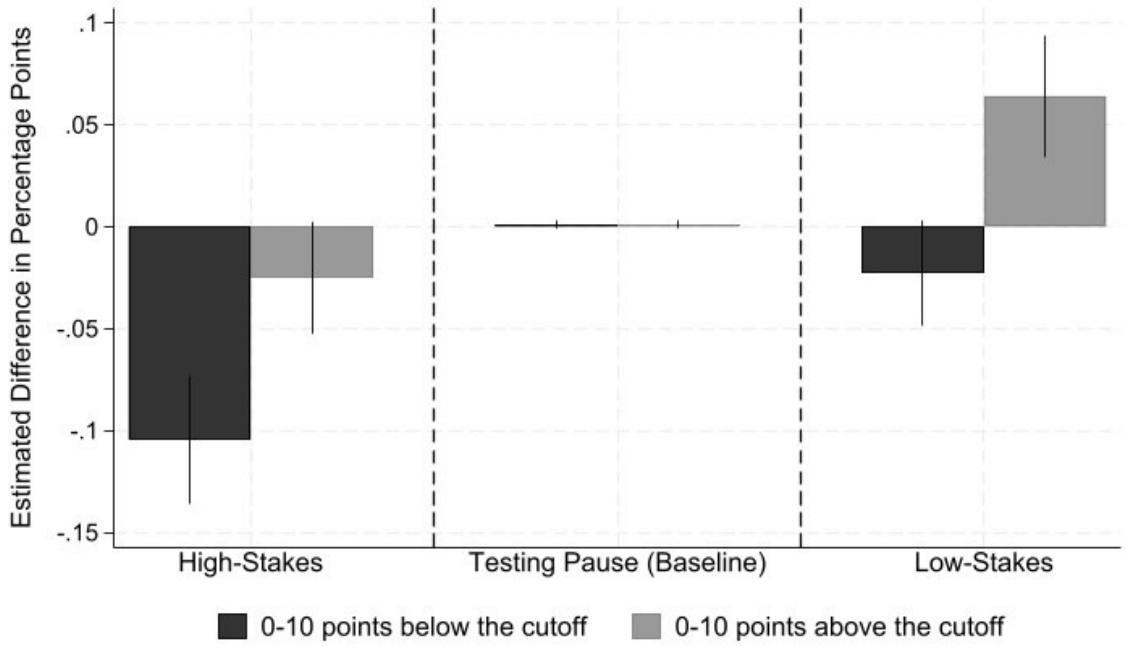
Dif-in-dif estimates – remedial ELA class



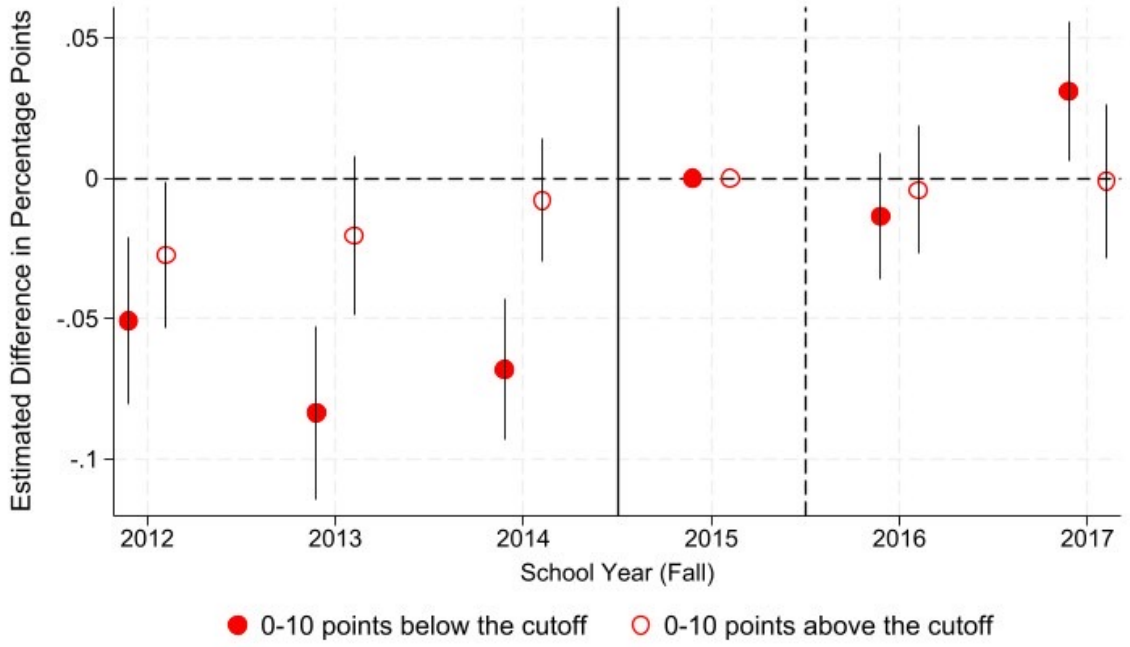
Event study estimates – advanced ELA class



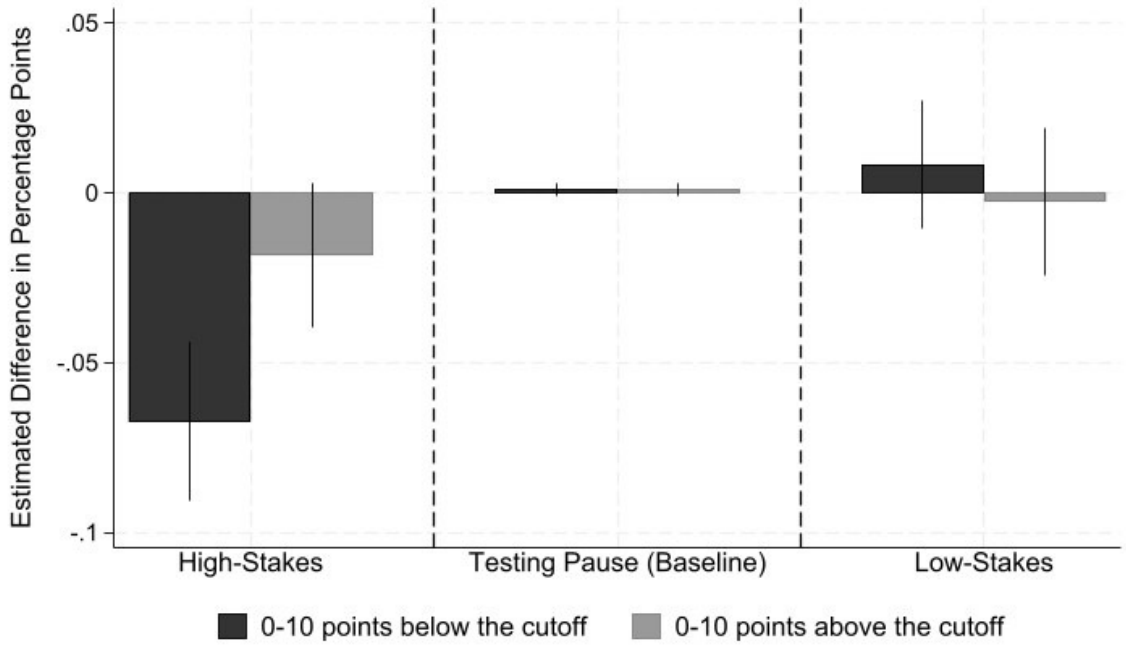
Dif-in-dif estimates – advanced ELA class



Event study estimates – advanced non-ELA class



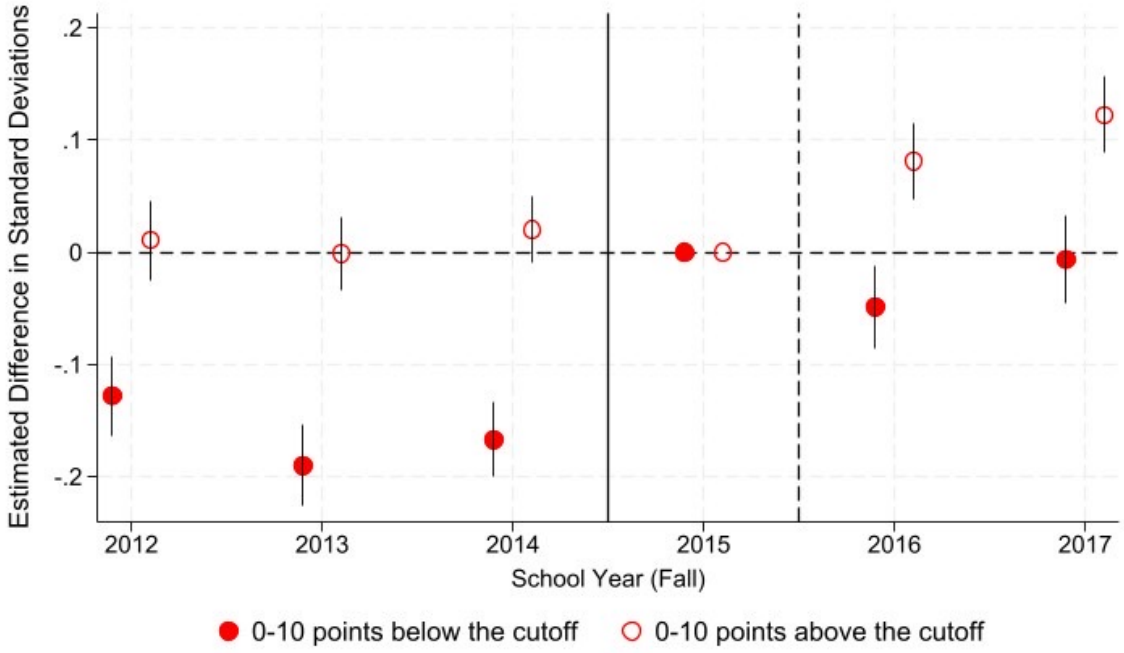
Dif-in-dif estimates – advanced non-ELA class



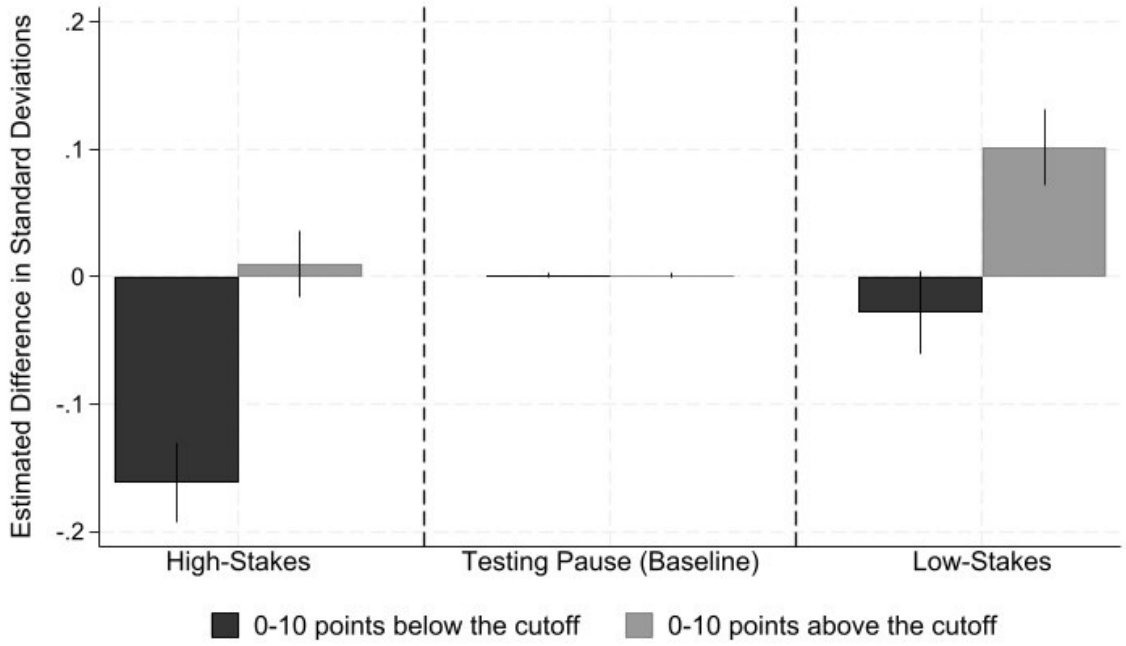
Results so far in the presentation

- High-stakes regime: Apparent over-treatment of kids just below the line and under-treatment of kids just above the line
- Low-stakes regime: Apparent under-treatment of kids just above the line continued
- In the high-stakes regime, both kids just below and just above the line took fewer advanced non-ELA classes. Evidence of a signaling effect?

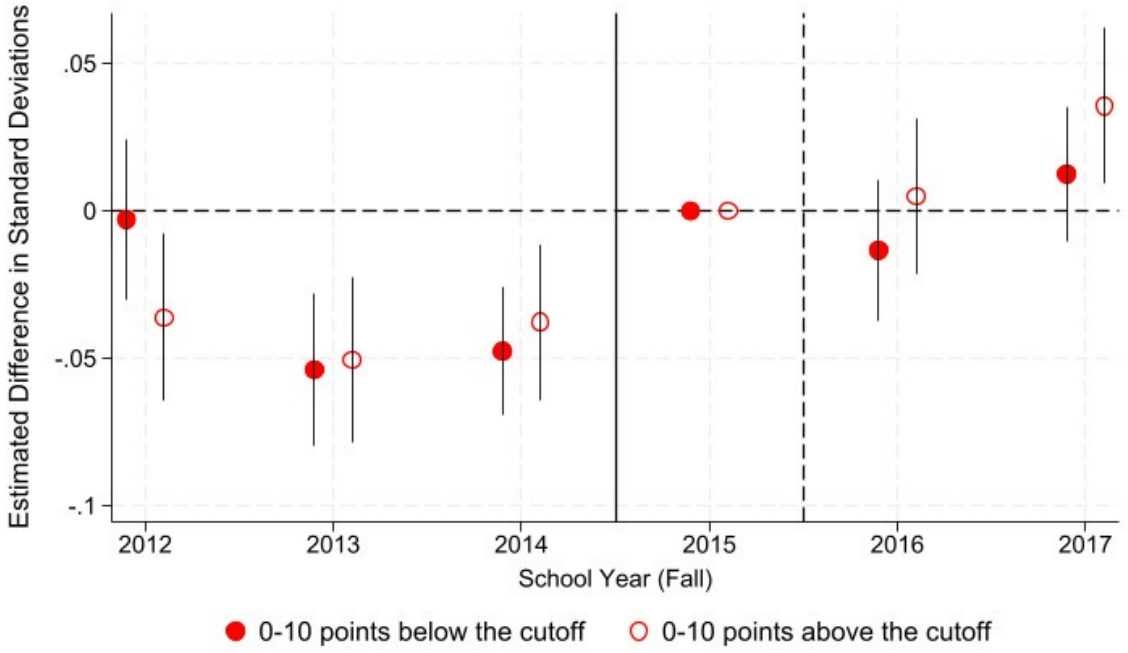
Consequences of bright lines for classroom peers (ELA classes)



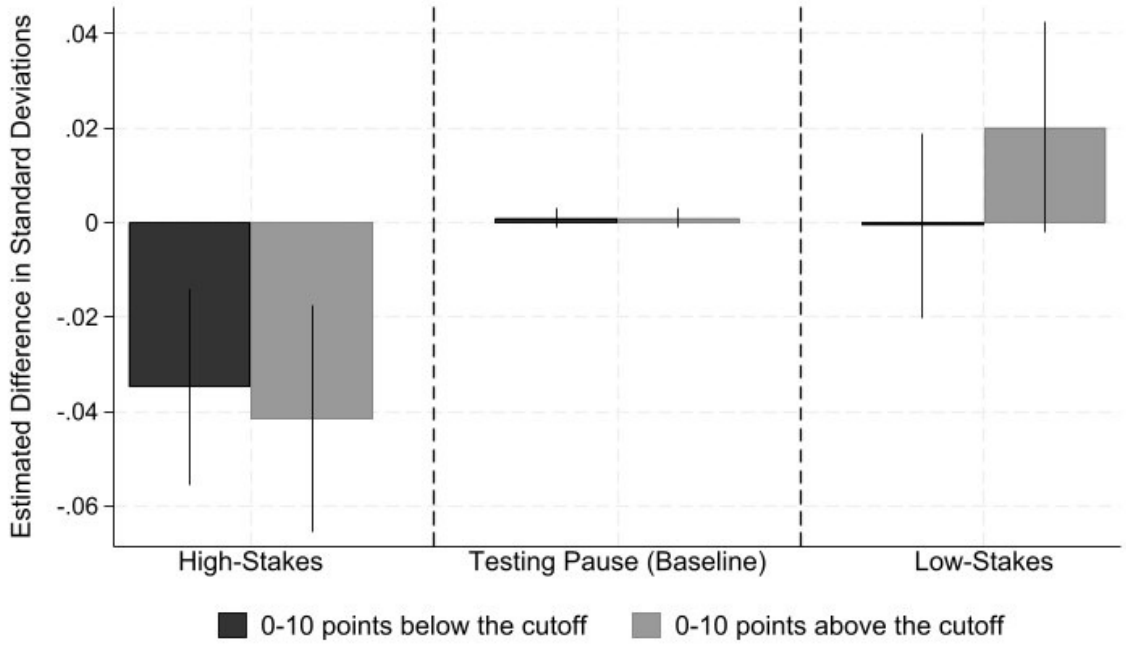
Consequences of bright lines for classroom peers (ELA classes)



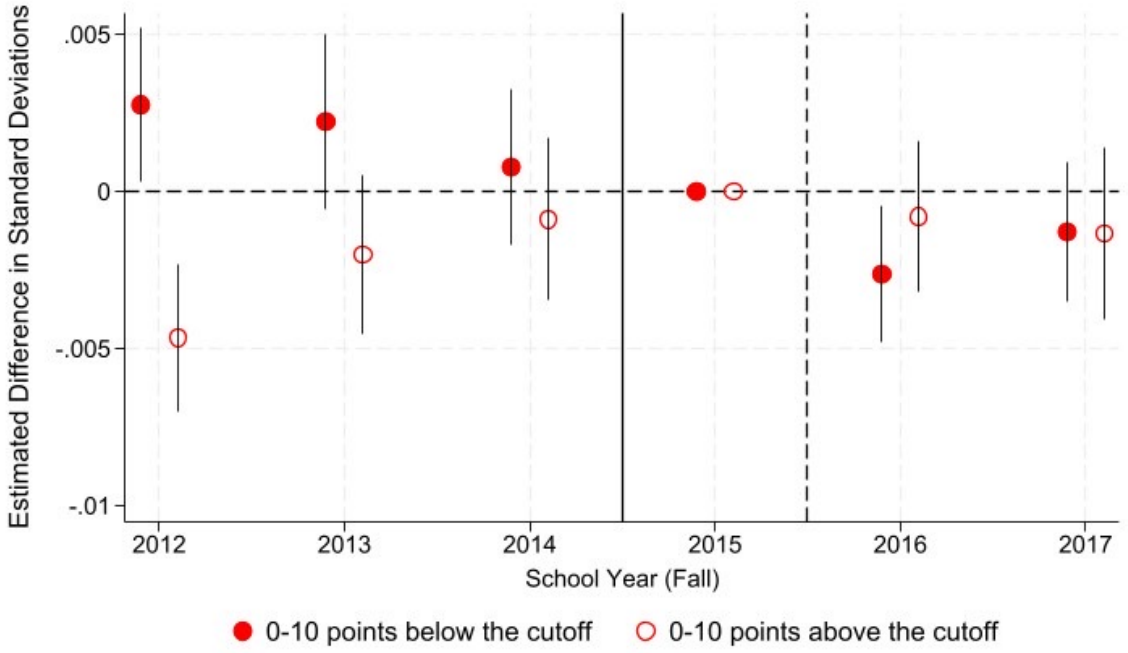
Consequences of bright lines for classroom peers (non-ELA classes)



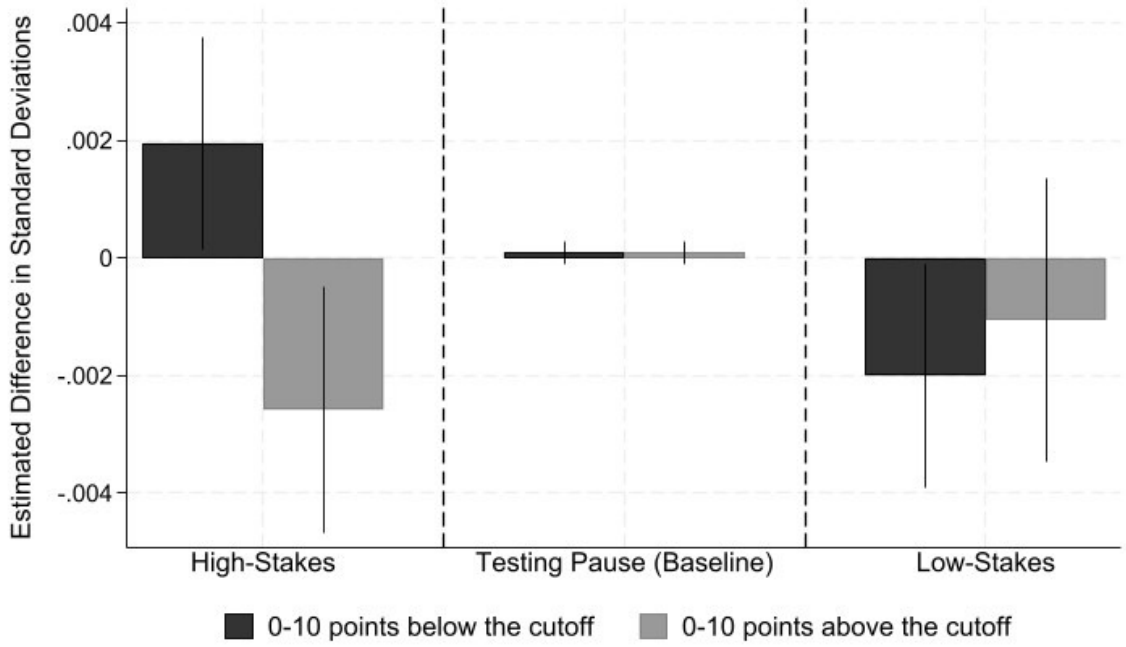
Consequences of bright lines for classroom peers (non-ELA classes)



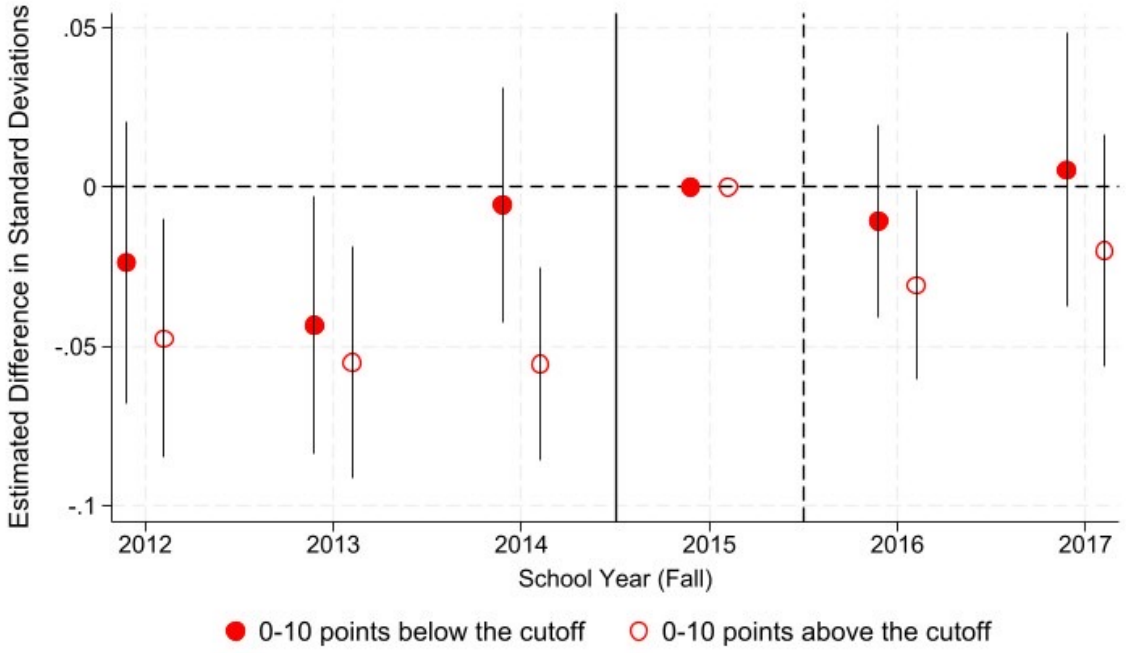
Consequences of bright lines for measured teacher value added in ELA classes



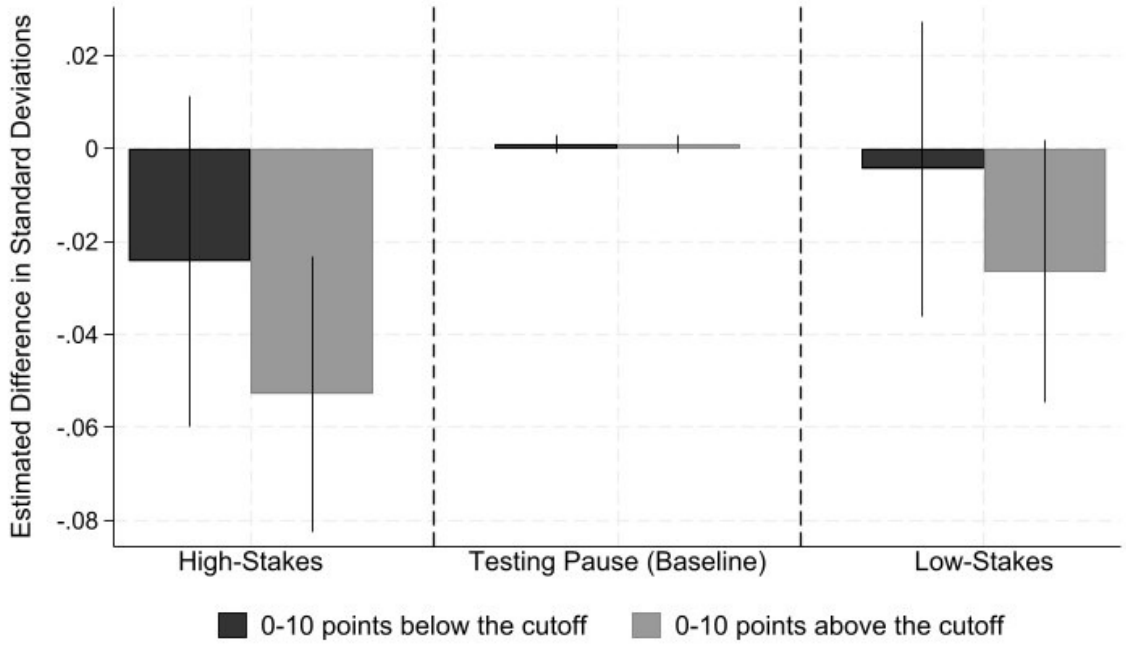
Consequences of bright lines for measured teacher value added in ELA classes



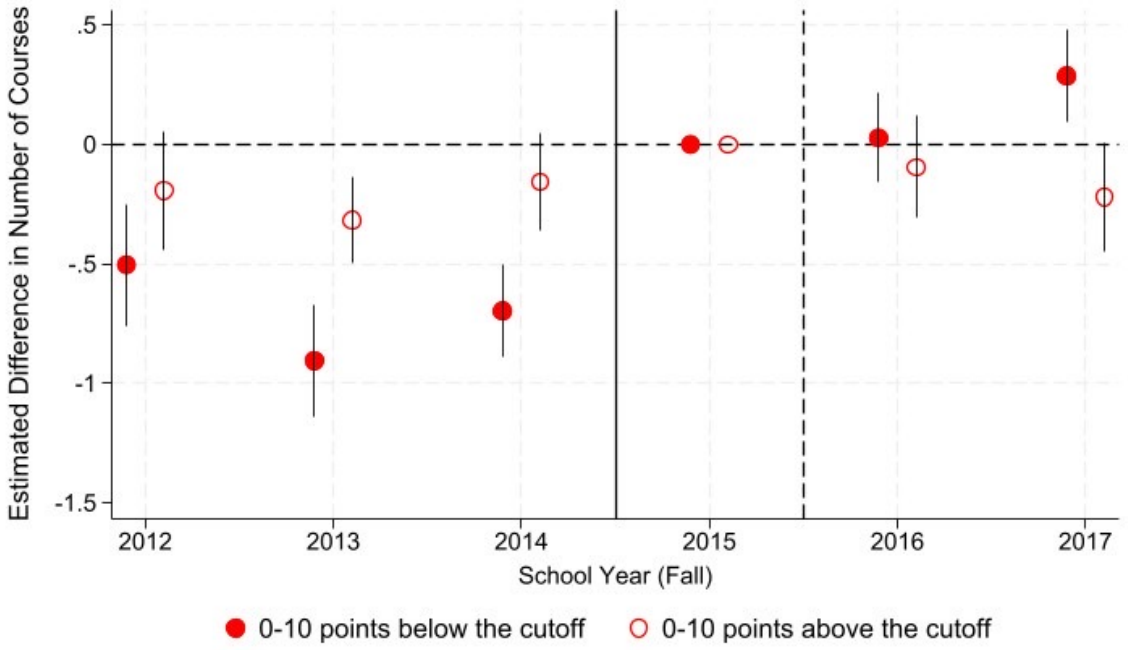
Consequences of bright lines for high school test scores (10th grade ELA score)



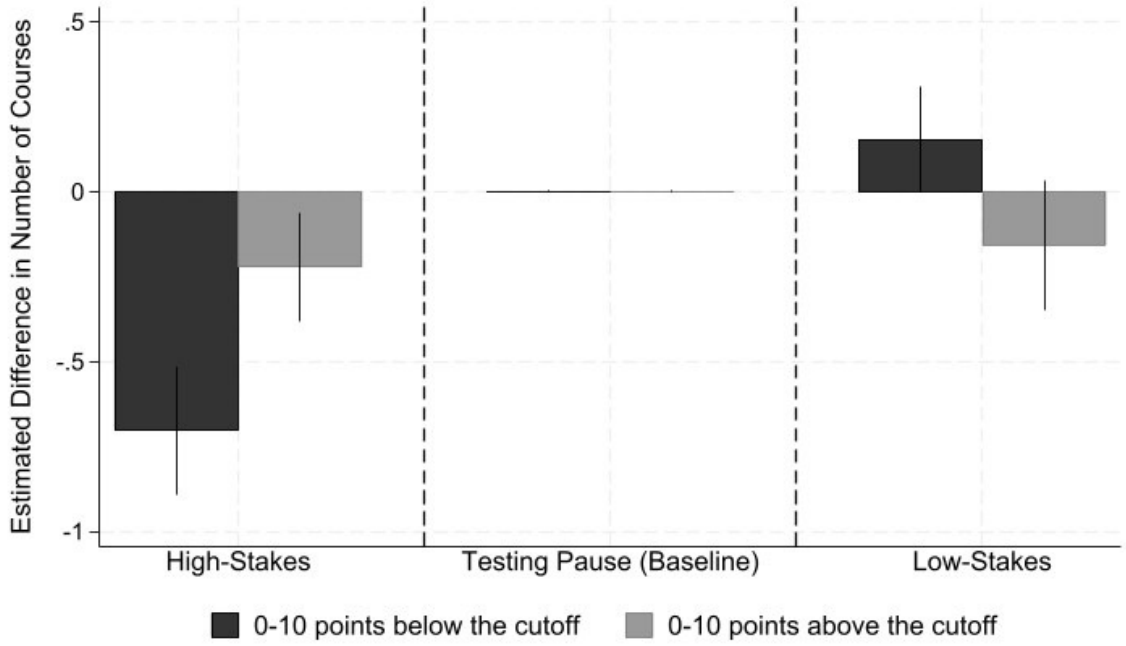
Consequences of bright lines for high school test scores (10th grade ELA score)



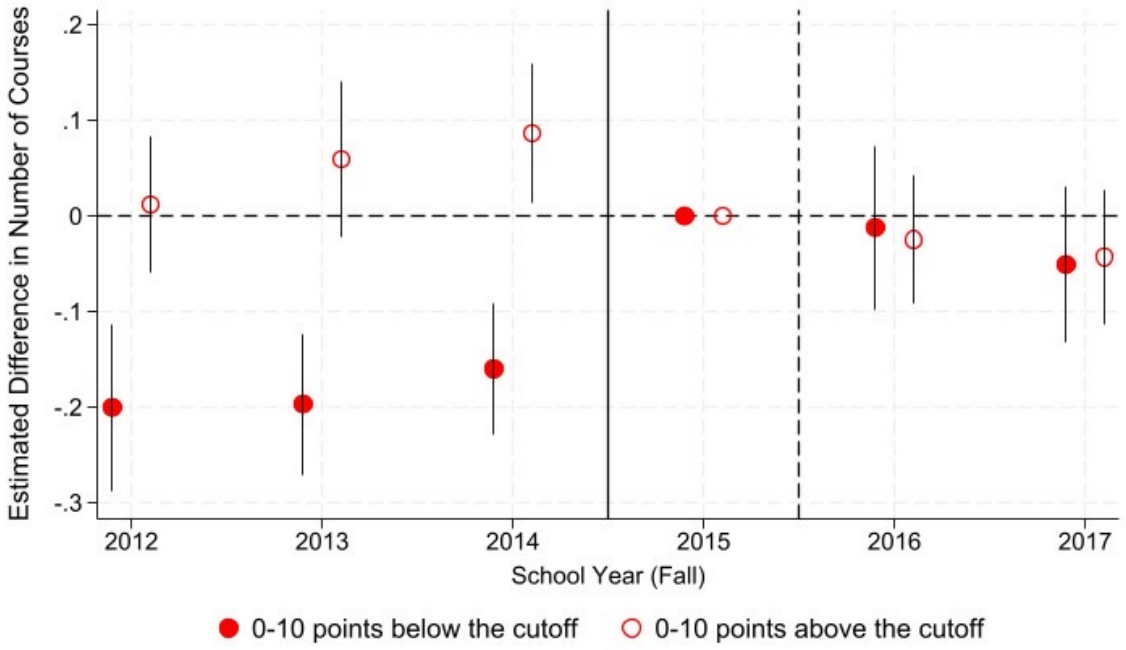
Consequences of bright lines for high school course-taking (number of advanced courses)



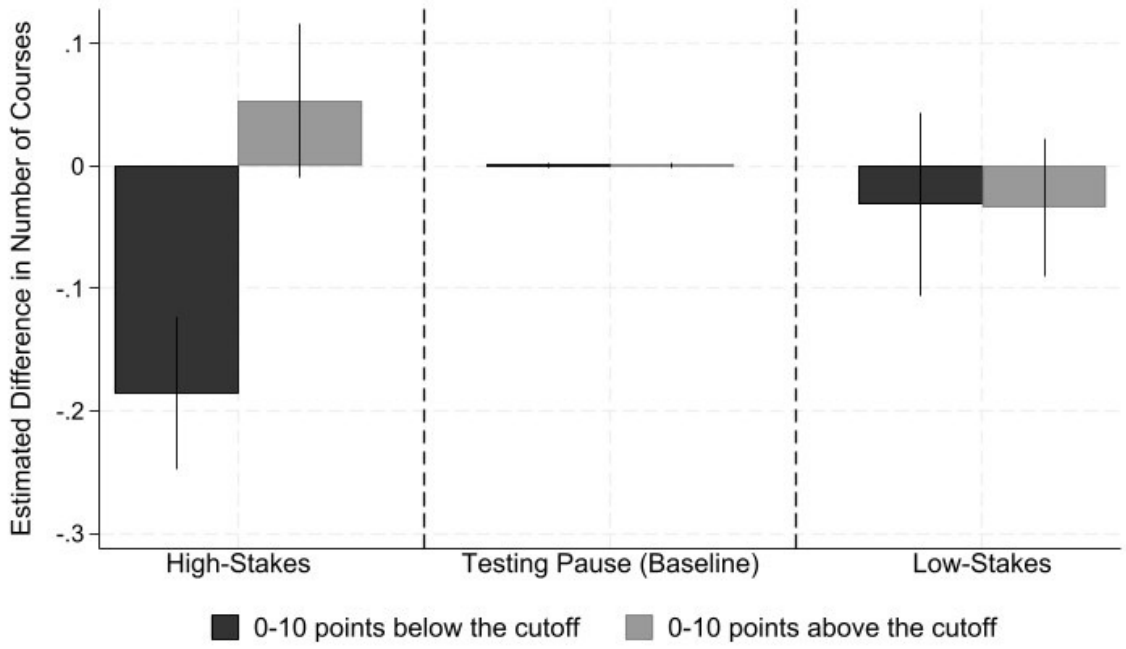
Consequences of bright lines for high school course-taking (number of advanced courses)



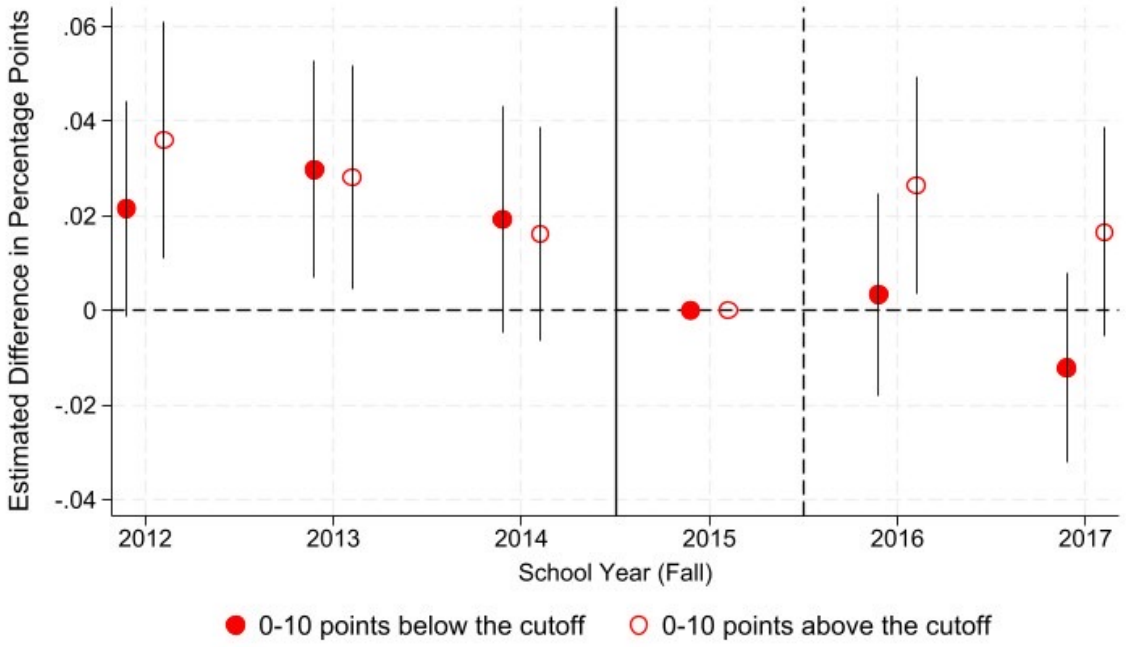
Consequences of bright lines for high school course-taking (number of CTE courses)



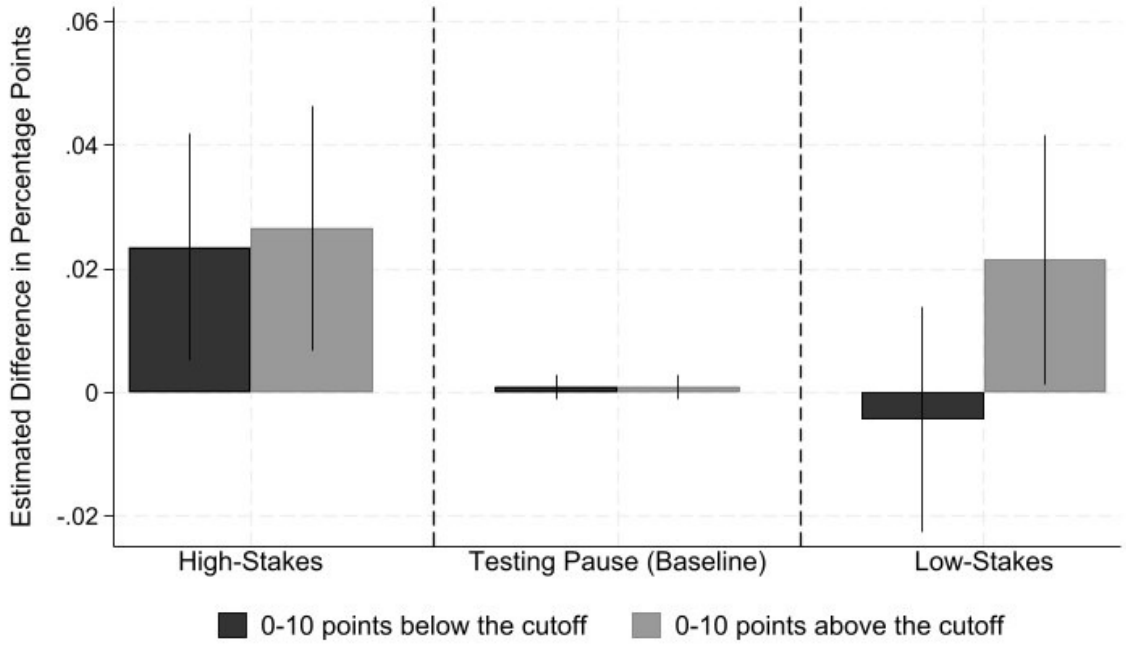
Consequences of bright lines for high school course-taking (number of CTE courses)



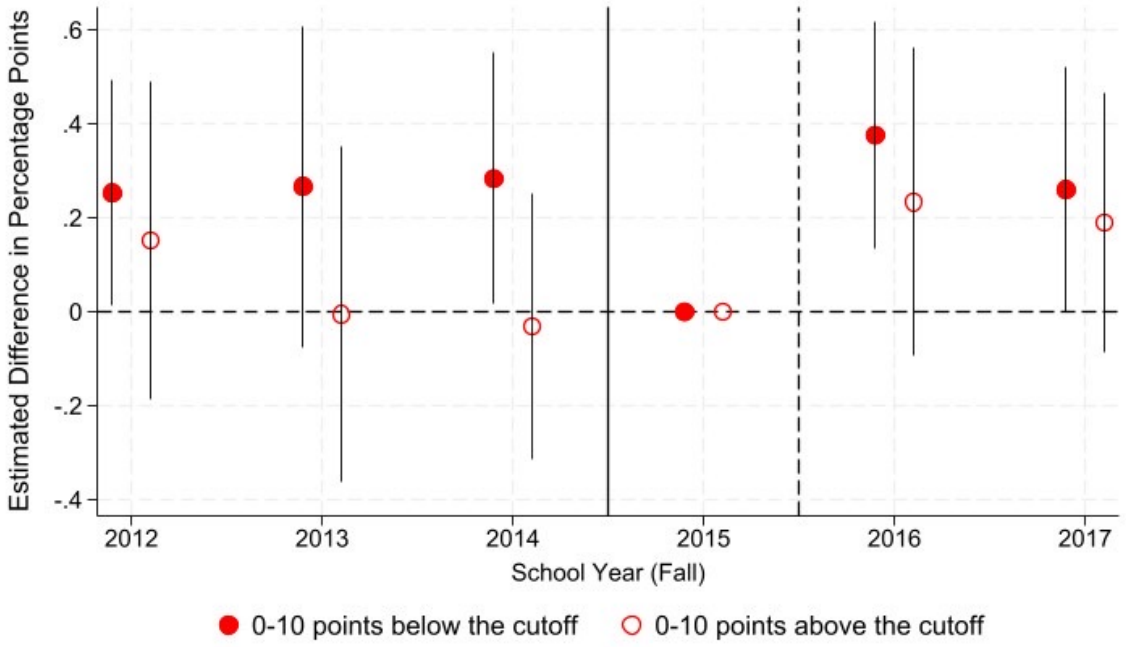
Consequences of bright lines for high school probability of being suspended



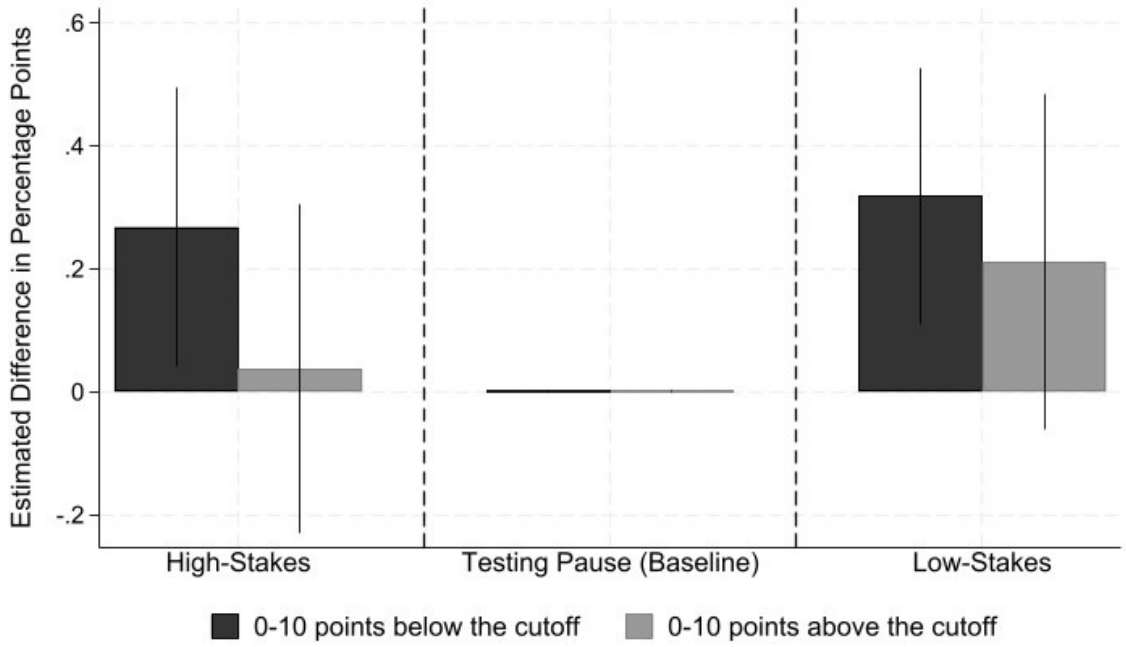
Consequences of bright lines for high school probability of being suspended



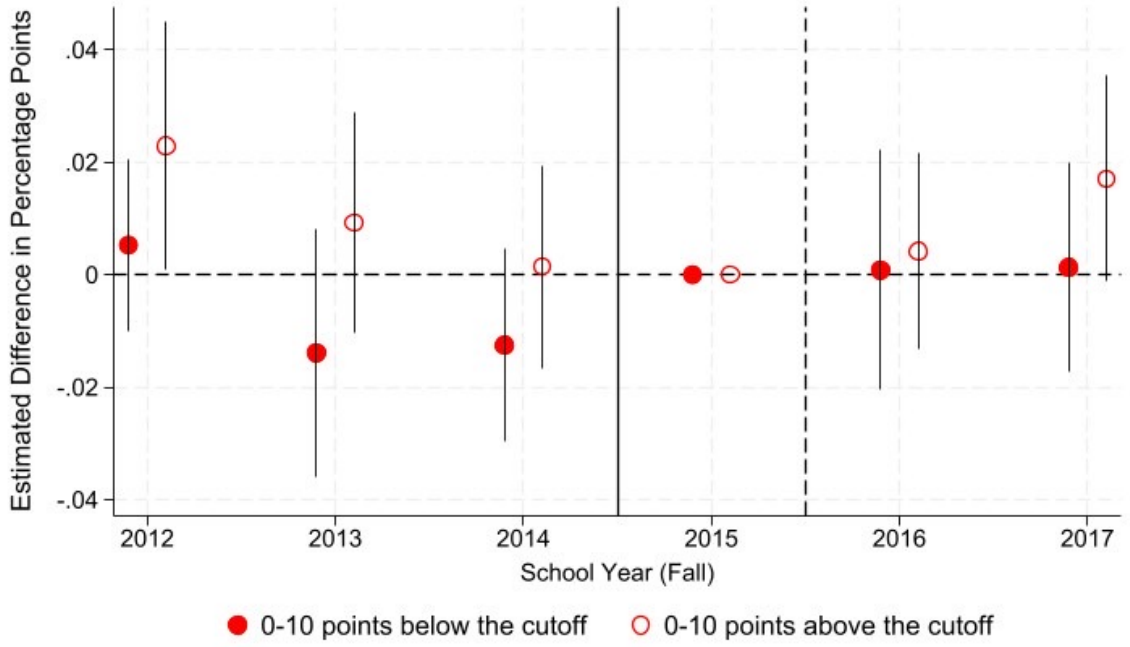
Consequences of bright lines for high school absence rate



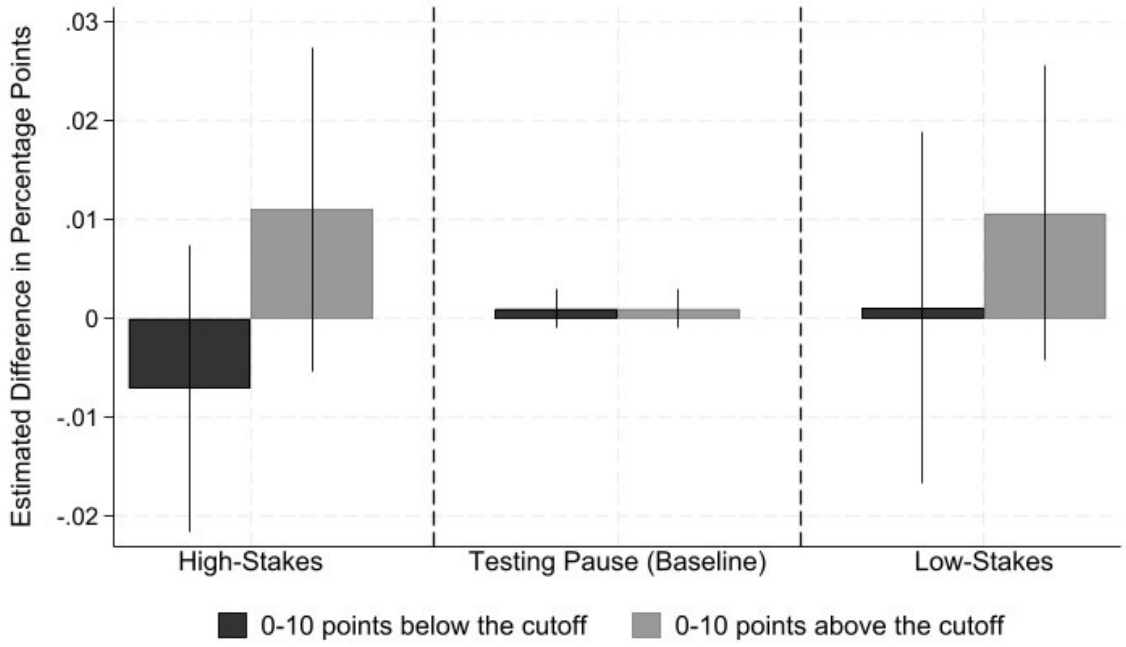
Consequences of bright lines for high school absence rate



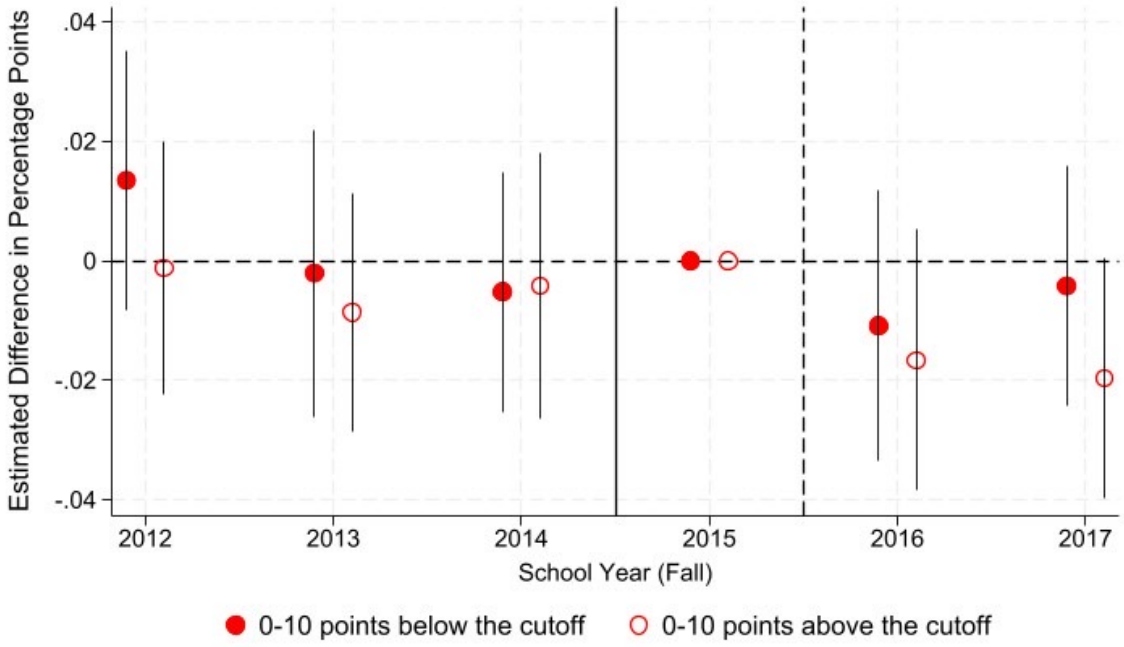
Consequences of bright lines for probability of earning a high school diploma



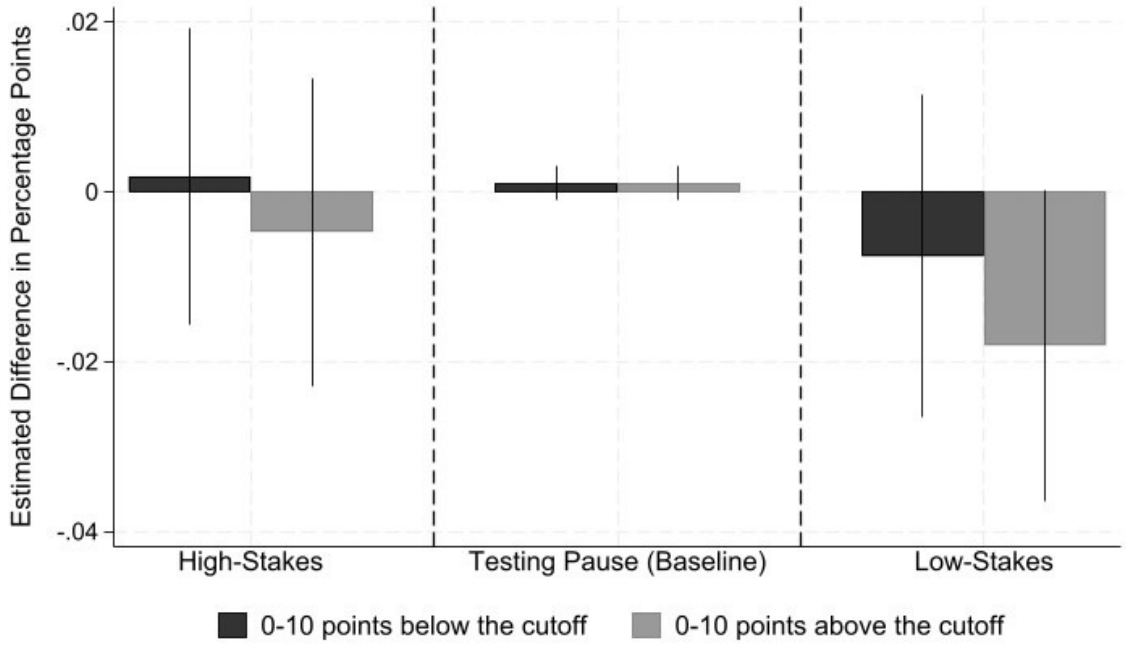
Consequences of bright lines for probability of earning a high school diploma



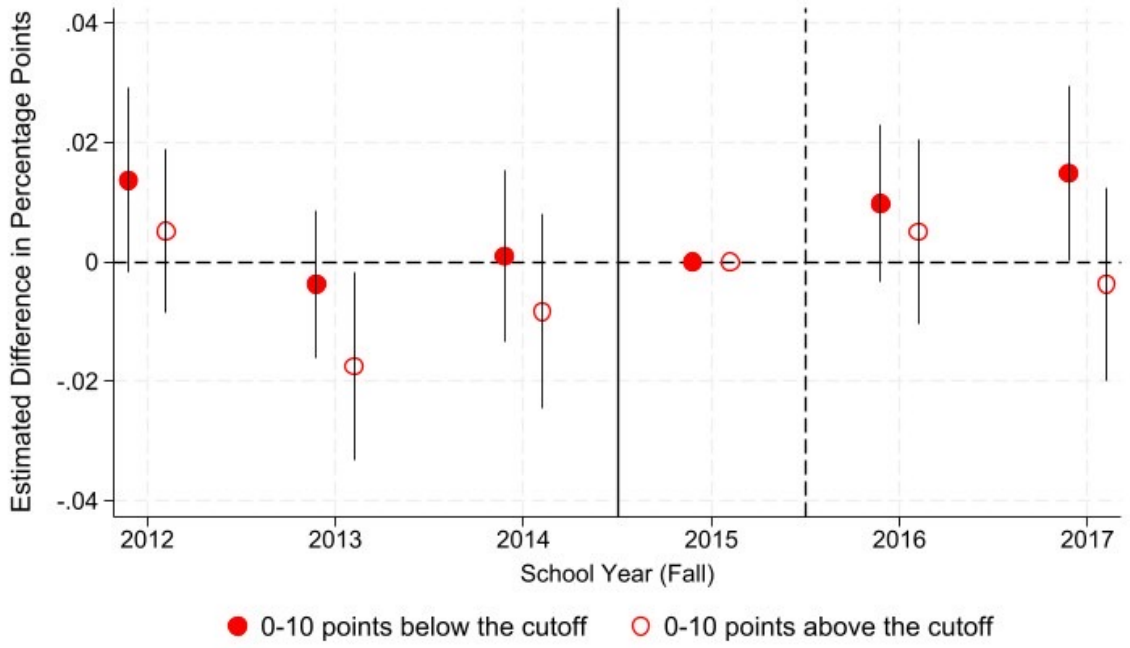
Consequences of bright lines for probability of postsecondary enrollment



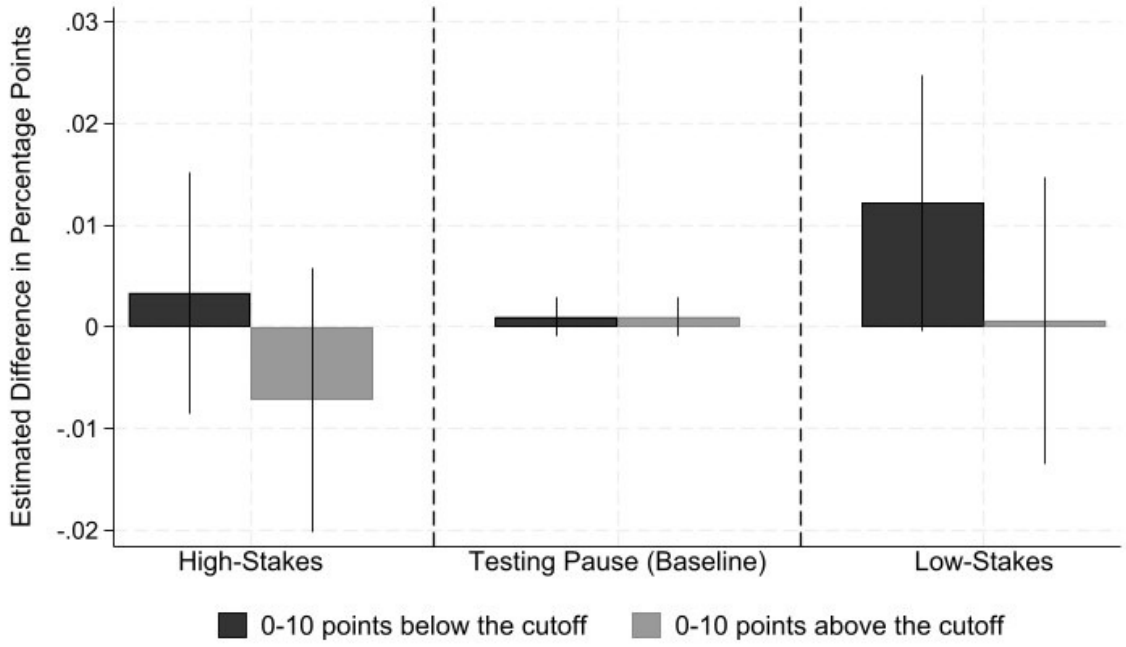
Consequences of bright lines for probability of postsecondary enrollment



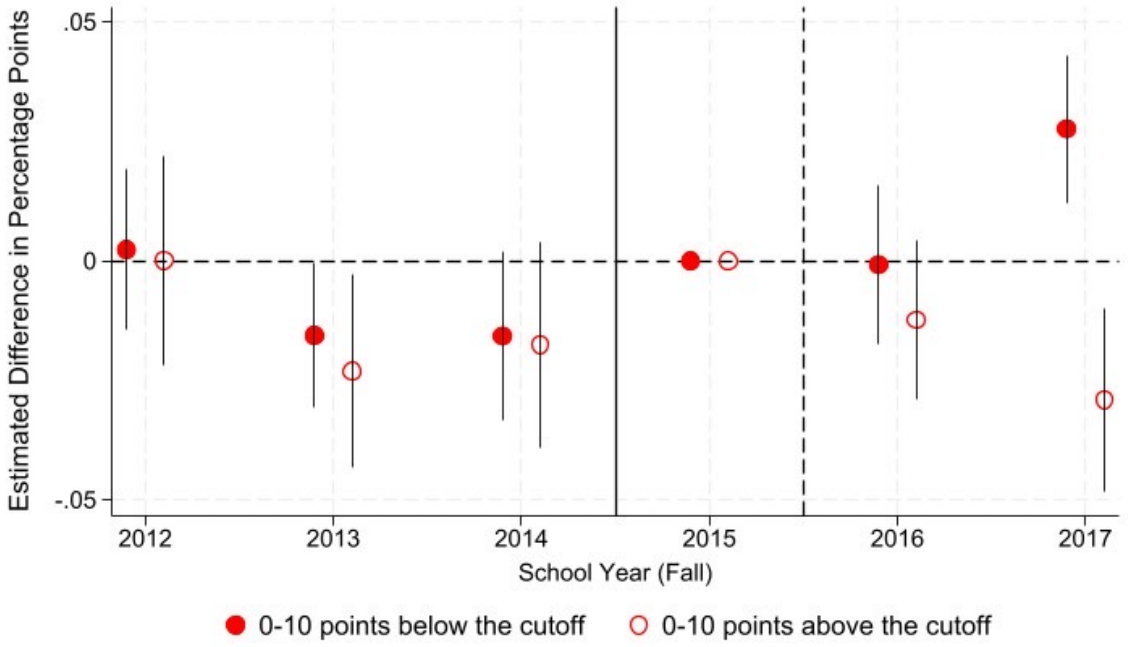
Consequences of bright lines for probability of enrolling in a highly-selective college



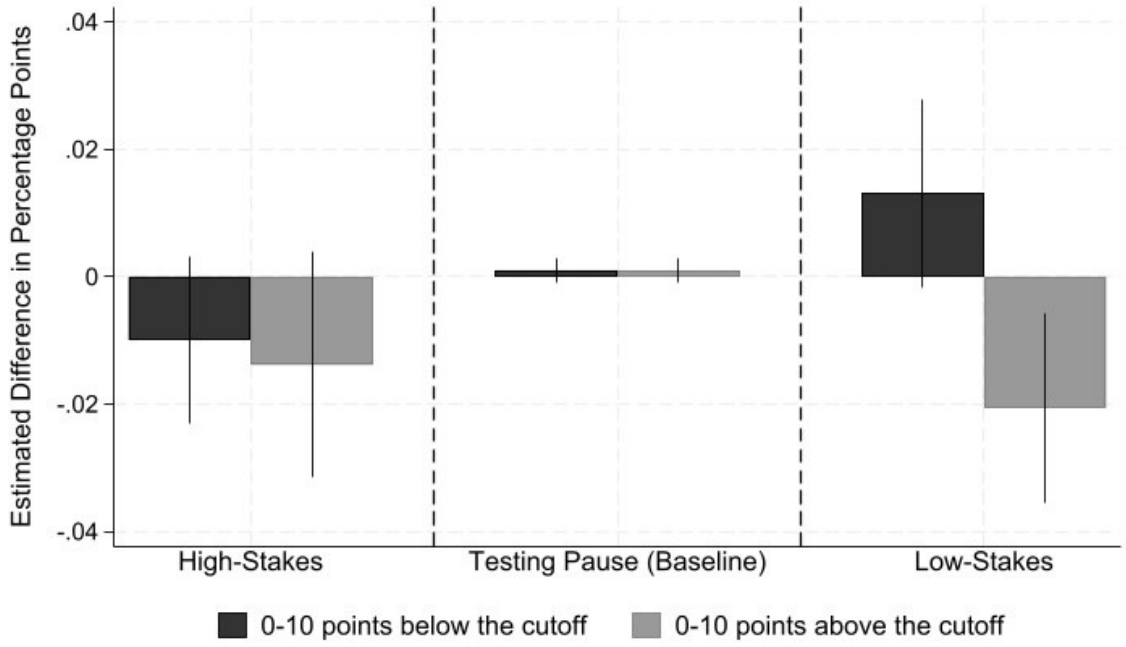
Consequences of bright lines for probability of enrolling in a highly-selective college



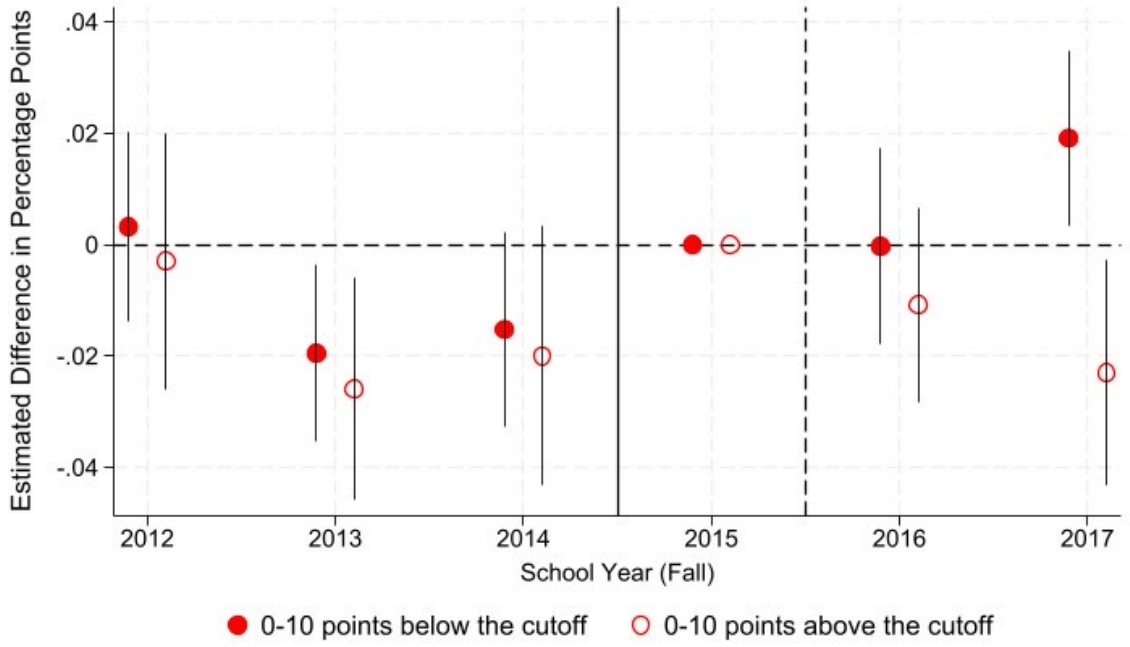
Consequences of bright lines for probability of persisting in college beyond second year



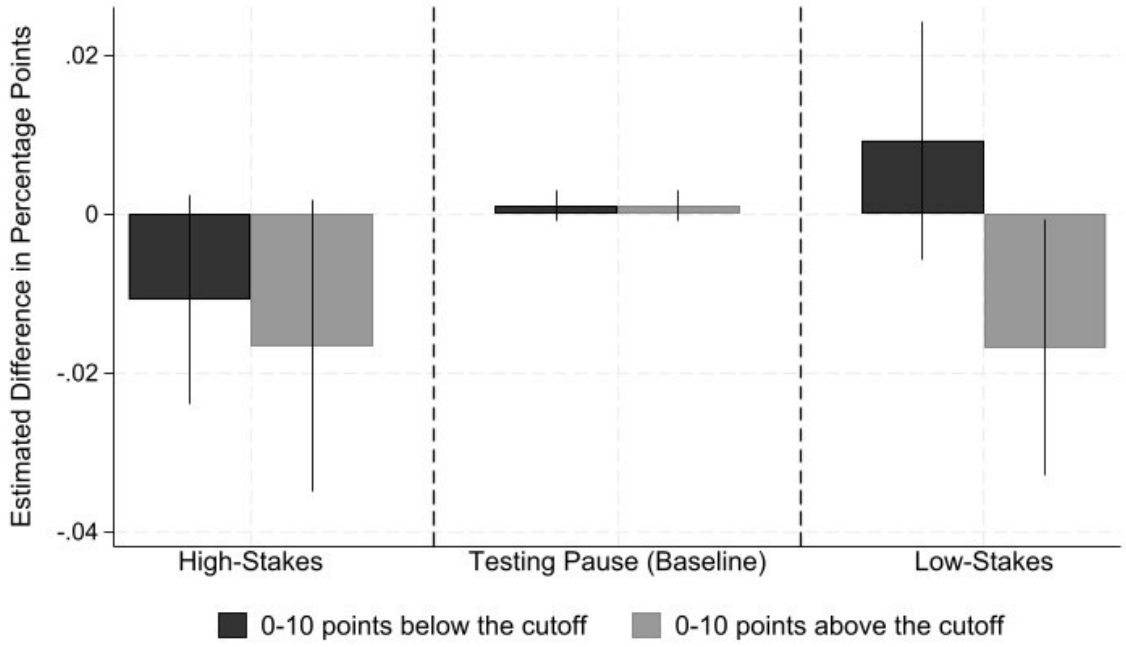
Consequences of bright lines for probability of persisting in college beyond second year



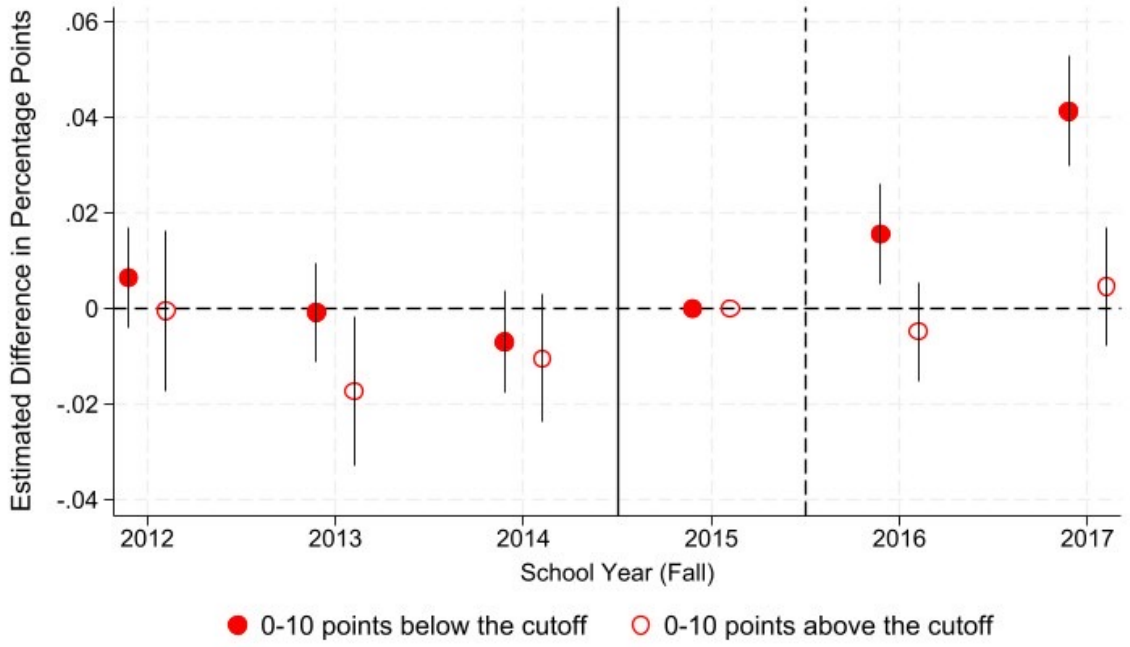
Consequences of bright lines for probability of earning a 2 or 4 year degree



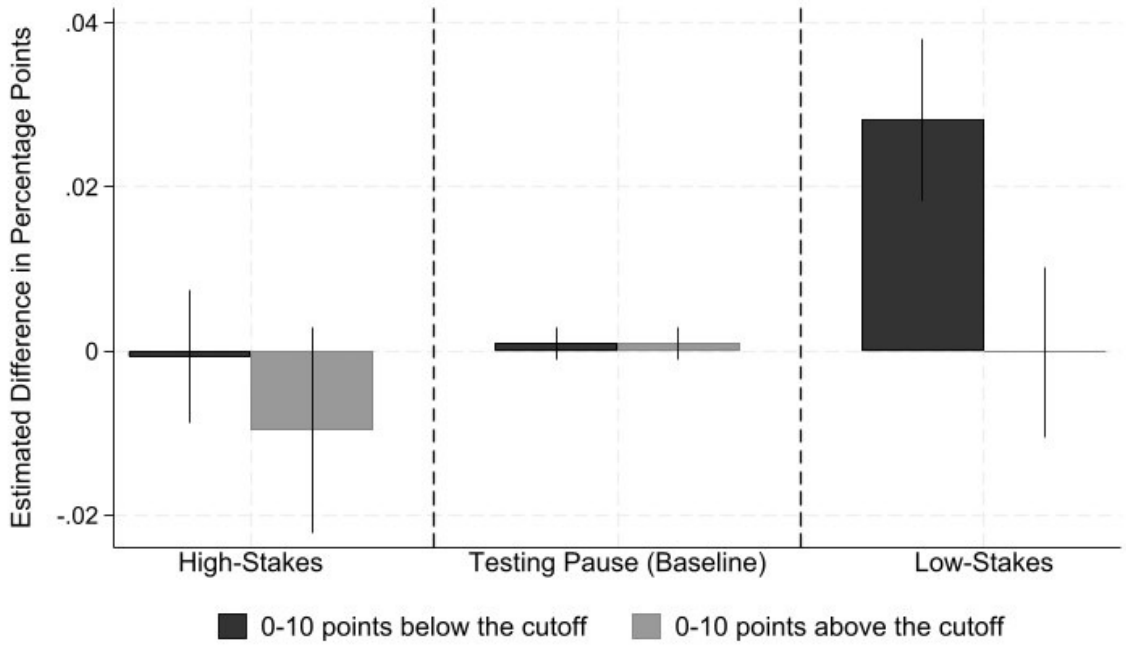
Consequences of bright lines for probability of earning a 2 or 4 year degree



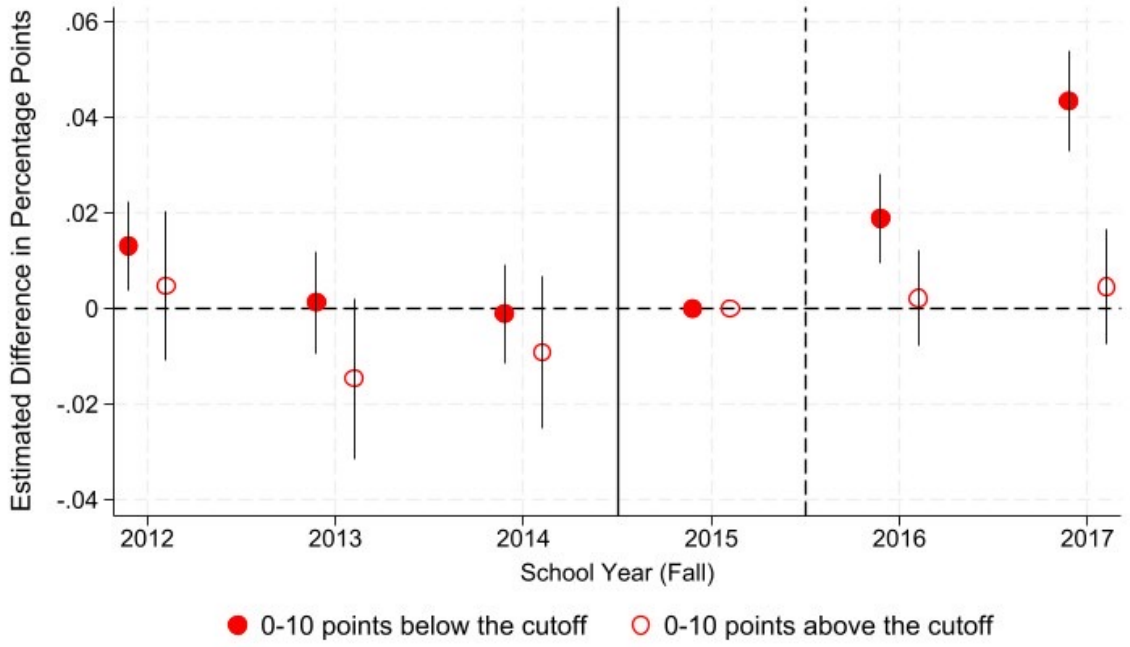
Consequences of bright lines for probability of earning a 4 year degree



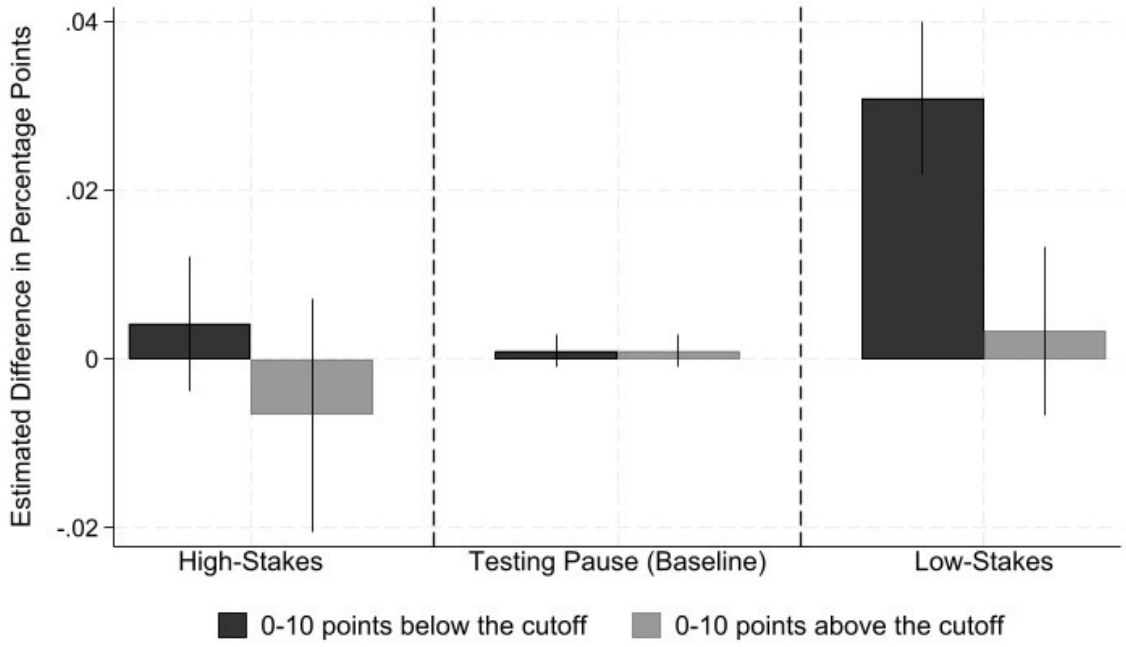
Consequences of bright lines for probability of earning a 4 year degree



Consequences of bright lines for probability of earning a degree from a highly selective inst



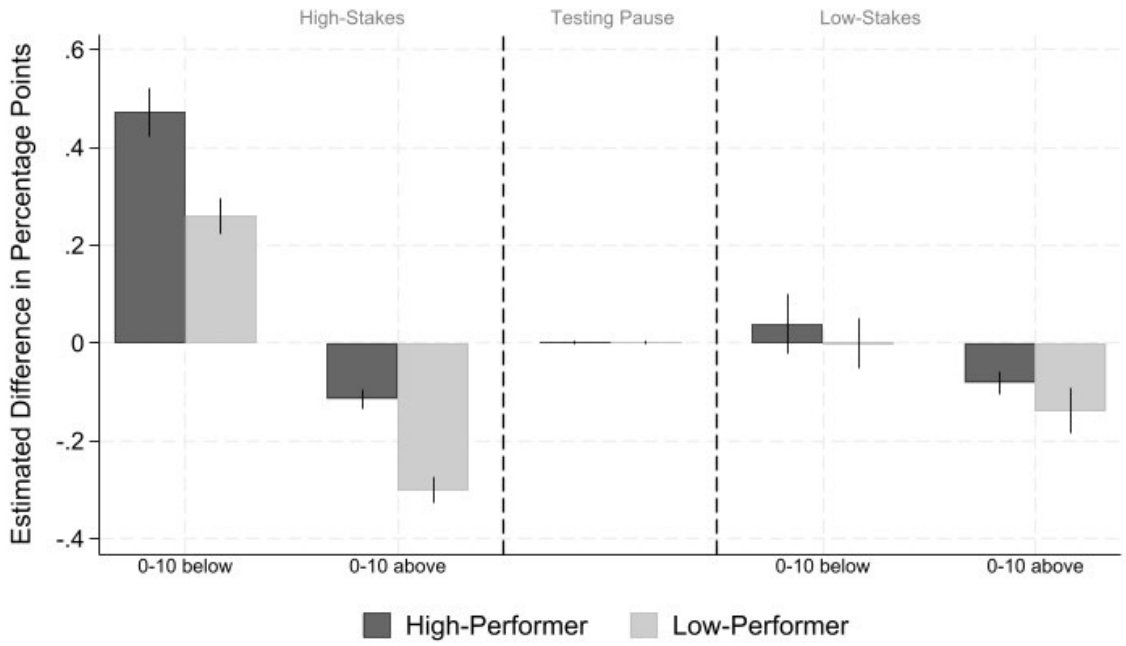
Consequences of bright lines for probability of earning a degree from a highly selective inst



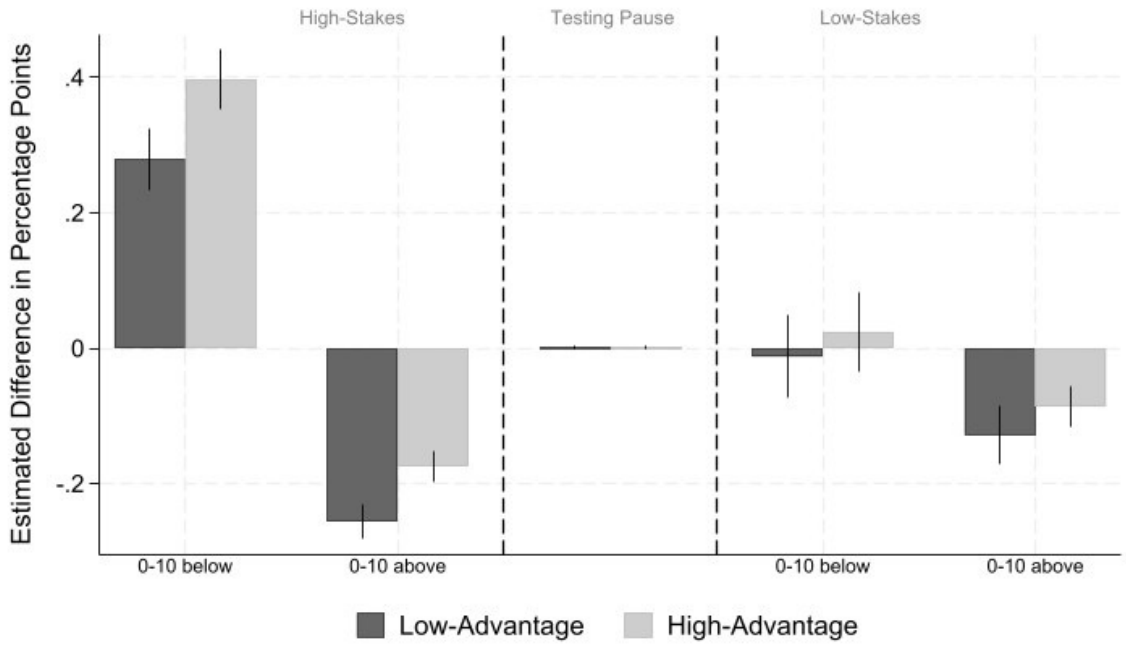
Results so far in the presentation

- High-stakes regime: Apparent over-treatment of kids just below the line and under-treatment of kids just above the line
- Low-stakes regime: Apparent under-treatment of kids just above the line continued
- Looking to longer-run outcomes, it appears that kids below the line have better LR outcomes under low-stakes regime, and kids above the line have better LR outcomes in the “pause” year
  - Both consistent with a story of over-treatment of kids just below the line and under-treatment of kids just above the line

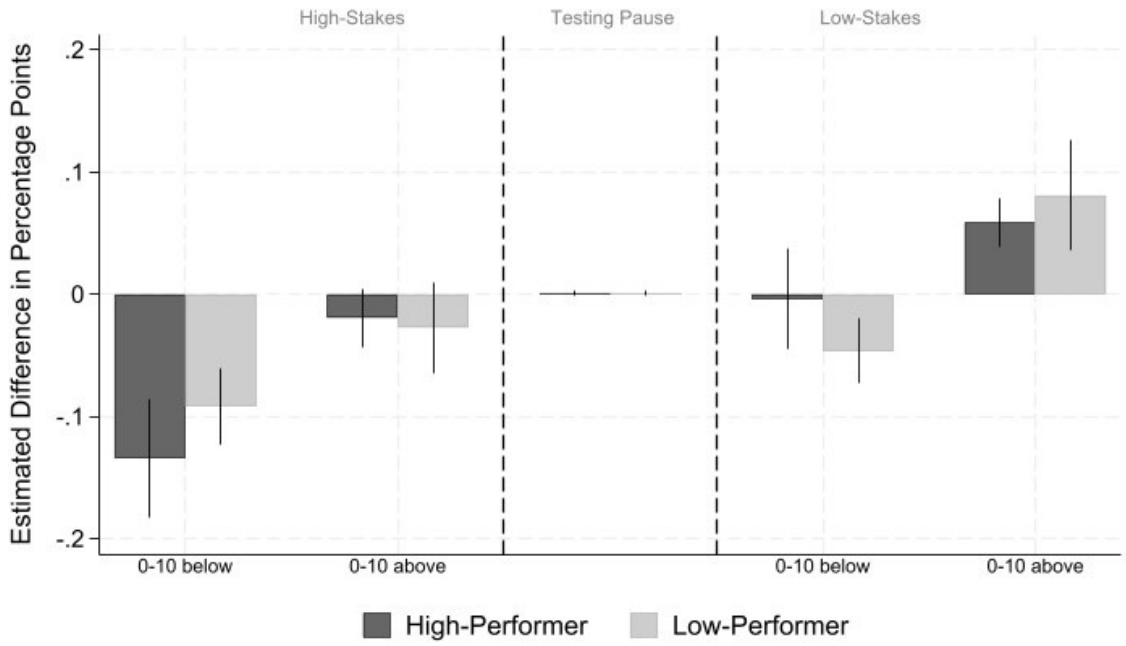
Heterogeneity in prob(remedial ELA class): Differences by historical math performance



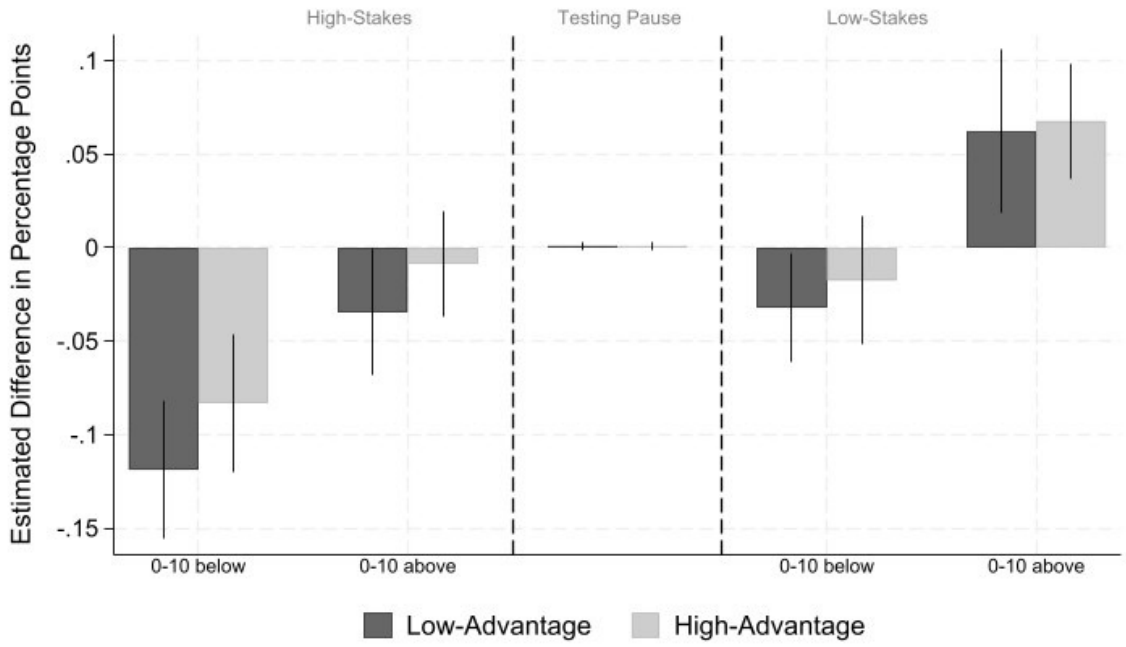
Heterogeneity in prob(remedial ELA class): Differences by family advantage



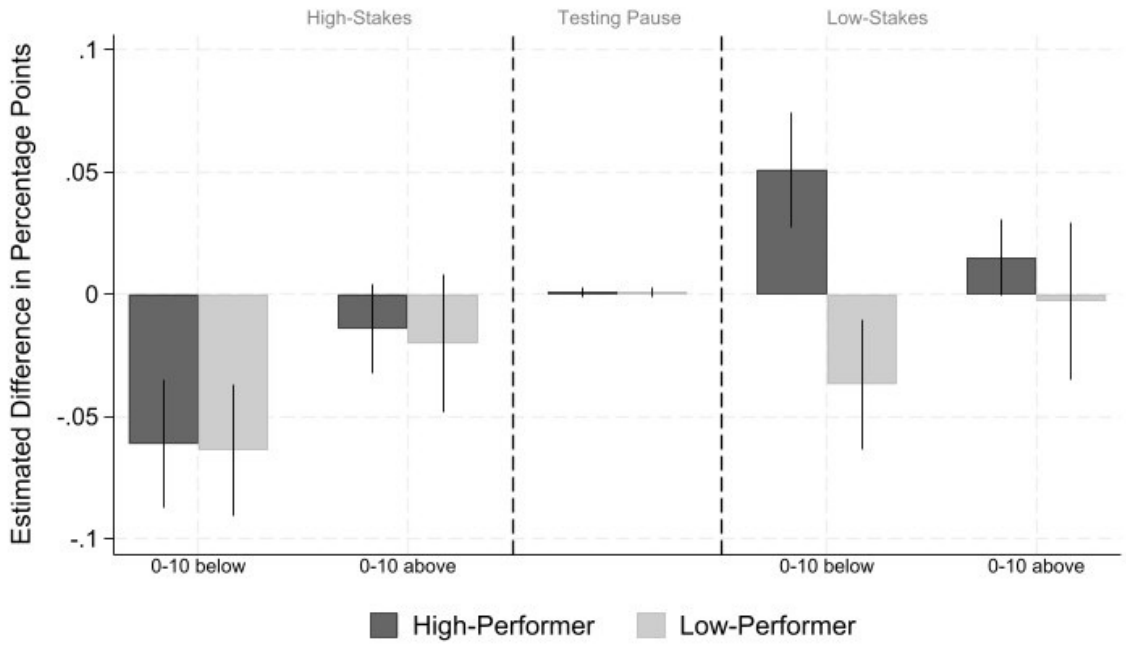
Heterogeneity in prob(advanced ELA class): Differences by historical math performance



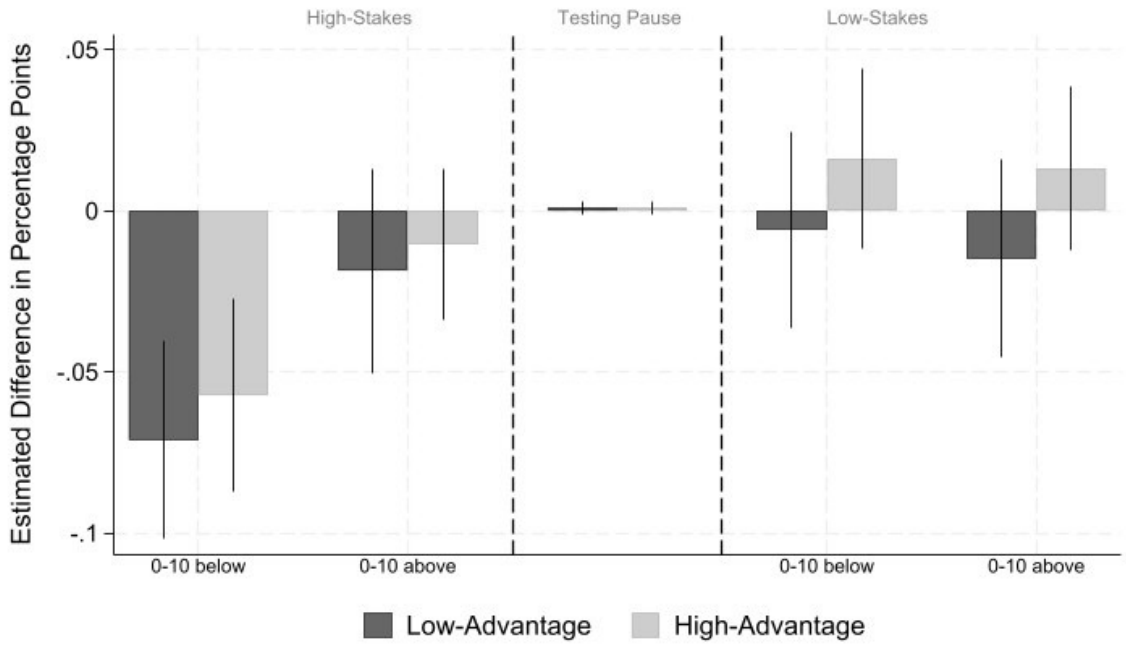
Heterogeneity in prob(advanced ELA class): Differences by family advantage



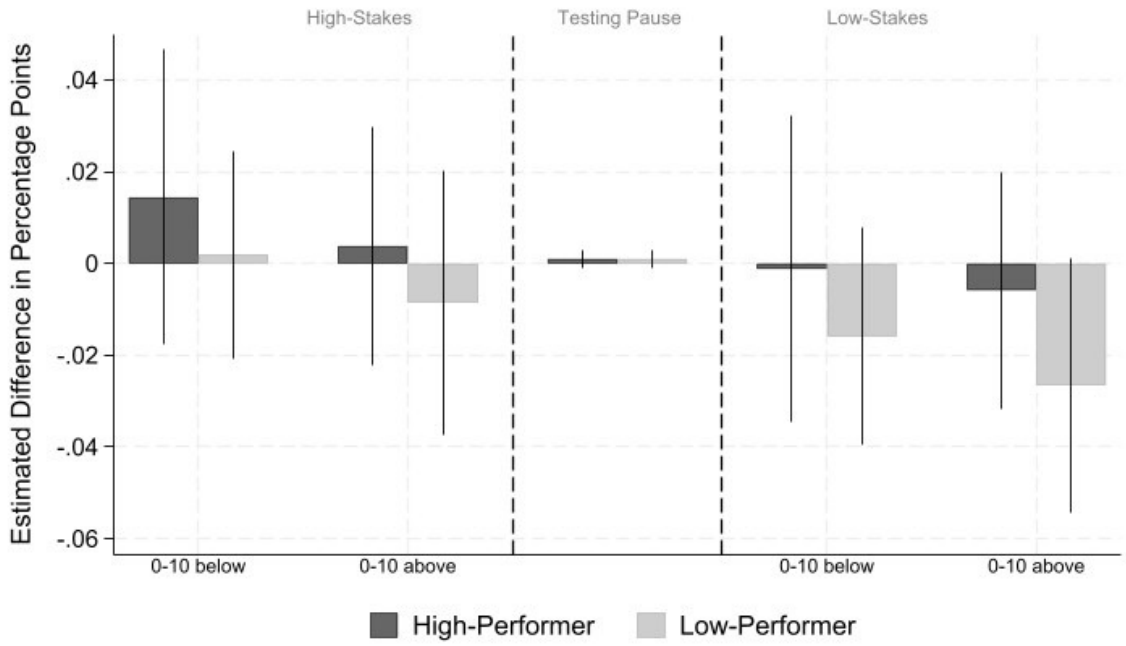
Heterogeneity in prob(advanced non-ELA class): Differences by historical math perform



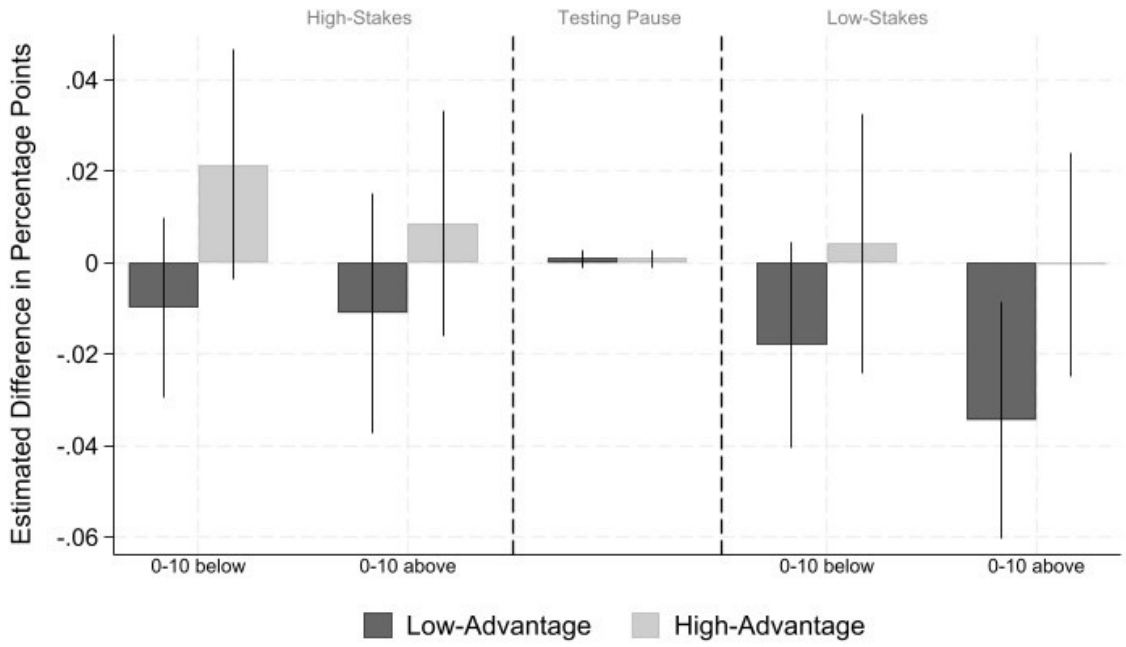
Heterogeneity in prob(advanced non-ELA class): Differences by family advantage



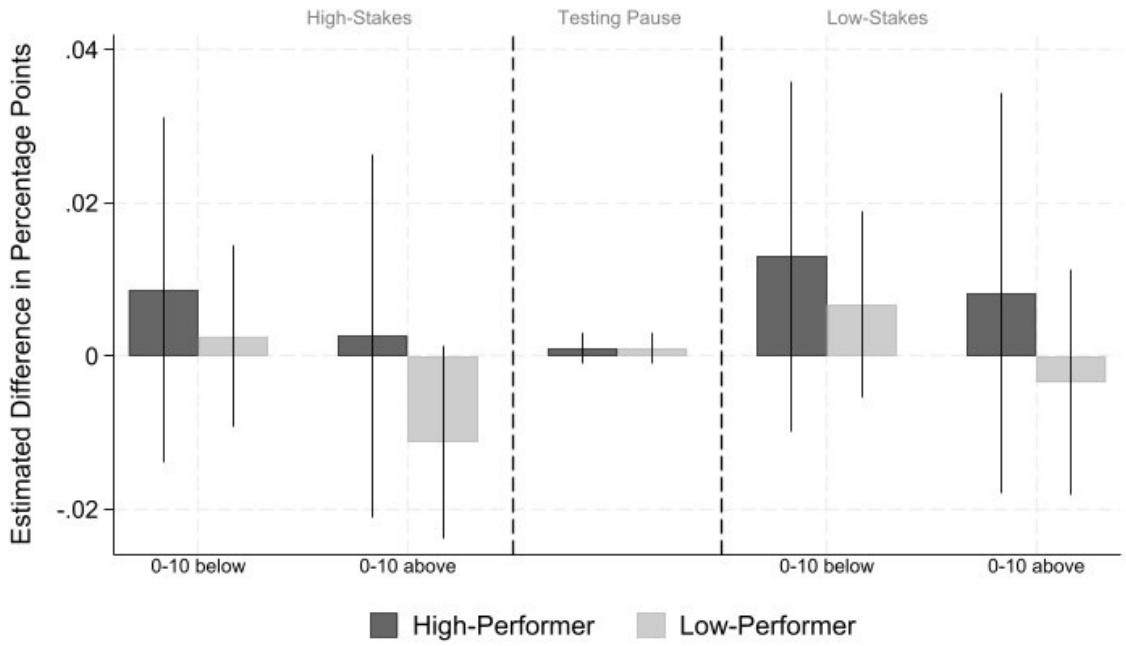
Heterogeneity in prob(enrolled 2/4 yr college): Differences by historical math performance



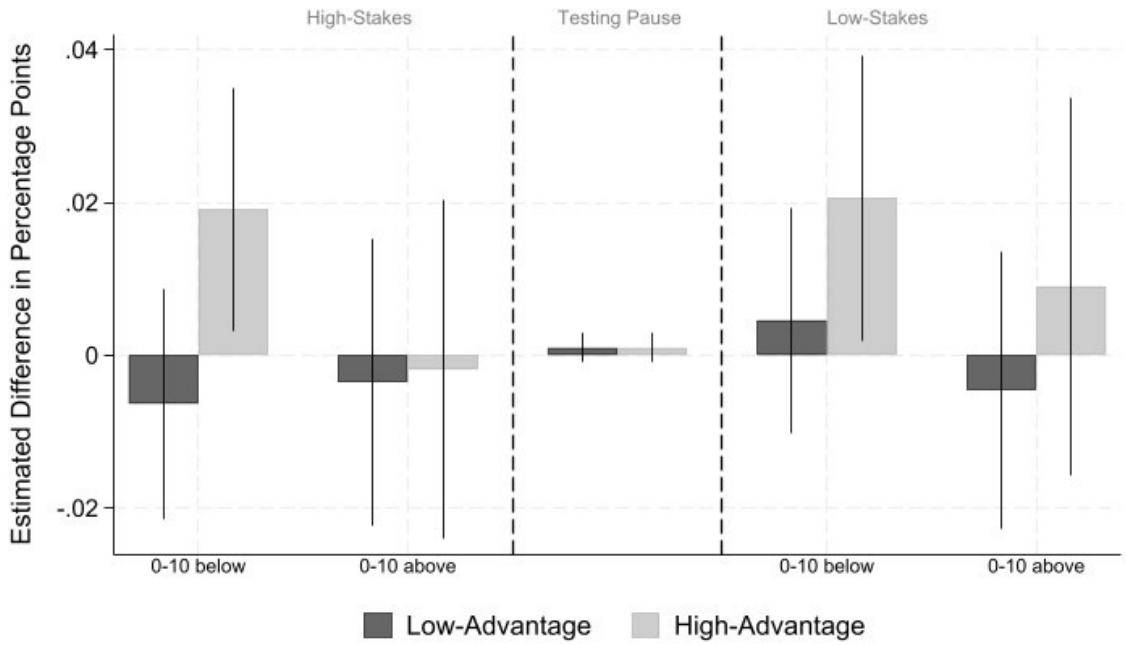
Heterogeneity in prob(enrolled 2/4 yr college): Differences by family advantage



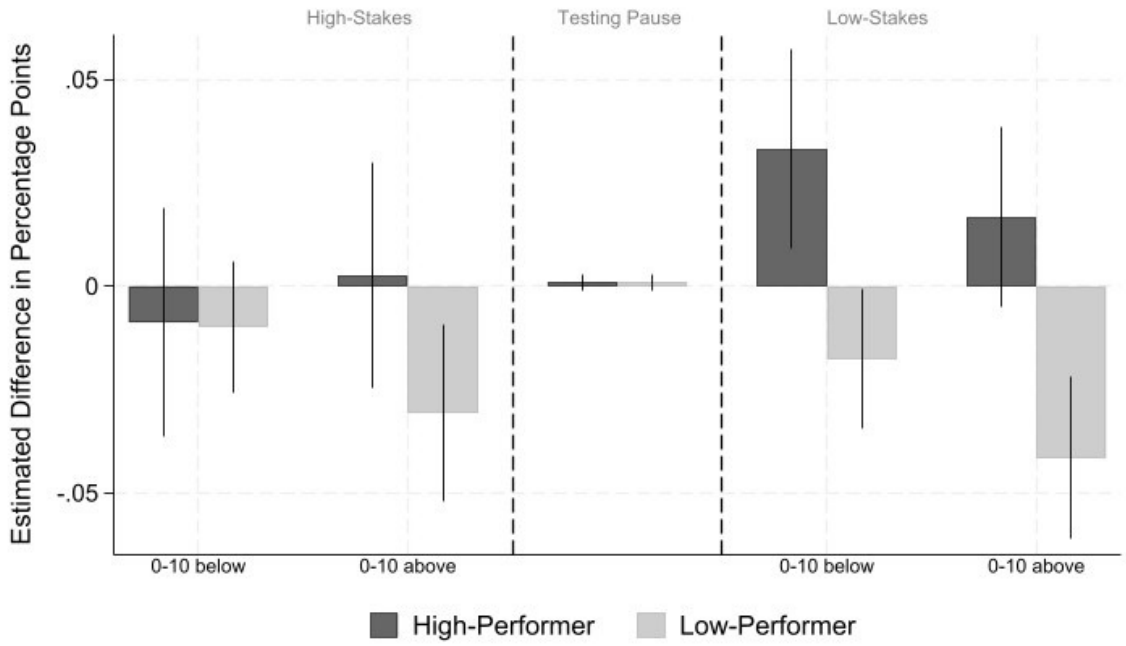
Heterogeneity in prob(enrolled high selective): Differences by historical math performance



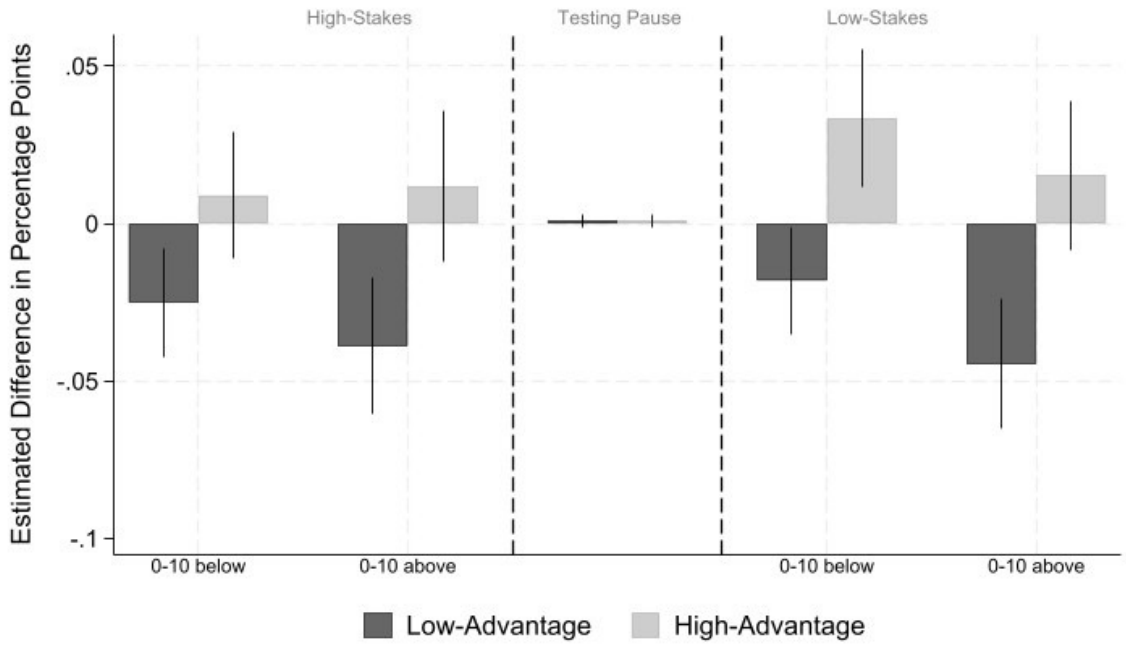
Heterogeneity in prob(enrolled high selective): Differences by family advantage



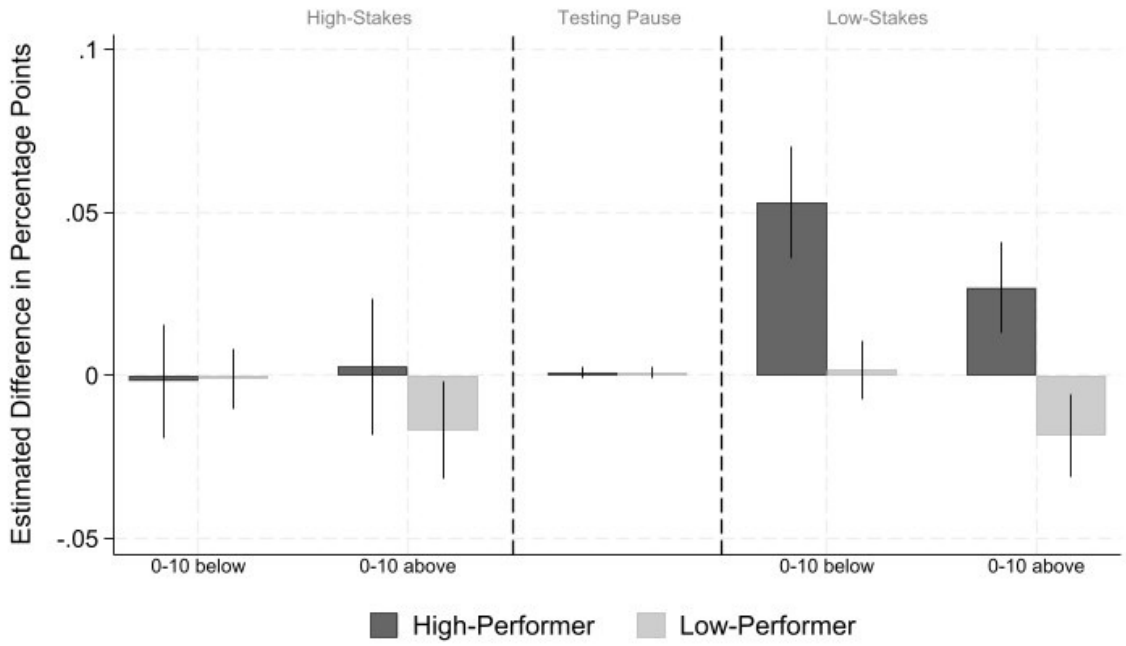
Heterogeneity in prob(earned 2/4 yr degree): Differences by historical math performance



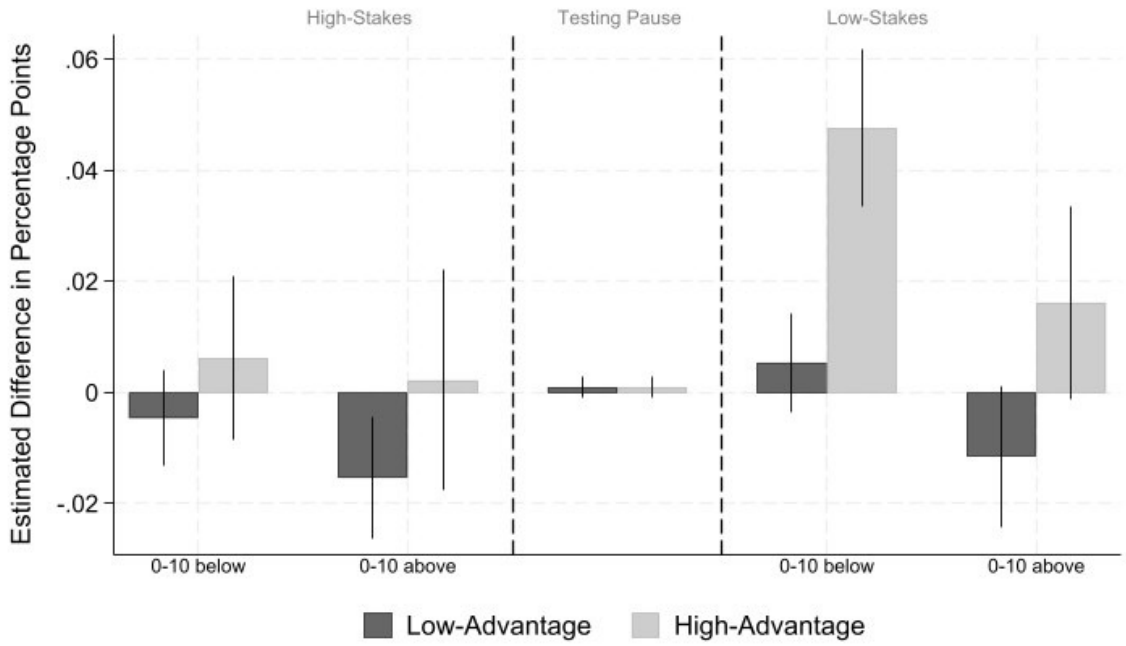
Heterogeneity in prob(earned 2/4 yr degree): Differences by family advantage



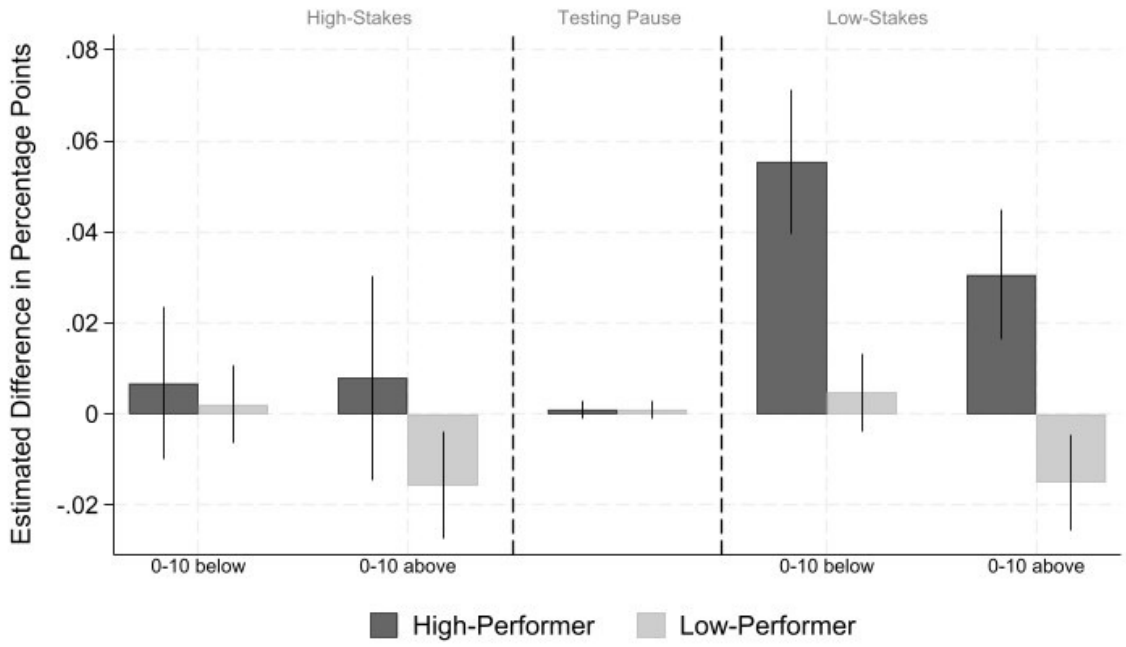
Heterogeneity in prob(earned 4 yr degree): Differences by historical math performance



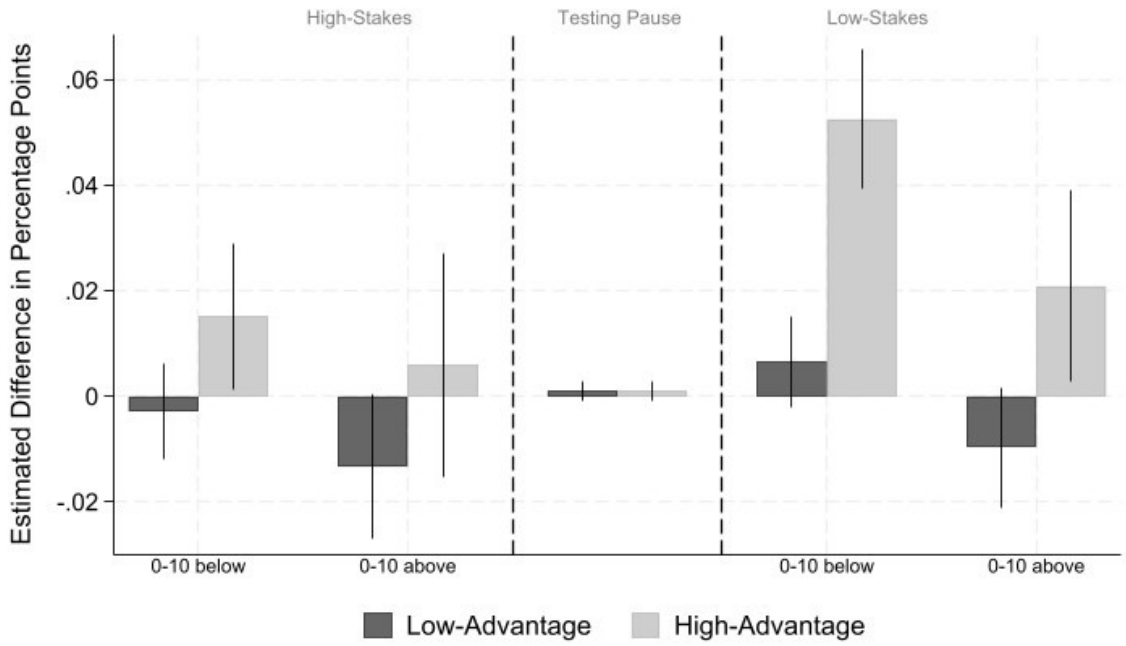
Heterogeneity in prob(earned 4 yr degree): Differences by family advantage



Heterogeneity in prob(earned high sel degree): Differences by historical math performance



Heterogeneity in prob(earned high sel degree): Differences by family advantage



Results so far in the presentation

- High-stakes regime: Apparent over-treatment of kids just below the line and under-treatment of kids just above the line
- Low-stakes regime: Apparent under-treatment of kids just above the line continued
- Looking to longer-run outcomes, it appears that kids below the line have better LR outcomes under low-stakes regime, especially if high math performer/high advantage, and kids above the line have better LR outcomes in the "pause" year, especially if low math performer
  - Both consistent with a story of over-treatment of kids just below the line and under-treatment of kids just above the line

Take-away messages 1

- Cutpoints matter – for students on both sides of the cutpoint
  - The RD comparisons that we typically make compare treatment for students on both sides of the cutpoint to one another…but not against their counterfactual condition that would have occurred absent the cutpoint itself
  - We observe evidence that, when stakes are high, kids below the cutpoint appear to be over-treated, and kids above the cutpoint appear to be under-treated…and the under-treatment above the cutpoint continues even when the stakes are low
  - Evidence that these differential treatments have consequences that continue into college outcomes
  - One more reason to think carefully about where proficiency cutpoints are set, above and beyond the standards argument

Take-away messages 2

- Assigning high stakes to cutpoints may lead to more differentiation between kids on either side of the cutpoint than might be ideal
- Moreover, cutpoints are anchoring devices that have important implications – not just for differentiating treatments between people on either side of the cutpoint, but also apparently for determining what a given place in the distribution "means"…
- While we don't saw whether the cutpoint is too high, too low, or "just right", this paper shows that where we draw the line matters – and not only for determining who is proficient and who is not!