Large Impacts of Grade Retention in New York City

Douglas Almond, Ajin Lee, Amy Ellen Schwartz

NBER Economics of Education Program Meeting, Fall 2023

November 30, 2023

Motivation

- Schools use grade retention to help students who are struggling academically
  - Forces the student to repeat the same grade
  - Gives another year to "catch-up" $\Rightarrow$ (short-term) achievement gains
  - Potential scarring effects in the longer term
- What we know about retention:
  - It affects a lot of students: Around 1 million students are retained each year in the US (NCES, 2019)
  - Short-term increases in test scores (Figlio and Özek, 2020; Greene and Wainters, 2007; Greene and Wainters, 2012; Hwang and Koedel, 2023; Jacob and Lefgren, 2004; Mariano and Martorell, 2013; Matsudaira, 2008; Schwerdt et al., 2017)
  - Negative effects on longer-term outcomes (Eren et al., 2017; Eren et al., 2022)
  - Null or mixed findings (Jacob and Lefgren, 2008; Schwerdt et al., 2017)

Standardized Testing and Grade Retention

- Impact of grade retention on longer-term outcomes
- Consider failing a New York's annual statewide proficiency exam in grades 3-8
  - Nationally, test-based grade retention policies are common
  - In 2019, 18 states and many large school districts used test-based such policies
- Retention regression discontinuity design
  - What happens when you "just fail" an annual standardized test?
  - Assumption is that only thing that is discontinuous at threshold is treatment (retention)
- If pass both Math and English exams, you will be promoted
- If fail either Math or English, you are *likely* to be promoted, but you are *at risk* of getting retained $\Rightarrow$ discretion
  - Fuzzy RDD
  - Among those who are eligible for retention according to the test-based retention policy, only about 16% of students are retained in our data

Standardized Testing and Special Education

- In our setting, retention is not the only treatment that is triggered at the threshold
- Students are also more likely to receive special education services the following year when they fail an exam
  - Not a policy, empirically true
- We identify a problem and a solution
- Resolve this identification challenge by pursuing a *regression kink design*
  - Special education classification is not additionally affected by the test score conditional on failing
  - While the retention decision continues to be affected by "how badly" the student failed the exam
- Isolate the impact of retention in NYC

Retention in Grade

- Parents face a tradeoff
  - Better fit academically v. social promotion
  - long-term effects? probably not well known
- Teachers
  - Short-term benefits may be salient
  - Won't have this low-performing student again; externality?
- Student performance does improve in short run for marginally retained
  - Most consistent empirical finding
  - Retain too much?
- Long-run outcomes are likely more valued by society, parents, and students themselves
- Here, have a shock to perception of appropriateness for grade
  - Won't be able to disentangle various channels (teacher v. parent v. student roles)

Outline

Background
- Retention Policy

Data and Sample

Naive RDD

RK Design and Impacts
- Density of Running Variable

Impact Heterogeneity

First Stage Heterogeneity
- Compliant Subpopulation

Conclusion

Outline

Background
- Retention Policy

Data and Sample

Naive RDD

RK Design and Impacts
- Density of Running Variable

Impact Heterogeneity

First Stage Heterogeneity
- Compliant Subpopulation

Conclusion

Test-Based Promotion Policy

- School principal makes promotion decisions based on multiple criteria, in consultation with parents and teachers
- Importantly, students must show progress towards meeting the English and Math Standards
- Students in grades 3-8 take New York State Math and English Language Arts (ELA) tests each spring
- Student test scores are classified into four levels
  - Level 1 not meeting state learning standards, or "failing"
  - Level 2 partially meeting state learning standards
  - Level 3 meeting state learning standards
  - Level 4 exceeding state learning standards
- The cutoffs vary by subject and by year
- Scoring Level 2 in both tests essentially guarantees promotion, while students who score Level 1 in either test are at risk of being retained

Data and Sample Construction

- Statewide proficiency test results and retention data for students in grades 3-8 between academic year (AY) 2007/2008 and 2010/2011
- Merge with graduation data between AY 2011/2012 and 2018/2019
- Focus on graduation outcomes measured within 4 years of students entering the 9th grade
- Drop students who would not be in 12th grade by AY 2018/2019 if retained once
  - The probability of appearing in the graduation data is smooth across the threshold
- Focus on the first time students appear in the analysis sample

More on the sample

Data and Sample Construction

- School choice data (2007-2013) to characterize high school programs based on admission methods
- Selective high schools
  - Specialized high schools (average SAT score = 1429 in 2018; enroll 7% of NYC public school students)
    - Bronx High School of Science
    - Brooklyn Latin School
    - Brooklyn Technical High School
    - Fiorello H. LaGuardia High School of Music and Art and Performing Arts
    - High School for Math, Science and Engineering at City College
    - High School of American Studies at Lehman College
    - Queens High School for the Sciences at York College
    - Staten Island Technical High School
    - Stuyvesant High School
  - High schools using screened admission (93 out of 472 schools; enroll 39% of NYC public school students)

Summary Statistics

	Below cutoff [-10,0)	Above cutoff [0,10]	Above 10 scale scores	Full sample
Retention in year t+1	0.116	0.006	0.002	0.015
Special education in year t+1	0.314	0.219	0.087	0.123
Graduated high school in 4 years				
Any high school	0.516	0.610	0.809	0.758
Specialized high school	0.001	0.002	0.084	0.068
High school with screened admission	0.252	0.285	0.393	0.368
Enrolled in 9th grade				
Any high school	0.940	0.962	0.981	0.973
Specialized high school	0.001	0.002	0.086	0.070
High school with screened admission	0.281	0.315	0.410	0.388
Female	0.454	0.470	0.517	0.505
Asian	0.087	0.088	0.187	0.171
Black	0.373	0.374	0.284	0.296
Hispanic	0.487	0.467	0.374	0.395
White	0.053	0.071	0.153	0.135
Subsidized lunch	0.950	0.936	0.830	0.854
Observations	31,187	79,246	881,268	1,105,315

Outline

Background
    Retention Policy

Data and Sample

Naive RDD

RK Design and Impacts
    Density of Running Variable

Impact Heterogeneity

First Stage Heterogeneity
    Compliant Subpopulation

Conclusion

RD-Instrumental Variables Estimation

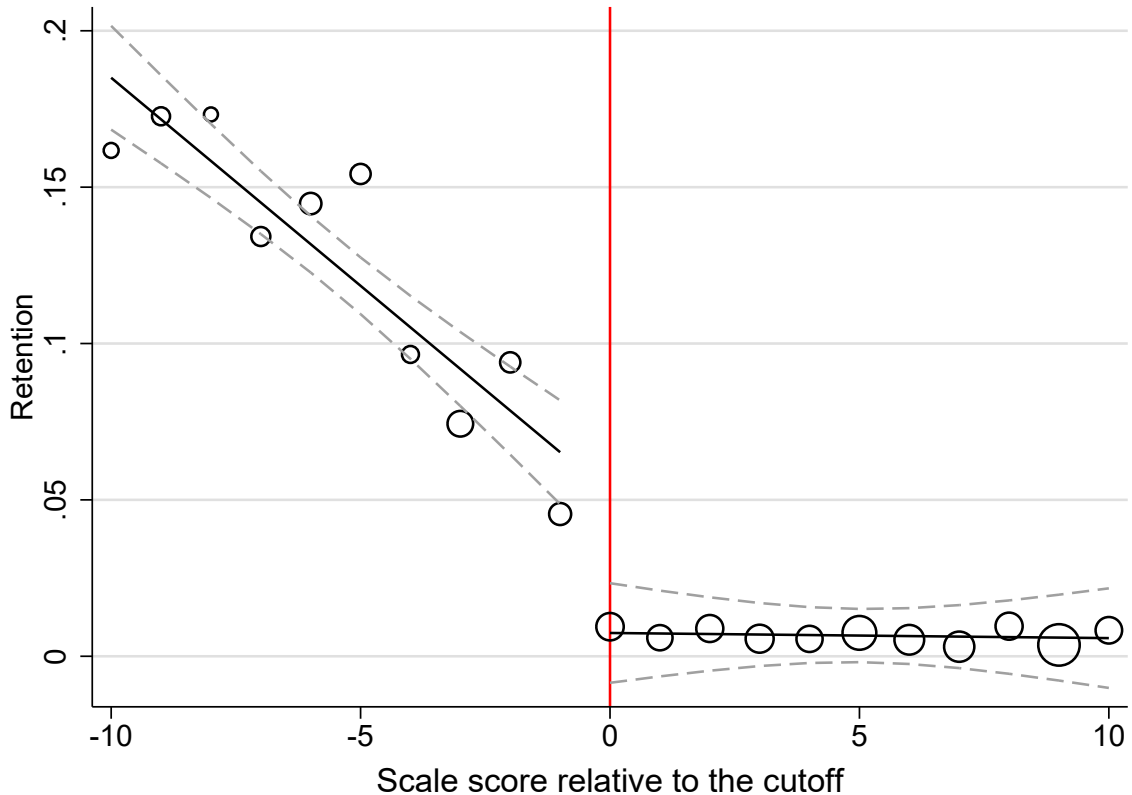
$$R_{igs,t+1} = \alpha_0 + \alpha_1 1[X_{igst} < 0] + \alpha_2 X_{igst} + \alpha_3 X_{igst} 1[X_{igst} < 0] + \theta_{gst} + \epsilon_{igst} \quad (1)$$

$$Y_{igs,t+x} = \beta_0 + \beta_1 1[X_{igst} < 0] + \beta_2 X_{igst} + \beta_3 X_{igst} 1[X_{igst} < 0] + \eta_{gst} + \nu_{igst} \quad (2)$$

- i individual, g grade, s subject, t year
- $R_{igs,t+1} = 1[\text{retained} = 1]$
- $Y_{igs,t+x}$ indicates graduation outcomes
- $X_{igst} = \min[\text{Math scale score - cutoff, ELA scale score - cutoff}]$
- θ_{gst} and η_{gst}: year×grade×subject dummies (cutoff fixed effects)
- Bandwidth: 10 scale scores below and above the threshold
- Robust standard errors
- $\frac{\hat{\beta}_1}{\hat{\alpha}_1}$ **represents the RD-IV estimate**

First Stage Effect of Exam Failure on Retention

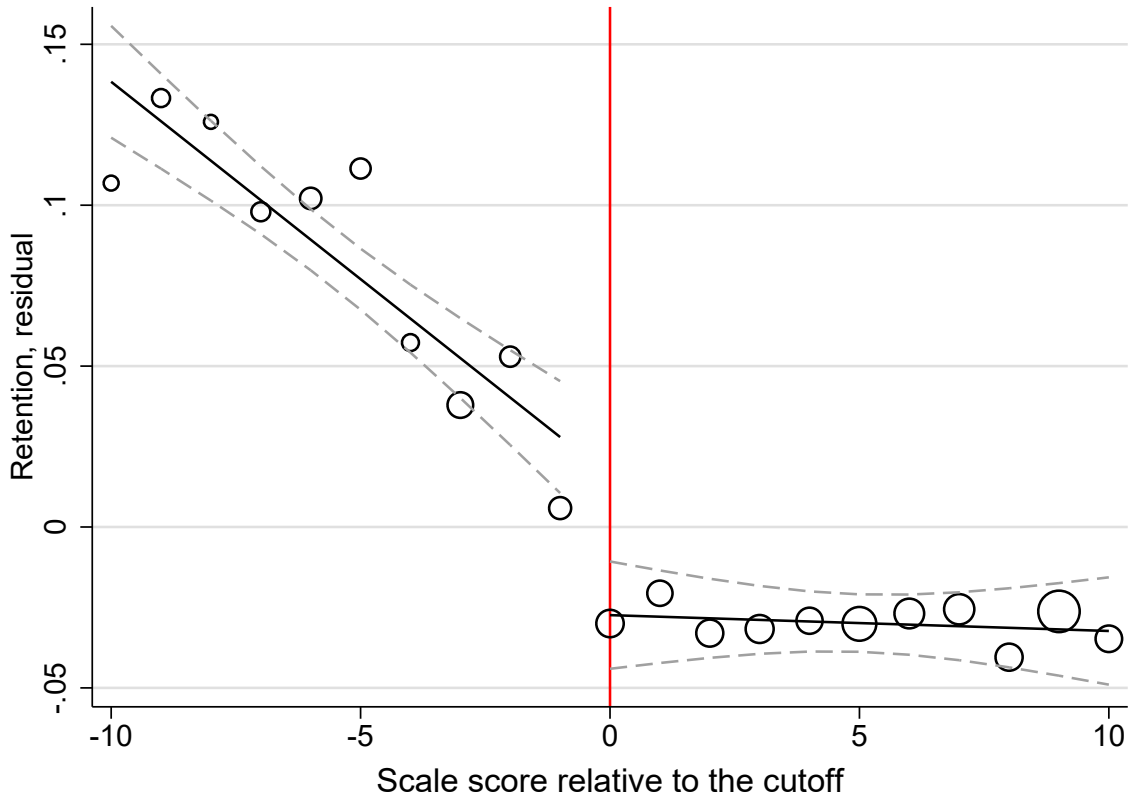
Raw data



- $\hat{\alpha_1} = 0.038^{***}$ (0.003)

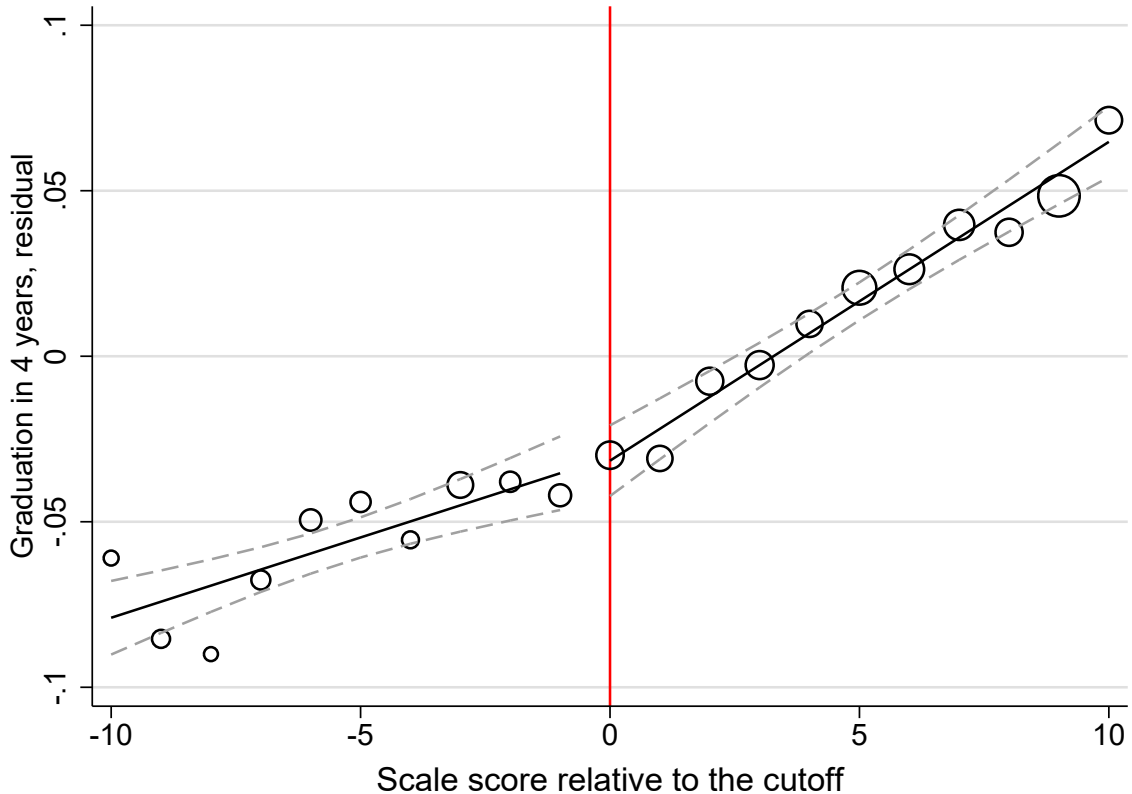
First Stage Effect of Exam Failure on Retention

Residual



- $\hat{\alpha_1} = 0.038^{***}$ (0.003)

Reduced Form: Graduated High School in 4 Years



- $\hat{\beta}_1 = 0.002$ (0.007)

IV Impact of Retention Using Naive RDD

	RD-IV Graduated from high school in 4 years
Retention	0.044 (0.172)
Observations	110432
Mean below cutoff	0.516
Mean above cutoff	0.610
F statistic	142.313

IV Impact of Retention Using Naive RDD

	RD-IV Graduated from high school in 4 years
Retention	0.044 (0.172)
Observations	110432
Mean below cutoff	0.516
Mean above cutoff	0.610
F statistic	142.313

- So why retain if no long-term benefit?

Brief Literature Review

- Jacob and Lefgren (2004): Short-term achievement
  - Examine the causal effect of remedial education programs in Chicago public schools (using RDD)
  - Positive effects of retention among 3rd graders; mixed impacts on 6th graders
- Mariano and Martorell (2013): New York City
  - Similar design, exploiting test score cutoffs used in assignment to summer school and retention
  - 2004-2008 data on 5th graders failing exam in 2004-2006
  - Document large and positive effects of grade retention on both Math and English
- Other settings: Florida (Schwerdt et al., 2017; Figlio and Özek, 2020); Indiana (Hwang and Koedel, 2023)
  - LiCalsi, Özek, and Figlio (2019) document that Florida's retention policy was enforced differently depending on maternal education

Short Run Impacts: Our Data

	A. RD-IV		B. RK-IV	
	Math next year	ELA next year	Math next year	ELA next year
Retention	29.373***	18.119***	39.829***	21.238***
	(8.222)	(4.393)	(5.133)	(2.749)
Observations	84168	84155	84168	84155
Mean below cutoff	656.164	642.697	656.164	642.697
Mean above cutoff	662.764	648.507	662.764	648.507
F statistic	133.825	135.627	280.011	280.181

Figures

Brief Literature Review

- Jacob and Lefgren (2009): High school completion
  - Examine the effect of retention on high school completion
  - Retaining 8th grade students substantially increases high school drop-out rates
  - Retention among 6th grade students does not have an effect on high school completion
- No effect on high school graduation in Florida (Schwerdt et al. 2017); increase in high school dropout and adult crime in Louisiana (Eren et al. 2017, Eren et al. 2022)
- Longer-term endpoints probably more important *per se* than shorter-term achievement
  - Do I care more about my child's performance in 4th grade or whether they completed high school?
  - Also don't have compositional effects: endogeneity of the specific test taken
- We focus on long-term outcomes such as high school quality and high school completion

Outline

Background
- Retention Policy

Data and Sample

Naive RDD

RK Design and Impacts
- Density of Running Variable

Impact Heterogeneity

First Stage Heterogeneity
- Compliant Subpopulation

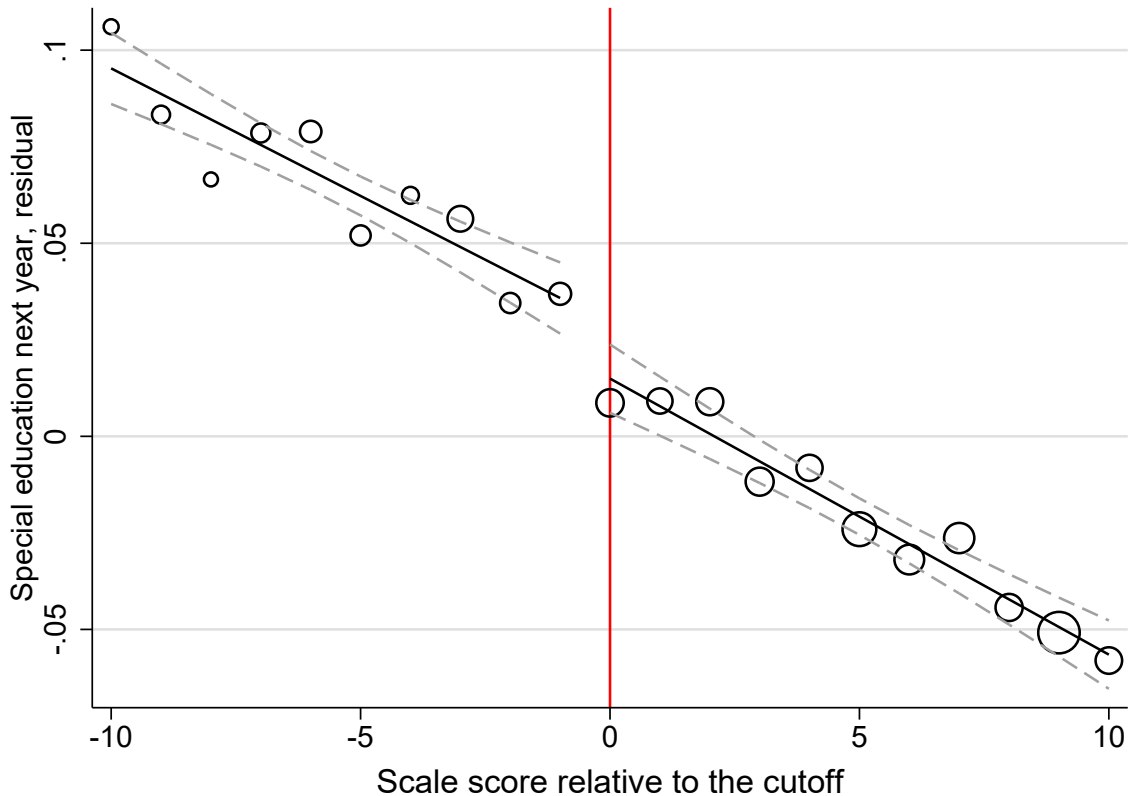
Conclusion

First Stage Effect of Exam Failure on Special Education

- Empirically, we find that exam failure affects the probability of receiving special education services

First Stage Effect of Exam Failure on Special Education

- Empirically, we find that exam failure affects the probability of receiving special education services



Special education in New York City

- Multiple steps are involved in the identification and placement of students in special education services:
  1. A parent or a Department of Education (DOE) school official must refer a child for special education services
  2. The DOE evaluates a child in all areas related to a suspected disability
  3. An Individualized Education Program (IEP) team (which includes a parent) reviews information from the evaluation and determines a child's eligibility for special education services
  4. If found eligible, the team develops an IEP, and a child is placed for special education services
- Our results suggest that exam failure may trigger a referral for an evaluation for special education, resulting in increased identification

First Stage Effect of Exam Failure on Special Education

Dependent Variable: **Special Education**

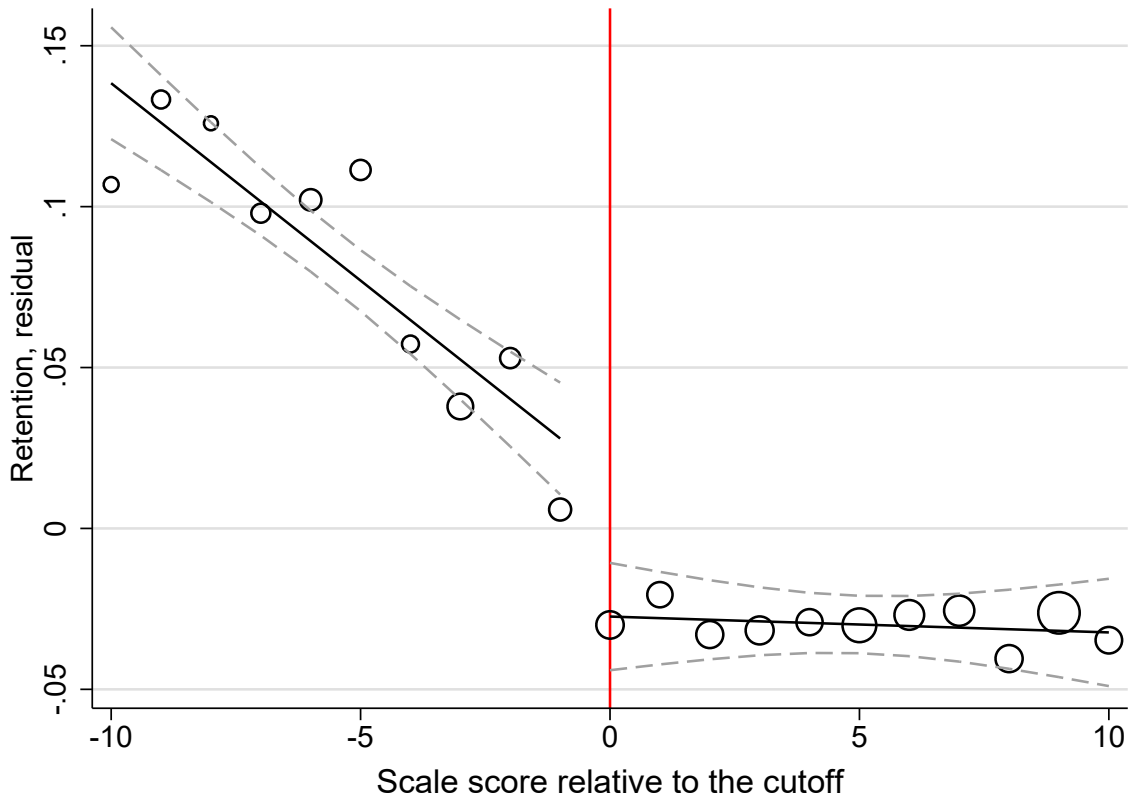
RHS:

Below cutoff	0.015**
	(0.006)
Running variable	-0.007***
	(0.000)
Below cutoff * running variable	0.000
	(0.001)
Observations	110413
Mean below cutoff	0.314
Mean above cutoff	0.219

Failure of Exclusion Restriction

- RDD for retention premised on retention being the only treatment that changed at threshold
- That special education also changes at discontinuity implies a failure of the exclusion restriction for IV
- We can only hope to get the combined effect of special education and retention from the RDD
- Reduced form RDD not of great policy interest
- Schools and policymakers want to know which decision to make/remedial service to provide
  - Retention viewed as particularly costly by parents and students

First Stage Effect of Exam Failure on Retention



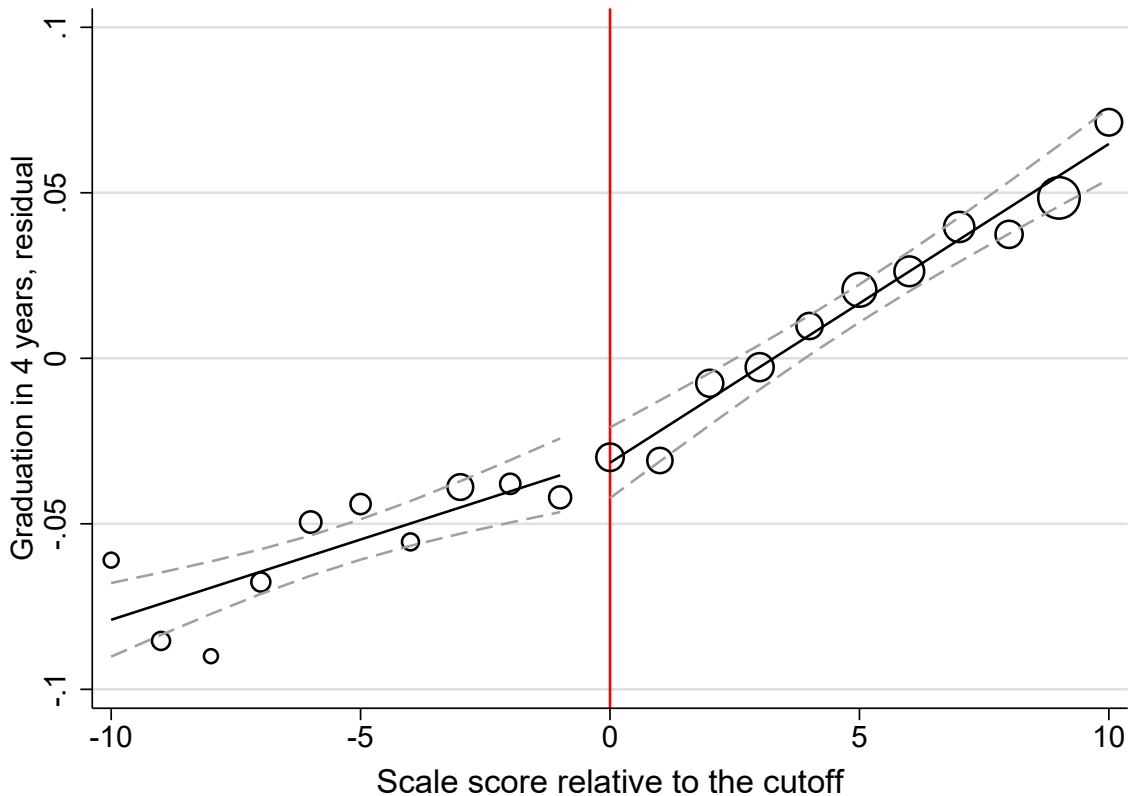
First Stage Effect of Exam Failure on Retention

Dependent Variable: **Retention**

RHS:

Below cutoff	0.038***
	(0.003)
Running variable	-0.000**
	(0.000)
Below cutoff * running variable	-0.014***
	(0.001)
Observations	110432
Mean below cutoff	0.116
Mean above cutoff	0.006

Reduced Form: Graduated High School in 4 Years



Reduced Form: Graduated High School in 4 Years

Dependent Variable: **Graduated High School in 4 Years**

RHS:

Below cutoff	0.002
	(0.007)
Running variable	0.010***
	(0.001)
Below cutoff * running variable	-0.004***
	(0.001)
Observations	110432
Mean below cutoff	0.516
Mean above cutoff	0.610

RK-Instrumental Variables Estimation

$$R_{igs,t+1} = \alpha_0 + \alpha_1 1[X_{igst} < 0] + \alpha_2 X_{igst} + \alpha_3 X_{igst} 1[X_{igst} < 0] + \theta_{gst} + \epsilon_{igst} \quad (3)$$

$$Y_{igs,t+x} = \beta_0 + \beta_1 1[X_{igst} < 0] + \beta_2 X_{igst} + \beta_3 X_{igst} 1[X_{igst} < 0] + \eta_{gst} + \nu_{igst} \quad (4)$$

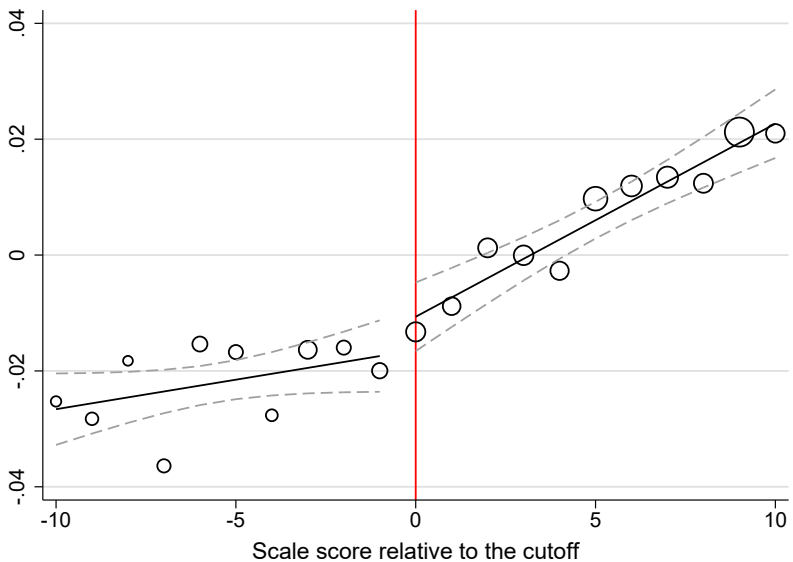
- i individual, g grade, s subject, t year
- $R_{igs,t+1} = 1[\text{retained} = 1]$
- $Y_{igs,t+x}$ indicates graduation outcomes
- $X_{igst} = \min[\text{Math scale score - cutoff, ELA scale score - cutoff}]$
- θ_{gst} and η_{gst}: year×grade×subject dummies (threshold fixed effects)
- Bandwidth: 10 scale scores below and above the threshold
- Robust standard errors
- $\frac{\hat{\beta}_3}{\hat{\alpha}_3}$ **represents the RK-IV estimate**

RK-Instrumental Variables Estimates

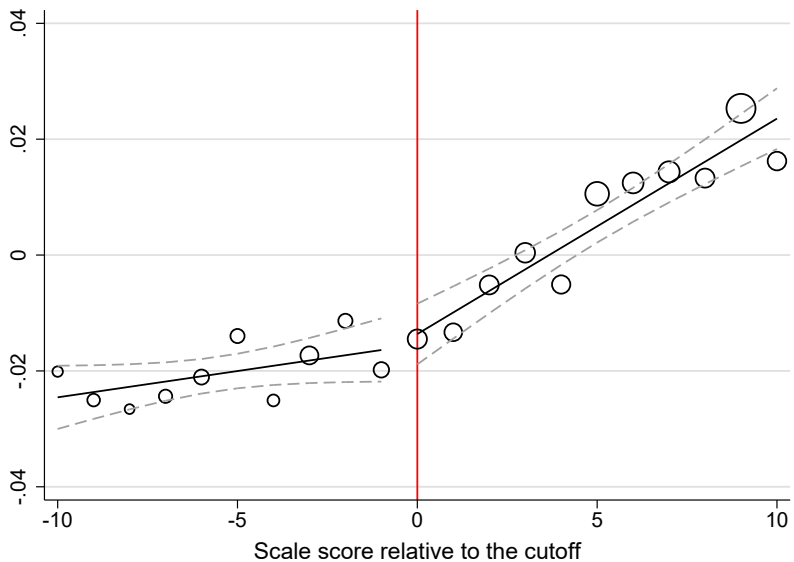
	RK-IV Graduated high school in 4 years
Retention	0.331***
	(0.087)
Observations	110432
Mean below cutoff	0.516
Mean above cutoff	0.610
F statistic	411.855

Reduced Form: Screened School Enrollment & Graduation

(a) Enrollment in a high school with screened admission in 9th grade



(b) Graduation from a high school with screened admission in 4 years



RK-Instrumental Variables Estimates: Screened High School Enrollment & Graduation

	RK-IV	
	Enrollment in a high school with screened admission in 9th grade	Graduation from a high school with screened admission in 4 years
Retention	0.158**	0.220***
	(0.079)	(0.078)
Observations	105553	110202
Mean below cutoff	0.281	0.252
Mean above cutoff	0.315	0.285
F statistic	418.174	413.777

Outline

Background
- Retention Policy

Data and Sample

Naive RDD

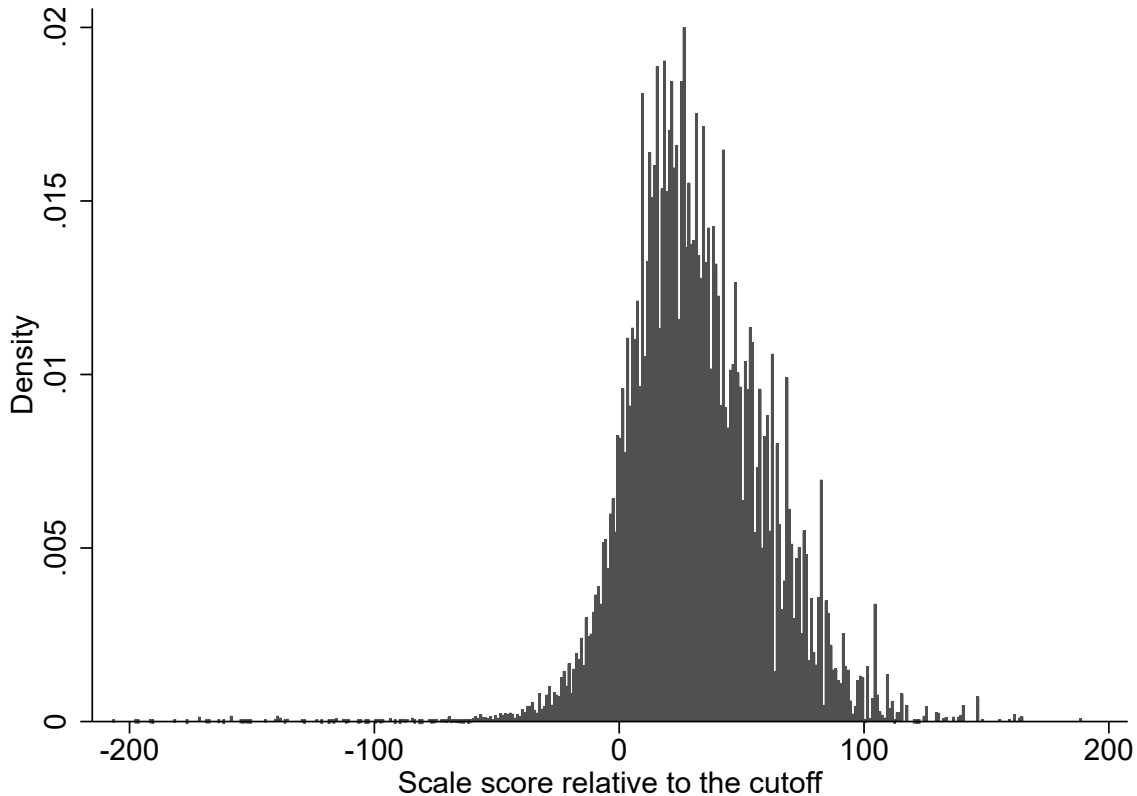
RK Design and Impacts
- Density of Running Variable

Impact Heterogeneity

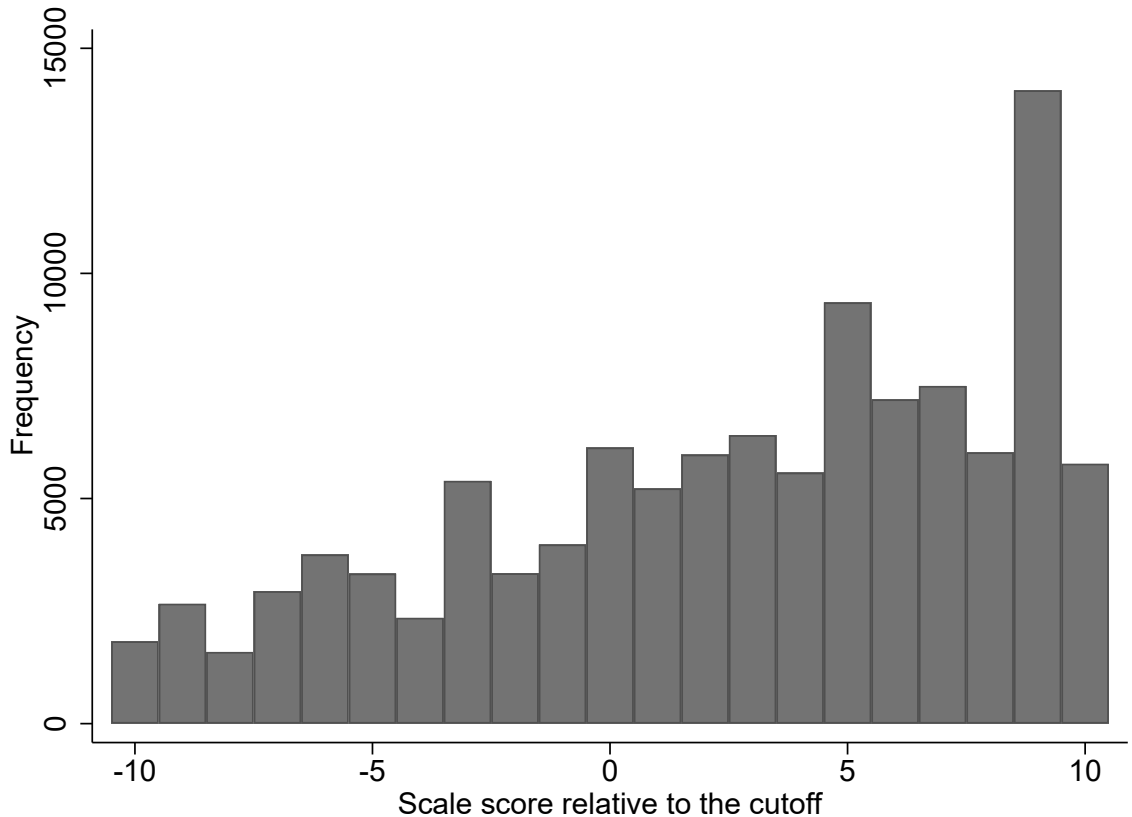
First Stage Heterogeneity
- Compliant Subpopulation

Conclusion

Sample Density by Running Variable

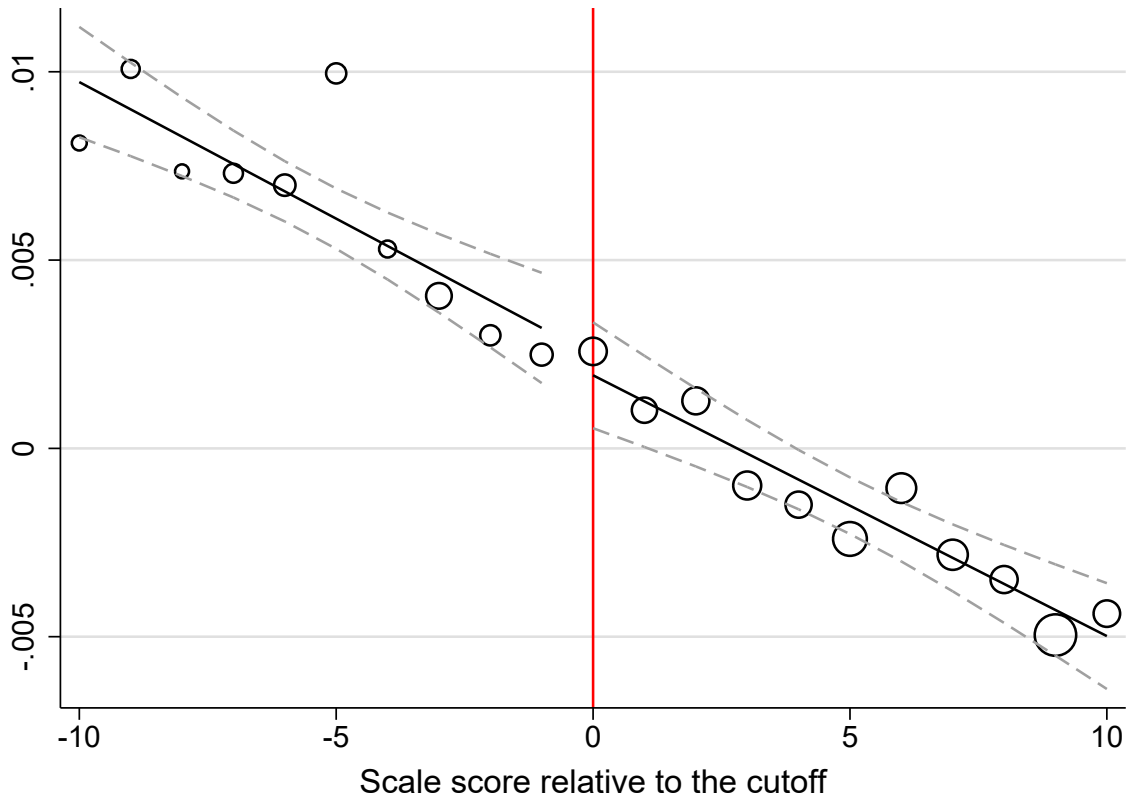


Sample Density by Running Variable



Balance of Baseline Covariates

More to point, no change in slope of predicted probability of retention:



Predicted probability of retention is calculated based on student gender, race/ethnicity, age in months, BMI, height, weight, free or reduced-price lunch eligibility, baseline special education participation, previous Math and English scale scores, and school fixed effects. Robust standard errors are used.

No Evidence of Differential Exit Below the Threshold

	Probability of having records	
	In the following year	For 4-year graduation outcomes
Below cutoff	0.000	0.005
	(0.001)	(0.005)
Running variable	0.000	0.001*
	(0.000)	(0.000)
Below cutoff * running variable	0.000	0.001
	(0.000)	(0.001)
Observations	137391	137391
Mean below cutoff	0.989	0.800
Mean above cutoff	0.992	0.805

Outline

Background
- Retention Policy

Data and Sample

Naive RDD

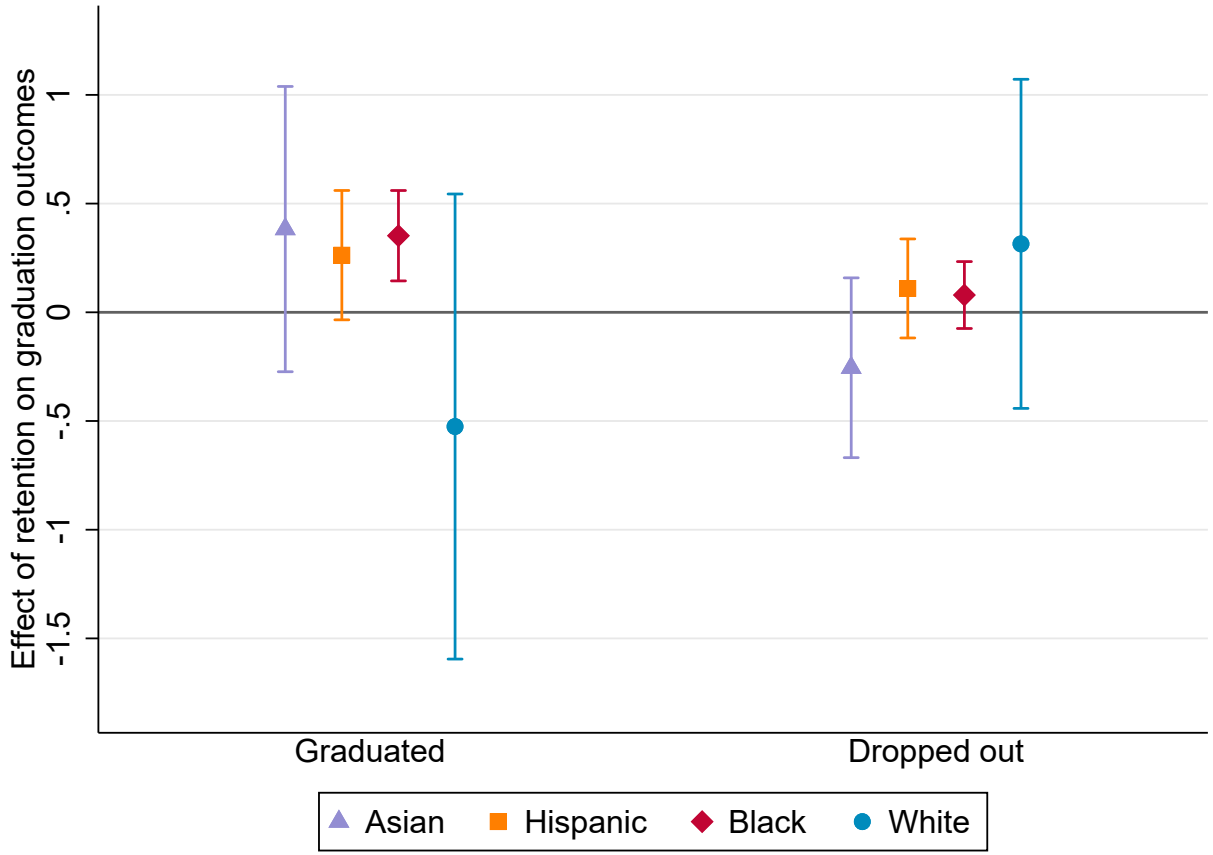
RK Design and Impacts
- Density of Running Variable

Impact Heterogeneity

First Stage Heterogeneity
- Compliant Subpopulation

Conclusion

Heterogeneity by Race



(Noting the confidence intervals)

Regression estimates for Black students

Other Dimentions

- We consider various student characteristics, such as gender, subsidized lunch eligibility, age for grade, height for grade, weight categories
- The positive effect of retention on graduation is large for girls than for boys, although the estimates are not statistically different Table
- The increased graduation effect is much more pronounced for students who are eligible for free or reduced-priced lunch Table
  - Note, however, that most students in our sample are eligible for subsidized lunch
- Little heterogeneity across student age and physical characteristics
- Overall, students with relatively disadvantaged backgrounds may be more likely to benefit from grade retention in a long run

Heterogeneity by grade

Outline

Background
- Retention Policy

Data and Sample

Naive RDD

RK Design and Impacts
- Density of Running Variable

Impact Heterogeneity

First Stage Heterogeneity
- Compliant Subpopulation

Conclusion

First Stage: Retention as Outcome

Note: Even in subgroups, the relationship between test score and retention probability becomes significantly steeper below the threshold!

		A. Gender		B. Race/ethnicity			
	All	Female	Male	Asian	Hispanic	Black	White
Below cutoff	0.038***	0.047***	0.031***	0.016*	0.037***	0.045***	0.026**
	(0.003)	(0.005)	(0.004)	(0.009)	(0.004)	(0.006)	(0.011)
Running variable	-0.000**	-0.000*	-0.000	-0.000	-0.000	-0.000**	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Below cutoff	-0.014***	-0.013***	-0.014***	-0.011***	-0.011***	-0.019***	-0.009***
* running variable	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)
Observations	110432	51371	59059	9696	52177	41293	7245
Mean below cutoff	0.116	0.122	0.110	0.075	0.103	0.148	0.073
Mean above cutoff	0.006	0.005	0.007	0.003	0.006	0.007	0.006

First Stage: Retention as Outcome

Note: Even in subgroups, the relationship between test score and retention probability becomes significantly steeper below the threshold!

		A. Gender		B. Race/ethnicity			
	All	Female	Male	Asian	Hispanic	Black	White
Below cutoff	0.038***	0.047***	0.031***	0.016*	0.037***	0.045***	0.026**
	(0.003)	(0.005)	(0.004)	(0.009)	(0.004)	(0.006)	(0.011)
Running variable	-0.000**	-0.000*	-0.000	-0.000	-0.000	-0.000**	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Below cutoff	-0.014***	-0.013***	-0.014***	-0.011***	-0.011***	-0.019***	-0.009***
* running variable	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)
Observations	110432	51371	59059	9696	52177	41293	7245
Mean below cutoff	0.116	0.122	0.110	0.075	0.103	0.148	0.073
Mean above cutoff	0.006	0.005	0.007	0.003	0.006	0.007	0.006

First Stage: Retention as Outcome

Note: Even in subgroups, the relationship between test score and retention probability becomes significantly steeper below the threshold!

		A. Gender		B. Race/ethnicity			
	All	Female	Male	Asian	Hispanic	Black	White
Below cutoff	0.038***	0.047***	0.031***	0.016*	0.037***	0.045***	0.026**
	(0.003)	(0.005)	(0.004)	(0.009)	(0.004)	(0.006)	(0.011)
Running variable	-0.000**	-0.000*	-0.000	-0.000	-0.000	-0.000**	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Below cutoff	-0.014***	-0.013***	-0.014***	-0.011***	-0.011***	-0.019***	-0.009***
* running variable	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)
Observations	110432	51371	59059	9696	52177	41293	7245
Mean below cutoff	0.116	0.122	0.110	0.075	0.103	0.148	0.073
Mean above cutoff	0.006	0.005	0.007	0.003	0.006	0.007	0.006

First Stage: Special Education as Outcome

Note: Even in subgroups, there is no kink in the probability of special education placement!

		A. Gender		B. Race/ethnicity			
	All	Female	Male	Asian	Hispanic	Black	White
Below cutoff	0.015**	0.001	0.029***	-0.019	0.018**	0.016*	0.024
	(0.006)	(0.008)	(0.008)	(0.018)	(0.009)	(0.010)	(0.027)
Running variable	-0.007***	-0.007***	-0.007***	-0.007***	-0.008***	-0.007***	-0.010***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Below cutoff	0.000	-0.001	0.002	0.005	0.002	-0.003	0.005
* running variable	(0.001)	(0.001)	(0.001)	(0.003)	(0.002)	(0.002)	(0.005)
Observations	110413	51356	59055	9694	52170	41286	7242
Mean below cutoff	0.314	0.259	0.361	0.194	0.325	0.307	0.467
Mean above cutoff	0.219	0.179	0.255	0.165	0.228	0.198	0.338

Other heterogeneity

First Stage: Special Education as Outcome

Note: Even in subgroups, there is no kink in the probability of special education placement!

		A. Gender		B. Race/ethnicity			
	All	Female	Male	Asian	Hispanic	Black	White
Below cutoff	0.015**	0.001	0.029***	-0.019	0.018**	0.016*	0.024
	(0.006)	(0.008)	(0.008)	(0.018)	(0.009)	(0.010)	(0.027)
Running variable	-0.007***	-0.007***	-0.007***	-0.007***	-0.008***	-0.007***	-0.010***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Below cutoff	0.000	-0.001	0.002	0.005	0.002	-0.003	0.005
* running variable	(0.001)	(0.001)	(0.001)	(0.003)	(0.002)	(0.002)	(0.005)
Observations	110413	51356	59055	9694	52170	41286	7242
Mean below cutoff	0.314	0.259	0.361	0.194	0.325	0.307	0.467
Mean above cutoff	0.219	0.179	0.255	0.165	0.228	0.198	0.338

Other heterogeneity

First Stage: Special Education as Outcome

Note: Even in subgroups, there is no kink in the probability of special education placement!

		A. Gender		B. Race/ethnicity			
	All	Female	Male	Asian	Hispanic	Black	White
Below cutoff	0.015**	0.001	0.029***	-0.019	0.018**	0.016*	0.024
	(0.006)	(0.008)	(0.008)	(0.018)	(0.009)	(0.010)	(0.027)
Running variable	-0.007***	-0.007***	-0.007***	-0.007***	-0.008***	-0.007***	-0.010***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Below cutoff	0.000	-0.001	0.002	0.005	0.002	-0.003	0.005
* running variable	(0.001)	(0.001)	(0.001)	(0.003)	(0.002)	(0.002)	(0.005)
Observations	110413	51356	59055	9694	52170	41286	7242
Mean below cutoff	0.314	0.259	0.361	0.194	0.325	0.307	0.467
Mean above cutoff	0.219	0.179	0.255	0.165	0.228	0.198	0.338

Other heterogeneity

Outline

Background
- Retention Policy

Data and Sample

Naive RDD

RK Design and Impacts
- Density of Running Variable

Impact Heterogeneity

First Stage Heterogeneity
- Compliant Subpopulation

Conclusion

Complier Characteristics – Retention

- Estimate compliers' *average* characteristics (Angrist & Pischke, 2009)
  - Empirically: ratio of magnitudes of first stage by subgroup

Complier Characteristics – Retention

- Estimate compliers' *average* characteristics (Angrist & Pischke, 2009)
  - Empirically: ratio of magnitudes of first stage by subgroup

Characteristic	Analysis sample mean	RDD complier ratio	RKD complier ratio
Female	0.465	1.219	0.988
Asian	0.088	0.404	0.779
Hispanic	0.472	0.973	0.818
Black	0.374	1.170	1.368
White	0.066	0.670	0.640
Subsidized lunch	0.940	1.040	1.002
Young for grade	0.334	1.111	0.995
Old for grade	0.320	0.958	0.933
Short for grade	0.300	1.065	0.918
Tall for grade	0.382	0.860	1.049
Underweight	0.033	0.962	1.008
Healthy	0.523	0.951	1.017
Overweight	0.444	1.064	0.979
Obese	0.251	1.180	0.924

Complier Characteristics – Retention

- Estimate compliers' *average* characteristics (Angrist & Pischke, 2009)
  - Empirically: ratio of magnitudes of first stage by subgroup

Characteristic	Analysis sample mean	RDD complier ratio	RKD complier ratio
Female	0.465	1.219	0.988
Asian	0.088	0.404	0.779
Hispanic	0.472	0.973	0.818
Black	0.374	1.170	1.368
White	0.066	0.670	0.640
Subsidized lunch	0.940	1.040	1.002
Young for grade	0.334	1.111	0.995
Old for grade	0.320	0.958	0.933
Short for grade	0.300	1.065	0.918
Tall for grade	0.382	0.860	1.049
Underweight	0.033	0.962	1.008
Healthy	0.523	0.951	1.017
Overweight	0.444	1.064	0.979
Obese	0.251	1.180	0.924

Complier Characteristics – Retention

- Estimate compliers' *average* characteristics (Angrist & Pischke, 2009)
  - Empirically: ratio of magnitudes of first stage by subgroup

Characteristic	Analysis sample mean	RDD complier ratio	RKD complier ratio
Female	0.465	1.219	0.988
Asian	0.088	0.404	0.779
Hispanic	0.472	0.973	0.818
Black	0.374	1.170	1.368
White	0.066	0.670	0.640
Subsidized lunch	0.940	1.040	1.002
Young for grade	0.334	1.111	0.995
Old for grade	0.320	0.958	0.933
Short for grade	0.300	1.065	0.918
Tall for grade	0.382	0.860	1.049
Underweight	0.033	0.962	1.008
Healthy	0.523	0.951	1.017
Overweight	0.444	1.064	0.979
Obese	0.251	1.180	0.924

Outline

Background
- Retention Policy

Data and Sample

Naive RDD

RK Design and Impacts
- Density of Running Variable

Impact Heterogeneity

First Stage Heterogeneity
- Compliant Subpopulation

Conclusion

Summary and Conclusion

- Roughly 50,000 students are "held back" in New York City each year; In the United States as a whole, around 2% or 1 million students are retained each year
- Little evidence on the long-term consequences of grade retention
  - We argue that progression in high school is more important to students and parents than shorter-term outcomes
- In New York, exam failure may also trigger referral to special education
- Conveniently, special education decisions ignore the exam score | failure
- In contrast, retention decisions start to pay attention to score when falls below failure threshold
- This allows us to isolate impact of retention using an RK design

Summary and Conclusion

- Leveraging this shock in a RK design, we find retention increases the likelihood of completing high school
- The magnitude of this impact is large, increasing high school completion by *33* percentage points!
  - Roughly equivalent to scoring average
- Minority students drive this large benefit
  1. Stronger first stage
  2. Larger IV estimate
- Boys and white students tend to avoid retention and are more likely to be classified as special education in response to poor exam performance
- Two contributions
  1. We show the importance of accounting for double treatments
  2. We find large, positive impacts on both high school quality and high school completion