The Effect of Test Score Performance Labels on Post-Secondary Educational Outcomes: Evidence from Michigan[1]

By Joshua Brownstein

Updated: June 2023

Abstract

Standardized test scores and the labels associated with those scores provide students and their parents highly credible information about a student's academic achievement. This information could cause students and their parents to change their beliefs regarding a student's academic ability. This may then change a student's future educational choices and thus their future educational outcomes. In this paper I use administrative data on Michigan students to look at the impact of receiving different labels summarizing a student's performance on standardized tests on a student's post-secondary educational outcomes. I use a regression discontinuity research design to compare students who have similar test scores but who receive different summary labels. While some of my estimates are significant, almost all lack of robustness to using another bandwidth and I am likely to find some spurious effects given the large number of estimates in this paper. I conclude that I do not find evidence of a large effect of performance labels on postsecondary outcomes.

[1] Thank you to Scott Imberman, Ajin Lee, Leslie Papke, and Kris Renn for providing me feedback on this project as members of my dissertation committee. Thank you to the Michigan Education Data Center (MEDC) for providing me with the data for this project. Thank you to Michigan Education Data Center employees Jasmina Camo-Biogradlija and Kyle Kwaiser for their repeated communications with me related to applying for and getting access to MEDC data for this project. This research result used data structed and maintained by the MERI-Michigan Education Data Center (MEDC). MEDC data is modified for analysis purposes using rules governed by MEDC and are not identical to those data collected and maintained by the Michigan Department of Education (MDE) and/or Michigan's Center for Educational Performance and Information (CEPI). Results, information and opinions solely represent the analysis, information, and opinions of the author(s) and are not endorsed by, or reflect the views or positions of, grantors, MDE and CEPI or any employee thereof.

I. Introduction

Governments throughout the world administer standardized exams to their students to learn about their academic achievement (Schleicher 2015). Much of the literature on standardized testing has focused on how schools (Figlio and Loeb 2011; Figlio and Ladd 2015) and teachers (Donaldson and Papay 2015) react to the test scores especially when performance on these exams leads to rewards or sanctions. A less studied aspect of standardized testing is how testing provides knowledge about a student's academic achievement to the student and their parents. This information may be a significantly more credible signal of a student's academic achievement than grades given the wide variation in grading practices among schools and teachers[2]. If parents and students make decisions about post-secondary education based on beliefs about the student's academic ability, and if how standardized test scores are described changes their beliefs, then which label a student receives will change which post-secondary education choices students make. By changing a student's education choices these labels may then change a student's postsecondary outcomes.

This chapter looks at the causal effect of getting different labels on standardized tests on post-secondary outcomes using administrative data on students in Michigan. I use a regression discontinuity research design to look at students who receive similar exam scores but different labels summarizing those scores. I look at students who are close to the cutoffs of receiving either the label associated with the highest or lowest scores for an 11^{th} grade math and reading exam. While some of my estimates are statistically significant, almost all lack robustness to using another bandwidth. Also, if no labels had any effect on postsecondary outcomes, I would be likely to find some statistically significant effects anyway given the large number of estimates in this chapter. I conclude that I do not find evidence of a large effect of performance labels on postsecondary outcomes.

II. Literature Review

The study most similar to my study is Papay, Murnane, and Willett (2016). They use Massachusetts administrative data to study the effect of test score labels in grades 8 and 10 on post-secondary enrollment. They focus their analysis on students who get free and reduced-price lunches and who live in urban school districts. Among those students near the cutoffs, being

[2] See Gershenson (2018) for evidence of differential grade inflation by the affluence of students in North Carolina. See Pattison, Grodsky, and Muller (2013) for evidence that while grades have risen over time, the signaling power of grades as measured by the variance of grades and predictive power of grades has not decreased over time.

labeled Advanced rather than Proficient on the 10th grade math exam causes a 5-percentage point increase in post-secondary attendance within a year after intended high school graduation. This effect is greatest among students who when surveyed before taking the 10th grade exam reported they did not plan on attending a 4-year college.

Two other papers study the effect of performance labels using regression discontinuity research designs on K – 12 outcomes. Avery and Goodman (2021) study the causal effect of receiving an Advanced label on a 10th grade math test on the probability of taking an Advanced Placement Calculus course for Massachusetts students. They find that for Black and Hispanic students, getting the Advanced label increases the probability a student will take an Advanced Placement Calculus course by 2.5 percentage points. Beuchert, Eriksen, and Krægpøth (2020) study the effect of 3rd grade test score labels[3] for children in Denmark. Pooling the results of different labeling cutoffs, they find that getting a label associated with a lower score on the 3rd grade math exam causes a 6% of a standard deviation increase in scores on the 6th grade math exam.

There is one study that looks at the effect of students being informed about their academic ability on a mock standardized exam on high school outcomes. It provides evidence that information about academic ability can change academic outcomes by changing a student's choice about where to go to school. Bobba and Frisancho (2019) study the effect of providing information about academic ability on the secondary schooling choices of students from high poverty neighborhoods in Mexico City. Schools are randomly assigned one of three treatments: no intervention, a mock secondary school admissions exam without informing students of their scores, or a mock secondary school admissions exam with informing students of their scores. They find the combination of the mock exam and the information about the exam score made high scoring students more likely to go to academic (college prep) schools and low scoring students more likely to go to non-academic (vocational/technical) schools. This new sorting of students to schools led to an increase in the on time high school graduation rate for students in the exam and information group of 8 percentage points compared to the no intervention group.

Papay, Murnane, and Willett (2016), Beuchert, Eriksen, and Krægpøth (2020), and Avery

[3] In Denmark scores receive one of the following 5 labels ordered from the lowest scoring exams to the highest scoring exams: Considerably Below Average, Below Average, Average, Above Average, and Considerably Above Average. The effect size around the cutoff between Considerably Below Average and Below Average is greater than the effect size at any of the other cutoffs.

and Goodman (2021) use regression discontinuity research designs to look at the effect of test score labels on the future educational outcomes of tested students. In all those papers, the labels did not carry any consequences in terms of things like the ability to graduate or the ability to take certain classes. Also, in those papers parents are sent reports about their child's test performance that include the label that corresponds to their child's score. This is like the institutional setting for this paper. In the case of the Danish score report, parents are not given information about the underlying scale score that determines the label. This is different from the reports in Massachusetts and Michigan that show parents the underlying scale score in the report.

My study builds on those studies in several ways. It is the first study to look at the effect of test score labels in Michigan. Because of Michigan's large population[4], this study can detect smaller effects than the prior literature. Compared to Papay, Murnane, and Willett (2016) and Avery and Goodman (2021) this paper studies an exam taken in 11th grade rather than an exam taken in 10th grade. The closer the information is received relative to high school graduation, the more impact it might have under the assumption that events closer in time to the measured outcome have a greater effect on that outcome than events further away in time from it.

III. Institutional Setting

In 2002 the No Child Left Behind Act was passed[5] . The act required all U.S. states to administer standardized exams to students in math and reading in grades 3 through 8 and once in high school[6].

From the 2007 - 2008 school year to the 2013 – 2014 school year 11th grade students in Michigan are required to take standardized exams in Math, Reading, Science, Social Studies, and Writing as part of the Michigan Merit Exam. For each exam each student is assigned a scale score to indicate how well they did on the exam. Students who have higher scale scores did better on the exam generally by answering a higher proportion of the exam's multiple-choice questions correctly.

[4] The population of Denmark in Q1 2020 was 5,822,763. The estimated population of Massachusetts in 2019 was 6,892,503. The estimated population of Michigan in 2019 was 9,986,857. See https://www.dst.dk/en/Statistik/emner/befolkning-og-valg/befolkning-og-befolkningsfremskrivning/folketal for Denmark population and https://www.census.gov/data/tables/time-series/demo/popest/2010s-state-total.html for the population of Massachusetts and Michigan.

[5] https://en.wikipedia.org/wiki/No_Child_Left_Behind_Act

[6] In 2015 President Obama signed the Every Student Succeeds Act. While this law officially repealed the No Child Left Behind Act, it has its own set of requirements to test students in grades 3 to 8 and once in high school. See https://www.edweek.org/policy-politics/the-every-student-succeeds-act-an-essa-overview/2016/03 for more information.

Students' performance on each exam is summarized by a performance label. Which label a student receives is based on their scale score. The performance labels from lowest scores to highest scores are: Not Proficient, Partially Proficient, Proficient, and Advanced. Scale scores are mapped to performance labels based on scores being in non-overlapping intervals. This means that for any given year all students who receive a lower performance label, such as Not Proficient, have lower scale scores than all students who receive a higher performance label, such as Partially Proficient.

IV. Data and Sample

The data for this project comes from the Michigan Education Data Center (MEDC). MEDC houses student level data for all K - 12 students who attend public schools in Michigan including data on test scores and demographic information such as a student's race and gender. It also has data on post-secondary enrollment and degree completion from the National Student Clearinghouse.

This chapter uses data on all students in Michigan who have 11th grade test scores from the 2007 – 2008 school year to the 2013 – 2014 school year[7]. Students whose data on their race, their gender, or if they are economically disadvantaged are missing are not included in my sample[8]. I construct 4 samples to look at students near the cutoffs between the following pairs of performance labels: Proficient and Advanced on the math exam, Proficient and Advanced on the reading exam, Not Proficient and Partially Proficient on the math exam, and Not Proficient and Partially Proficient on the reading exam. For each sample I only include students who receive one of the performance labels in the sample's name. For example, the Math Proficient/Advanced sample only includes students who receive a Proficient or Advanced label on their 11th grade math exam.

[7] The 2007 – 2008 school year is the earliest year that the Michigan Education Data Center has test score data for. I choose my last year to be 2013 – 2014 so I could analyze similar exam data across years. Starting in Spring 2015 Michigan made large changes to its 11th grade standardized exams changing from the Michigan Merit Exam to the Michigan Student Test of Education Progress. See https://medc.miedresearch.org/dataset/k-12-student-assessments.

[8] I start with a sample of 803,798 students who were in 11th grade from school year 2007 -2008 to school year 2013 – 2014. Of those students 87,609 are missing data on their race and gender, 28,253 are missing data on if they are economically disadvantaged, and 1,054 are missing data on their reading and math scores. Some of those groups of students overlap. Once all students with missing data are removed, I have a sample of 716,694 students.

Table 1 - Summary Statistics

Variable	Math Proficient/Advanced	Reading Proficient/Advanced	Math Not Proficient/Partially Proficient	Reading Not Proficient/Partially Proficient
Female Indicator	0.46	0.53	0.52	0.46
	(0.50)	(0.50)	(0.50)	(0.50)
White Indicator	0.88	0.85	0.72	0.67
	(0.32)	(0.36)	(0.45)	(0.47)
Black Indicator	0.03	0.07	0.19	0.24
	(0.16)	(0.26)	(0.39)	(0.43)
Hispanic Indicator	0.02	0.03	0.04	0.05
	(0.13)	(0.16)	(0.20)	(0.21)
Asian Indicator	0.05	0.03	0.02	0.02
	(0.22)	(0.17)	(0.13)	(0.14)
Two or More Races Indicator	0.01	0.02	0.02	0.02
	(0.12)	(0.13)	(0.14)	(0.14)
Native American Indicator	0.00	0.01	0.01	0.01
	(0.07)	(0.08)	(0.09)	(0.09)
Hawaiian Indicator	0.00	0.00	0.00	0.00
	(0.03)	(0.03)	(0.03)	(0.03)
Economically Disadvantaged Indicator	0.15	0.24	0.42	0.47
	(0.36)	(0.43)	(0.49)	(0.50)
N	198,116	393,103	508,594	317,140

Notes: The table shows the mean outcome for each sample above the standard deviation for that outcome in parentheses.

Table 1 shows summary statistics for the 4 samples. The main differences between the samples are between the Advanced/Proficient samples and between the Not Proficient/Partially Proficient samples. A higher proportion of the Not Proficient/Partially Proficient samples are black and economically disadvantaged. A lower proportion of the Not Proficient/Partially Proficient samples are white. The differences are smaller for proportion female and proportion of the other races in the data.

V. Empirical Framework

My goal in this paper is to look at how receiving different performance labels changes a student's post-secondary outcomes. I do this by using a sharp regression discontinuity research design to compare the outcomes of students near cutoffs to receive different performance labels. By doing this, for students close to a cutoff, I can estimate the average treatment effect of a student receiving a label associated with higher scale scores compared to receiving the label associated with lower scale scores on the other side of the cutoff.

For my main results I use the following estimating equation.

(1) $Outcome_{isy} = \beta_0 + \beta_1 HigherLabel_{isy} + \beta_2(ScaleScore_{isy} - Cutoff_s) + \beta_3 HigherLabel_{isy}(ScaleScore_{isy} - Cutoff_s) + \theta_y + \boldsymbol{\beta X}_i + \epsilon_{isy}$

In the equation individual i takes exam subject s in school year y. $HigherLabel_{isy}$ is an indicator variable for the student receiving the performance label in the sample associated with higher scales scores (either Advanced or Partially Proficient). $Cutoff_s$ is the lowest scale score a student needs to receive the higher label. In my sample, cutoffs vary based on the subject (math or reading) of the exam and which labels are on either side of the cutoff. However, the cutoffs do not vary depending on the year of the exam. θ_y is a fixed effect for the year the exam was taken. $\mathbf{X_i}$ are covariates. Covariates are indicator variables for a student's race, gender, and if they are economically disadvantaged. I cluster standard errors at the year of exam level.

The equation assumes a linear relationship between the outcome variable and the scale score of the exam allowing for the slope of the line to vary on either side of the cutoffs. The coefficient of interest is β_1 which is the average outcome for students near the cutoffs if they get the label associated with higher scale scores minus the counterfactual average outcome of those students if they got the label on the other side of the cutoff associated with receiving lower scale scores. I refer to this as the treatment effect of receiving the higher label.

For each regression I limit my sample to students whose scale scores are within a certain number of points of the cutoff. This value is called bandwidth. I choose a bandwidth for each sample based on the following procedure. First, for a given sample, I calculate the mean squared optimal bandwidth for each of my 6 outcome variables (ever enrolling in any post-secondary institution, ever enrolling in a 2-year institution, ever enrolling in a 4-year institution, having any post-secondary degree, having an associate degree, having a bachelor's degree) using the method in Calonico, Cattaneo, and Titiunik (2014). Bandwidths are chosen using a uniform kernel accounting for indicators related to a student's race, a student's gender, and if the student is economically disadvantaged being in the regression. The bandwidth I use for each sample is the average of the 6 calculated bandwidths rounded to the nearest whole number. As a robustness check, I redo my analysis using bandwidths that are 0.5 times and 1.5 times the value of the chosen bandwidths rounded to the nearest whole number. The results using these other bandwidths are presented in Appendix A and discussed in Section 3.10.

For β_1 to be the treatment effect of receiving the higher label, it must be the case that a

student having a scale score be close to and above a cutoff or close to and below a cutoff is as good as random. In that case students near the cutoff will have, on average, the same observable and unobservable characteristics. This may not be the case if, for example, students or the individuals who assign students scale scores precisely manipulate the scores so students receive a specific performance label. In this case I would not only be measuring the treatment effect of receiving a higher label, but the willingness or ability to manipulate scores to be above or below a cutoff. To check for this, I use a modified version of Equation 1 where the outcome is an indicator variable for a student being female, being a certain race, or being economically disadvantaged and other covariates are excluded from the regression. In those regressions β_1 is the discontinuity in the proportion of students with that characteristic at the cutoff. A significant coefficient would be evidence that the traits of students change suddenly at the cutoff and would be consistent with scores being manipulated to get a specific performance label.

To look at heterogeneity, I use the following estimating equation.

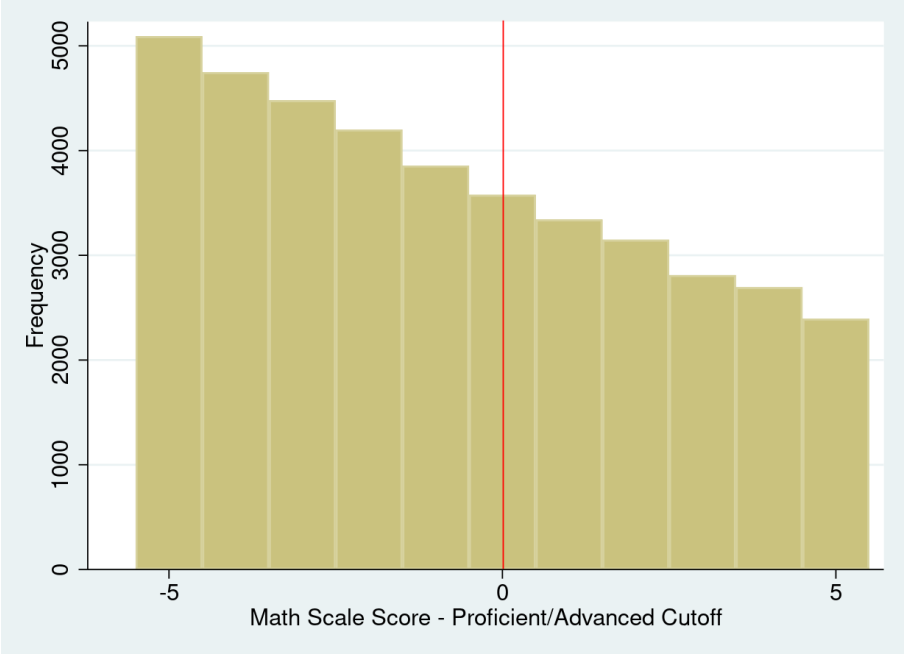
(2) $$Outcome_{igsy} = \beta_0 + \beta_1 HigherLabel_{igsy} + \beta_2 HigherLabel_{igsy} Subgroup_g + \beta_3 Subgroup_g + \beta_4 (ScaleScore_{igsy} - Cutoff_s) + \beta_5 HigherLabel_{igsy} (ScaleScore_{igsy} - Cutoff_s) + \beta_6 Subgroup_g (ScaleScore_{igsy} - Cutoff_s) + \beta_7 HigherLabel_{igsy} Subgroup_g (ScaleScore_{igsy} - Cutoff_s) + \theta_y + \epsilon_{igsy}$$

In Equation 2 I look at student i in subgroup g whose exam score is in subject s for exam taken in school year y. $Subgroup_g$ equals 1 if a student is a member of the subgroup and 0 otherwise. I look at 3 different subgroups: female students, black students, and economically disadvantaged students. In the regressions where the subgroup is black students only students who are either white or black are included in the regressions. Equation 2 assumes a linear relationship between the outcome variable and the scale score whose slope can be different both above and below the cutoff and for students who are and are not members of the subgroup. The estimates of interest are β_1, which is the higher label treatment effect for students who are not members of the subgroup, and $\beta_1 + \beta_2$, which is the higher label treatment effect for students who are members of the subgroup. β_2 is the difference between the two treatment effects.

VI. Math Proficient/Advanced Cutoff Results

A. Identification Test: Discontinuity in Density

Figure 1 – Histogram Students Close to the Cutoff Math Proficient/Advanced Sample



Notes: N = 40,349. Each bar in this histogram shows the number of students in the sample who received a different scale score.

Figure 1 shows the number of students in the sample who receive different scale scores for values of the scale score close to the cutoff. A sudden change in the number of students at 0 would be consistent with scores being manipulated so the student receives a different performance label. Based on Figure 1, I do not find evidence of this as the change in the number of observations is smooth at the cutoff.

B. Identification Test: Discontinuity in Covariates

Table 2 – Discontinuity in Covariates Math Proficient/Advanced Sample

	Female	White	Black	Hispanic
Advanced Label	-0.0016	-0.0042	0.0016	0.0026
	(0.0138)	(0.0048)	(0.0022)	(0.0036)
Year Fixed Effects	Y	Y	Y	Y
Mean Outcome	0.46	0.88	0.03	0.02
	Asian	Two or More Races	Native American	Hawaiian
Advanced Label	0.0002	0.0004	-0.0004	-0.0003
	(0.0023)	(0.0019)	(0.0013)	(0.0005)
Year Fixed Effects	Y	Y	Y	
Mean Outcomes	0.05	0.01	0.00	0.00
	Economically Disadvantaged			
Advanced Label	-0.0013			
	(0.0085)			
Year Fixed Effects	Y			
Mean Outcomes	0.15			

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. N = 40,349. Bandwidth = 5 scale score points. Standard errors are clustered at the year level. The outcomes are indicator variables for being female, being a specific race or being economically disadvantaged. Mean outcomes for the Math Proficient/Advanced Sample are shown.

In Table 2 I estimate discontinuities in the proportion of students who have different observable characteristics at the Proficient/Advanced cutoffs. I find that all discontinuities are small and statistically insignificant. Based on this and on Figure 1 I conclude there is no manipulation of scale scores and my estimates are treatment effects of receiving an Advanced label on the 11th grade math exam.

C. Higher Label Treatment Effect Math Proficient/Advanced Sample

Table 3 – Effect of Receiving a Higher Label Postsecondary Outcomes Math Proficient/Advanced Sample

	Any Postsecondary Enrollment	Any Postsecondary Enrollment	2-Year Enrollment	2-Year Enrollment	4-Year Enrollment	4-Year Enrollment
Advanced Label	-0.0020	-0.0020	-0.0003	-0.0002	-0.0018	-0.0018
	(0.0036)	(0.0038)	(0.0107)	(0.0107)	(0.0042)	(0.0048)
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Covariates	N	Y	N	Y	N	Y
Mean Outcome	0.95	0.95	0.52	0.52	0.86	0.86
	Any Postsecondary Degree	Any Postsecondary Degree	Associate Degree	Associate Degree	Bachelor's Degree	Bachelor's Degree
Advanced Label	-0.0099	-0.0097*	-0.0030	-0.0029	-0.0096	-0.0094
	(0.0054)	(0.0043)	(0.0020)	(0.0022)	(0.0051)	(0.0051)
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Covariates	N	Y	N	Y	N	Y
Mean Outcome	0.70	0.70	0.10	0.10	0.62	0.62

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. N = 40,349. Bandwidth = 5 scale score points. Standard errors are clustered at the year level. Mean outcomes for the Math Proficient/Advanced Sample are shown.

In Table 3 I estimate the treatment effect of receiving an Advanced label using the Math Proficient/Advanced Sample. For all the outcomes I check I estimate that the treatment effect is small and statistically insignificant both with and without covariates. I conclude that there is no effect on average of receiving an Advanced label on the 11th grade math exam on postsecondary outcomes.

Table 4 – Male and Female Treatment Effect Math Advanced Label

	Any Postsecondary Enrollment	Two-Year Enrollment	Four-Year Enrollment	Any Postsecondary Degree	Associate Degree	Bachelor's Degree
Advanced	-0.0007	0.0112	0.0042	0.0009	-0.0010	-0.0000
	(0.0054)	(0.0082)	(0.0086)	(0.0082)	(0.0053)	(0.0095)
Advanced * Female	-0.0032	-0.0271	-0.0137	-0.0246	-0.0049	-0.0218
	(0.0052)	(0.0197)	(0.0108)	(0.0150)	(0.0107)	(0.0136)
P(Advanced + Interaction)	0.26	0.46	0.08	0.05	0.38	0.03
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Mean Outcome Males	0.94	0.52	0.83	0.63	0.09	0.55
Mean Outcome Females	0.97	0.53	0.90	0.78	0.10	0.70

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. N = 40,349. Bandwidth = 5 scale score points. Mean outcomes are for students in the Math Proficient/Advanced Sample. All regressions include year fixed effects. Standard errors are clustered at the year level.

In Table 4 I estimate the treatment effect of receiving an Advanced label using the Math Proficient/Advanced Sample for male and female students. For most of the outcomes I check I estimate that the treatment effect for both groups of students and the difference between the two

treatment effects is small and statistically insignificant. However, I estimate that receiving an Advanced label causes female students to be significantly less likely to complete any postsecondary degree (2.4 percentage points) and less likely to complete a bachelor's degree (2.2 percentage points). While the treatment effect for males is very close to 0, for neither outcome are the effects for male and female students significantly different.

Table 5 – White and Black Treatment Effect Math Advanced Label

	Any Postsecondary Enrollment	Two-Year Enrollment	Four-Year Enrollment	Any Postsecondary Degree	Associate Degree	Bachelor's Degree
Advanced	-0.0022	0.0034	-0.0027	-0.0143	-0.0042*	-0.0147
	(0.0036)	(0.0121)	(0.0046)	(0.0081)	(0.0019)	(0.0089)
Advanced * Black	0.0363*	-0.0027	0.0363*	0.1348*	0.0240	0.1412*
	(0.0176)	(0.0570)	(0.0185)	(0.0657)	(0.0331)	(0.0637)
P(Advanced + Interaction)	0.10	0.99	0.13	0.09	0.57	0.07
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Mean Outcome White	0.95	0.53	0.86	0.70	0.10	0.62
Mean Outcome Black	0.96	0.52	0.88	0.58	0.06	0.52

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. N = 36,634. Only white and black students are included in the regressions. Bandwidth = 5 scale score points. Mean outcomes are for students in the Math Proficient/Advanced Sample. All regressions include year fixed effects. Standard errors are clustered at the year level.

In Table 5 I estimate the treatment effect of receiving an Advanced label using the Math Proficient/Advanced Sample for white and black students. No treatment effect nor any difference in treatment effect for white and black students is significant at the 5% level. However, point estimates for the treatment effect for black students are large and significant at the 10% level for both any postsecondary degree (12 percentage points) and bachelor's degree (13 percentage points). These degree treatment effect estimates are much larger than the effect estimates for ever enrolling in a postsecondary institution (3 percentage points) or enrolling in a 4-year institution (3 percentage points) respectively. Assuming these are real effects rather than estimates being due to random variation, then it would mean getting a higher label would increase the probability of a student getting a bachelor's degree for some black students who were already planning on enrolling in a postsecondary institution. I conclude that getting an Advanced label has little effect on postsecondary outcomes for white students but that it may make black students more likely to get a bachelor's degree.

Table 6 – Difference by Economically Disadvantage Treatment Effect Math Advanced Label

	Any Postsecondary Enrollment	Two-Year Enrollment	Four-Year Enrollment	Any Postsecondary Degree	Associate Degree	Bachelor's Degree
Advanced	-0.0002	0.0014	0.0027	-0.0059	-0.0002	-0.0069
	(0.0035)	(0.0114)	(0.0049)	(0.0035)	(0.0028)	(0.0049)
Advanced *	-0.0173	-0.0168	-0.0402*	-0.0404	-0.0245	-0.0296
Economically	(0.0148)	(0.0430)	(0.0185)	(0.0238)	(0.0186)	(0.0255)
Disadvantaged						
P(Advanced +	0.29	0.72	0.07	0.11	0.19	0.17
Interaction)						
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Mean Outcome Not	0.96	0.52	0.88	0.73	0.09	0.66
Economically						
Disadvantaged						
Mean Outcome	0.91	0.57	0.74	0.51	0.12	0.41
Economically						
Disadvantaged						

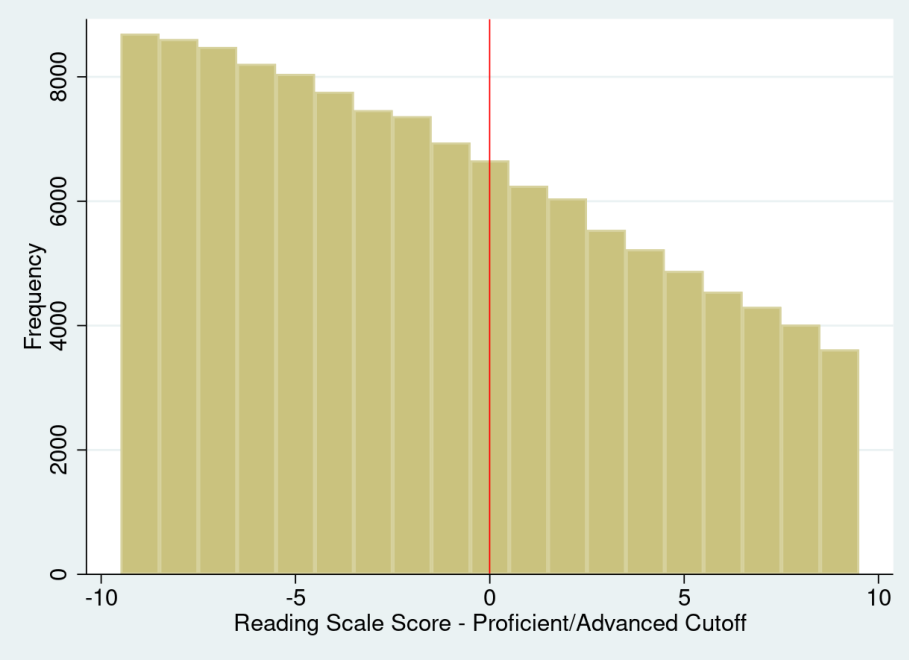
Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. N = 40,349. Bandwidth = 5 scale score points. Mean outcomes are for students in the Math Proficient/Advanced Sample. All regressions include year fixed effects. Standard errors are clustered at the year level.

In Table 6 I estimate the treatment effect of receiving an Advanced label using the Math Proficient/Advanced Sample for students who are and are not economically disadvantaged. None of the treatment effects for students who are or are not economically disadvantaged nor the difference between the treatment effects are statistically significant. I conclude that getting an Advanced label has little average effect on postsecondary outcomes for either group of students.

VII. Reading Proficient/Advanced Cutoff Results

A. Identification Test: Discontinuity in Density

Figure 2 – Histogram Students Close to the Cutoff Reading Proficient/Advanced Sample



Notes: N = 122,628. Each bar in this histogram shows the number of students in the sample who received a different scale score.

Figure 2 shows the number of students in the sample who receive different scale scores for values of the scale score close to the Proficient/Advanced cutoffs for the reading exam. Like for the Math Proficient/Advanced sample, I find no visual evidence of a discontinuity in the density of observations at the cutoffs.

B. Identification Test: Discontinuities in Covariates

Table 7 – Discontinuity in Covariates Reading Proficient/Advanced Sample

	Female	White	Black	Hispanic
Advanced Label	0.0010	-0.0001	0.0004	-0.0001
	(0.0068)	(0.0028)	(0.0020)	(0.0018)
Year Fixed Effects	Y	Y	Y	Y
Mean Outcome	0.53	0.85	0.07	0.03
	Asian	Two or More Races	Native American	Hawaiian
Advanced Label	-0.0003	0.0003	-0.0001	-0.0000
	(0.0016)	(0.0007)	(0.0010)	(0.0004)
Year Fixed Effects	Y	Y	Y	Y
Mean Outcomes	0.03	0.02	0.01	0.00
	Economically Disadvantaged			
Advanced Label	0.0008			
	(0.0057)			
Year Fixed Effects	Y			
Mean Outcomes	0.24			

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. N = 122,628. Bandwidth = 9 scale score points. Standard errors are clustered at the year level. The outcomes are indicator variables for being female, being a specific race or being economically disadvantaged. Mean outcomes for the Reading Proficient/Advanced Sample are shown.

In Table 7 I estimate discontinuities in the proportion of students who have different characteristics at the Proficient/Advanced cutoffs for the reading exam. I find that all discontinuities are small and statistically insignificant. Again, I conclude that discontinuities in outcomes at the cutoffs are due to the higher performance label rather than manipulation of students' scale scores.

C. Higher Label Treatment Effect Reading Proficient/Advanced Sample

Table 8 – Effect of Receiving a Higher Label Postsecondary Outcomes Reading Proficient/Advanced Sample

	Any Postsecondary Enrollment	Any Postsecondary Enrollment	2-Year Enrollment	2-Year Enrollment	4-Year Enrollment	4-Year Enrollment
Advanced Label	0.0005	0.0005	-0.0027	-0.0028	0.0013	0.0014
	(0.0028)	(0.0027)	(0.0077)	(0.0077)	(0.0047)	(0.0049)
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Covariates	N	Y	N	Y	N	Y
Mean Outcome	0.90	0.90	0.58	0.58	0.73	0.73
	Any Postsecondary Degree	Any Postsecondary Degree	Associate Degree	Associate Degree	Bachelor's Degree	Bachelor's Degree
Advanced Label	-0.0013	-0.0011	0.0104**	0.0104**	-0.0058	-0.0056
	(0.0062)	(0.0058)	(0.0032)	(0.0033)	(0.0078)	(0.0072)
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Covariates	N	Y	N	Y	N	Y
Mean Outcome	0.56	0.56	0.11	0.11	0.46	0.46

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. N = 122,628. Bandwidth = 9 scale score points. Standard errors are clustered at the year level. Mean outcomes for the Reading Proficient/Advanced Sample are shown.

In Table 8 I estimate the higher label treatment effect using the Reading Proficient/Advanced Sample. For 5 of 6 outcomes, I estimate that the treatment effect is small and statistically insignificant. However, I estimate that getting an Advanced label causes students to be a significant 1 percentage point more likely to earn an associate degree.

Table 9 – Male and Female Treatment Effect Reading Advanced Label

	Any Postsecondary Enrollment	Two-Year Enrollment	Four-Year Enrollment	Any Postsecondary Degree	Associate Degree	Bachelor's Degree
Advanced	0.0032	-0.0041	0.0060*	0.0060*	0.0142**	-0.0022
	(0.0026)	(0.0084)	(0.0030)	(0.0025)	(0.0051)	(0.0070)
Advanced * Female	-0.0050	0.0025	-0.0088	-0.0138	-0.0069	-0.0069
	(0.0044)	(0.0053)	(0.0077)	(0.0097)	(0.0067)	(0.0083)
P(Advanced + Interaction)	0.67	0.85	0.74	0.46	0.14	0.38
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Mean Outcome Males	0.88	0.55	0.69	0.50	0.10	0.41
Mean Outcome Females	0.92	0.60	0.77	0.61	0.13	0.50

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. N = 122,628. Bandwidth = 9 scale score points. Mean outcomes are for students in the Reading Proficient/Advanced Sample. All regressions include year fixed effects. Standard errors are clustered at the year level.

In Table 9 I estimate the higher label treatment effect using the Reading Proficient/Advanced Sample for male and female students. Consistent with my results for all students I find that getting an Advanced label causes male students to be 1.4 percentage points

more likely to earn an associate degree. The effect for female students, while not significantly different than the male student effect, is about half the magnitude and not statistically significant at the 10% level. I do not find evidence of a significant effect for either male or female students for the other outcomes I look at.

Table 10 – White and Black Treatment Effect Reading Advanced Label

	Any Postsecondary Enrollment	Two-Year Enrollment	Four-Year Enrollment	Any Postsecondary Degree	Associate Degree	Bachelor's Degree
Advanced	0.0016	-0.0001	0.0015	-0.0015	0.0127***	-0.0065
	(0.0023)	(0.0083)	(0.0044)	(0.0054)	(0.0033)	(0.0074)
Advanced * Black	-0.0328	-0.0304	-0.0431*	-0.0214	-0.0373**	-0.0082
	(0.0237)	(0.0226)	(0.0211)	(0.0193)	(0.0138)	(0.0235)
P(Advanced + Interaction)	0.25	0.17	0.12	0.33	0.14	0.60
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Mean Outcome White	0.90	0.58	0.73	0.57	0.12	0.47
Mean Outcome Black	0.91	0.61	0.72	0.39	0.07	0.31

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. N = 113,366. Only white and black students are included in the regressions. Bandwidth = 9 scale score points. Mean outcomes are for students in the Reading Proficient/Advanced Sample. All regressions include year fixed effects. Standard errors are clustered at the year level.

In Table 10 I look at the Advanced label treatment effect for white and black students. Again, the only significant coefficients are for associate degree completion. I find getting an Advanced label causes white students to be 1.2 percentage points more likely and black students 2.5 percentage points less likely to earn an associate degree. While the black student effect is not significant, the difference between the two effects is.

Table 11 – Difference by Economically Disadvantage Treatment Effect Reading Advanced Label

	Any Postsecondary Enrollment	Two-Year Enrollment	Four-Year Enrollment	Any Postsecondary Degree	Associate Degree	Bachelor's Degree
Advanced	-0.0012	-0.0023	0.0019	0.0009	0.0093**	-0.0020
	(0.0022)	(0.0064)	(0.0035)	(0.0060)	(0.0038)	(0.0079)
Advanced * Economically Disadvantaged	0.0101	-0.0007	-0.0032	-0.0118	0.0069	-0.0206**
	(0.0066)	(0.0153)	(0.0126)	(0.0127)	(0.0106)	(0.0073)
P(Advanced + Interaction)	0.29	0.87	0.94	0.46	0.12	0.04
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Mean Outcome Not Economically Disadvantaged	0.93	0.58	0.78	0.62	0.11	0.53
Mean Outcome Economically Disadvantaged	0.82	0.58	0.58	0.35	0.11	0.25

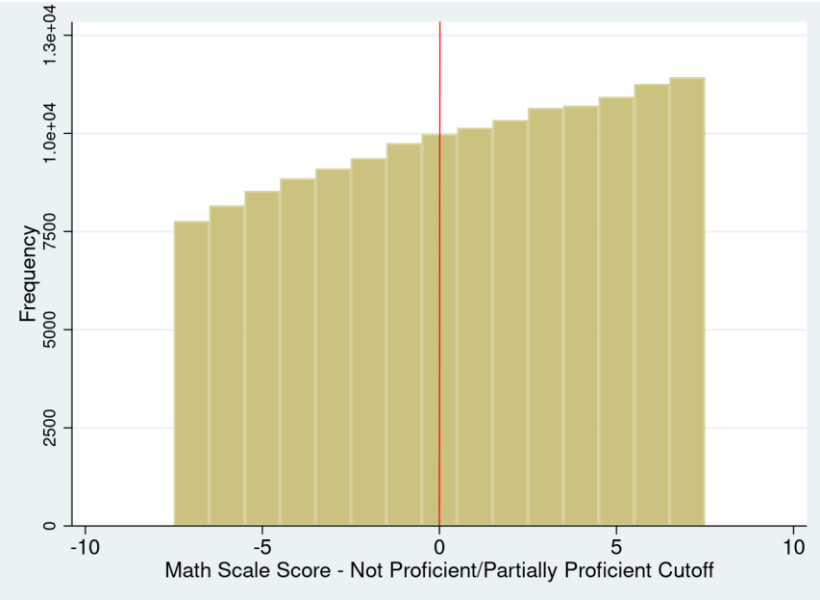
Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. N = 122,628. Bandwidth = 9 scale score points. Mean outcomes are for students in the Reading Proficient/Advanced Sample. All regressions include year fixed effects. Standard errors are clustered at the year level.

In Table 11 I look at the Advanced label treatment effect for students who are and are not economically disadvantaged. I find two significant treatment effects. Getting an Advanced label makes students who are not economically disadvantaged 0.93 percentage points more likely to earn an associate degree. I also find that getting an Advanced label makes economically disadvantaged students 2.3 percentage points less likely to earn a bachelor's degree. The only outcome where the effects for the two groups are significantly different is for earning a bachelor's degree.

VIII. Math Not Proficient/Partially Proficient Cutoff Results

A. Identification Test: Discontinuity in Density

Figure 3 – Histogram Students Close to the Cutoff Math Not Proficient/Partially Proficient Sample



Notes: N = 146,961. Each bar in this histogram shows the number of students in the sample who received a different scale score.

Figure 3 shows the number of students who receive different scale scores for scores close to the cutoffs. I find no visual evidence of a discontinuous change in the number of students at the cutoffs.

B. Identification Test: Discontinuities in Covariates

Table 12 – Discontinuity in Covariates Math Not Proficient/Partially Proficient Sample

	Female	White	Black	Hispanic
Partially Proficient Label	-0.0068 (0.0052)	0.0007 (0.0068)	-0.0055 (0.0054)	0.0015 (0.0022)
Year Fixed Effects	Y	Y	Y	Y
Mean Outcome	0.52	0.72	0.19	0.04
	Asian	Two or More Races	Native American	Hawaiian
Partially Proficient Label	-0.0004 (0.0009)	0.0024** (0.0008)	0.0007 (0.0011)	0.0005 (0.0003)
Year Fixed Effects	Y	Y	Y	Y
Mean Outcomes	0.02	0.02	0.01	0.00
	Economically Disadvantaged			
Partially Proficient Label	-0.0043 (0.0067)			
Year Fixed Effects	Y			
Mean Outcomes	0.42			

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. N = 146,961. Bandwidth = 7 scale score points. Standard errors are clustered at the year level. The outcomes are indicator variables for being female, being a specific race, or being economically disadvantaged. Mean outcomes for the Math Not Proficient/Partially Proficient Sample are shown.

In Table 12 I estimate discontinuities in the proportion of students who have different traits at the Not Proficient/Partially Proficient cutoffs for the math exam. For most of the characteristics I check the discontinuity is small and statistically insignificant. The exception to this is that I estimate the proportion of students who are two or more races increases by a significant 0.24 percentage points at the cutoff. Even if this is due to manipulation of the scale score, the manipulation seems to be only for a small percentage of students. This difference also probably will not have a big effect on my estimates because students who are two or more races are only 2% of the Math Not Proficient/Partially Proficient Sample.

C. Higher Label Treatment Effect Math Not Proficient/Partially Proficient Sample

Table 13 – Effect of Receiving a Higher Label Postsecondary Outcomes Math Not Proficient/Partially Proficient Sample

	Any Postsecondary Enrollment	Any Postsecondary Enrollment	2-Year Enrollment	2-Year Enrollment	4-Year Enrollment	4-Year Enrollment
Partially Proficient Label	-0.0033	-0.0021	-0.0059	-0.0053	0.0011	0.0028
	(0.0040)	(0.0048)	(0.0048)	(0.0052)	(0.0034)	(0.0043)
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Covariates	N	Y	N	Y	N	Y
Mean Outcome	0.75	0.75	0.59	0.59	0.45	0.45
	Any Postsecondary Degree	Any Postsecondary Degree	Associate Degree	Associate Degree	Bachelor's Degree	Bachelor's Degree
Partially Proficient Label	0.0027	0.0033	-0.0024	-0.0024	-0.0016	-0.0010
	(0.0039)	(0.0047)	(0.0026)	(0.0028)	(0.0029)	(0.0027)
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Covariates	N	Y	N	Y	N	Y
Mean Outcome	0.30	0.30	0.11	0.11	0.19	0.19

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. N = 146,961. Bandwidth = 7 scale score points. Standard errors are clustered at the year level. Mean outcomes for the Math Not Proficient/Partially Proficient Sample are shown.

In Table 13 I estimate the effect of receiving a Partially Proficient label on postsecondary outcomes. All the estimated treatment effects are small and statistically insignificant.

Table 14 – Male and Female Treatment Effect Math Partially Proficient Label

	Any Postsecondary Enrollment	Two-Year Enrollment	Four-Year Enrollment	Any Postsecondary Degree	Associate Degree	Bachelor's Degree
Partially Proficient	-0.0029	-0.0012	0.0006	0.0154**	0.0091**	0.0007
	(0.0078)	(0.0079)	(0.0046)	(0.0053)	(0.0036)	(0.0026)
Partially Proficient * Female	0.0009	-0.0077	0.0028	-0.0222*	-0.0209***	-0.0033
	(0.0086)	(0.0079)	(0.0091)	(0.0096)	(0.0038)	(0.0070)
P(Partially Proficient + Interaction)	0.62	0.09	0.65	0.39	0.01	0.68
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Mean Outcome Males	0.68	0.54	0.38	0.23	0.09	0.14
Mean Outcome Females	0.81	0.63	0.52	0.37	0.13	0.24

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. N = 146,691. Bandwidth = 7 scale score points. Mean outcomes are for students in the Math Not Proficient/Partially Proficient Sample. All regressions include year fixed effects. Standard errors are clustered at the year level.

In Table 14 I estimate the effect of receiving a Partially Proficient label for male and female students. For most outcomes I find that the male treatment effect, the female treatment effect, and the difference in the treatment effects is not statistically significant. I find that receiving a Partially Proficient label makes male students significantly more likely to earn a postsecondary degree (1.5 percentage points) and to earn an associate degree (0.91 percentage points). The effects for females are an insignificant 0.68 decrease in receiving any postsecondary degree and a significant 1.2 percentage point decrease in completing an associate degree. The difference in the associate degree treatment effects is statistically significant.

Table 15 – White and Black Treatment Effect Math Partially Proficient Label

	Any Postsecondary Enrollment	Two-Year Enrollment	Four-Year Enrollment	Any Postsecondary Degree	Associate Degree	Bachelor's Degree
Partially Proficient	-0.0028	-0.0080	-0.0016	0.0022	-0.0043*	-0.0009
	(0.0036)	(0.0055)	(0.0038)	(0.0029)	(0.0018)	(0.0044)
Partially Proficient * Black	0.0054	0.0071	0.0298	0.0077	0.0037	0.0061
	(0.0091)	(0.0095)	(0.0146)	(0.0064)	(0.0086)	(0.0036)
P(Partially Proficient + Interaction)	0.79	0.94	0.13	0.28	0.95	0.22
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Mean Outcome White	0.74	0.58	0.46	0.33	0.12	0.22
Mean Outcome Black	0.78	0.62	0.45	0.20	0.06	0.13

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. N = 133,804. Only white and black students are included in the regressions. Bandwidth = 7 scale score points. Mean outcomes are for students in the Math Not/Proficient Partially Proficient Sample. All regressions include year fixed effects. Standard errors are clustered at the year level.

In Table 15 I estimate the effect of receiving a Partially Proficient label for white and black students. None of the treatment effects are statistically significant.

Table 16 – Difference by Economically Disadvantage Treatment Effect Math Partially Proficient Label

	Any Postsecondary Enrollment	Two-Year Enrollment	Four-Year Enrollment	Any Postsecondary Degree	Associate Degree	Bachelor's Degree
Partially Proficient	0.0017	-0.0017	0.0065	0.0071	-0.0031	0.0007
	(0.0039)	(0.0058)	(0.0040)	(0.0047)	(0.0038)	(0.0039)
Partially Proficient * Economically Disadvantaged	-0.0131*	-0.0109*	-0.0143*	-0.0120**	0.0012	-0.0067
	(0.0056)	(0.0056)	(0.0064)	(0.0047)	(0.0061)	(0.0053)
P(Advanced + Interaction)	0.12	0.06	0.22	0.28	0.67	0.11
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Mean Outcome Not Economically Disadvantaged	0.80	0.62	0.52	0.38	0.13	0.26
Mean Outcome Economically Disadvantaged	0.67	0.54	0.35	0.19	0.08	0.10

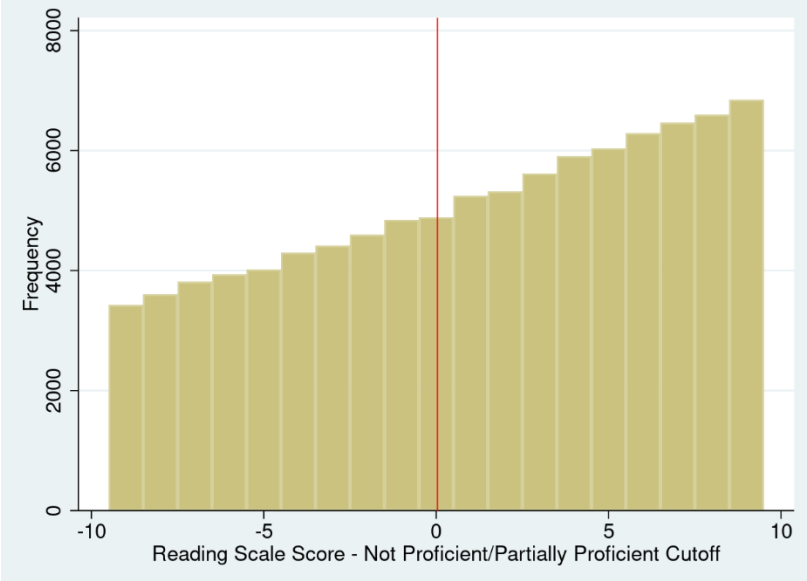
Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. N = 146,961. Bandwidth = 7 scale score points. Mean outcomes are for students in the Math Not Proficient/Partially Proficient Sample. All regressions include year fixed effects. Standard errors are clustered at the year level.

In Table 16 I estimate the math Partially Proficient treatment effect for students who are and are not economically disadvantaged. None of the treatment effects are statistically significant.

IX. Reading Not Proficient/Partially Proficient Cutoff

A. Identification Test: Discontinuity in Density

Figure 4 – Histogram Students Close to the Cutoff Reading Not Proficient/Partially Proficient Sample



Notes: N = 96,171. Each bar in this histogram shows the number of students in the sample who received a different scale score.

Figure 4 shows the number of students in the sample who receive different scale scores for values close to the cutoff. Visually there is no discontinuous change in the number of students at the cutoff.

B. Identification Check: Discontinuities in Covariates

Table 17 – Discontinuity in Covariates Reading Not Proficient/Partially Proficient Sample

	Female	White	Black	Hispanic
Partially Proficient Label	0.0046	0.0101	-0.0065	-0.0027
	(0.0033)	(0.0022)	(0.0080)	(0.0033)
Year Fixed Effects	Y	Y	Y	Y
Mean Outcome	0.46	0.67	0.24	0.05
	Asian	Two or More Races	Native American	Hawaiian
Partially Proficient Label	-0.0013	0.0010	-0.0009	0.0004
	(0.0016)	(0.0021)	(0.0014)	(0.0006)
Year Fixed Effects	Y	Y	Y	Y
Mean Outcomes	0.02	0.02	0.01	0.00
	Economically Disadvantaged			
Partially Proficient Label	0.0053			
	(0.0057)			
Year Fixed Effects	Y			
Mean Outcomes	0.47			

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. N = 96,169. Bandwidth = 9 scale score points. Standard errors are clustered at the year level. The outcomes are indicator variables for being female, being a specific race, or being economically disadvantaged. Mean outcomes for the Reading Not Proficient/Partially Proficient Sample are shown.

In Table 17 I estimate discontinuities in the proportion of students who are female, who are a specific race, or who are economically disadvantaged at the Reading Not Proficient/Partially Proficient cutoff. All the discontinuities are small and statistically insignificant.

C. Higher Label Treatment Effect Reading Not Proficient/Partially Proficient Sample

Table 18 – Effect of Receiving a Higher Label Postsecondary Outcomes Reading Not Proficient/Partially Proficient Sample

	Any Postsecondary Enrollment	Any Postsecondary Enrollment	2-Year Enrollment	2-Year Enrollment	4-Year Enrollment	4-Year Enrollment
Partially Proficient Label	-0.0023	-0.0011	-0.0047	-0.0036	-0.0029	-0.0021
	(0.0100)	(0.0096)	(0.0100)	(0.0096)	(0.0059)	(0.0060)
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Race and Gender Controls	N	Y	N	Y	N	Y
Mean Outcome	0.68	0.68	0.56	0.56	0.36	0.36
	Any Postsecondary Degree	Any Postsecondary Degree	Associate Degree	Associate Degree	Bachelor's Degree	Bachelor's Degree
Partially Proficient Label	-0.0033	-0.0033	-0.0002	-0.0005	-0.0012	-0.0010
	(0.0049)	(0.0048)	(0.0043)	(0.0042)	(0.0034)	(0.0035)
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Race and Gender Controls	N	Y	N	Y	N	Y
Mean Outcome	0.23	0.23	0.09	0.09	0.13	0.13

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. N = 96,169. Bandwidth = 9 scale score points. Standard errors are clustered at the year level. Mean outcomes for the Reading Not Proficient/Partially Proficient sample are shown.

In Table 18 I estimate treatment effects of receiving a Partially Proficient label on the reading exam. All the effects are small and statistically insignificant.

Table 19 – Male and Female Treatment Effect Reading Partially Proficient Label

	Any Postsecondary Enrollment	Two-Year Enrollment	Four-Year Enrollment	Any Postsecondary Degree	Associate Degree	Bachelor's Degree
Partially Proficient	-0.0094	-0.0089	-0.0091	-0.0086	0.0026	-0.0030
	(0.0107)	(0.0139)	(0.0059)	(0.0064)	(0.0050)	(0.0044)
Partially Proficient * Female	0.0144	0.0086	0.0127	0.0108	-0.0067	0.0035
	(0.0089)	(0.0113)	(0.0075)	(0.0104)	(0.0037)	(0.0085)
P(Partially Proficient + Interaction)	0.67	0.97	0.68	0.79	0.37	0.94
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Mean Outcome Males	0.63	0.51	0.32	0.19	0.08	0.11
Mean Outcome Females	0.74	0.61	0.41	0.27	0.11	0.15

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. N = 96,169. Bandwidth = 9 scale score points. Mean outcomes are for students in the Reading Not Proficient/Partially Proficient Sample. All regressions include year fixed effects. Standard errors are clustered at the year level.

Table 20 – White and Black Treatment Effect Reading Partially Proficient Label

	Any Postsecondary Enrollment	Two-Year Enrollment	Four-Year Enrollment	Any Postsecondary Degree	Associate Degree	Bachelor's Degree
Partially Proficient	-0.0033	-0.0028	-0.0085	-0.0009	-0.0004	-0.0001
	(0.0088)	(0.0097)	(0.0055)	(0.0059)	(0.0054)	(0.0020)
Partially Proficient * Black	0.0066	-0.0057	0.0207*	-0.0099	-0.0107	-0.0003
	(0.0077)	(0.0132)	(0.0102)	(0.0090)	(0.0077)	(0.0100)
P(Partially Proficient + Interaction)	0.82	0.58	0.33	0.16	0.10	0.96
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Mean Outcome White	0.67	0.54	0.36	0.26	0.11	0.15
Mean Outcome Black	0.74	0.62	0.37	0.16	0.05	0.08

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. N = 86,906. Only white and black students are included in the regressions. Bandwidth = 9 scale score points. Mean outcomes are for students in the Reading Not Proficient/Partially Proficient sample. All regressions include year fixed effects. Standard errors are clustered at the year level.

Table 21 – Difference by Economically Disadvantage Treatment Effect Reading Partially Proficient Label

	Any Postsecondary Enrollment	Two-Year Enrollment	Four-Year Enrollment	Any Postsecondary Degree	Associate Degree	Bachelor's Degree
Partially Proficient	0.0037	-0.0067	0.0052	0.0015	-0.0020	0.0029
	(0.0093)	(0.0115)	(0.0055)	(0.0074)	(0.0063)	(0.0036)
Partially Proficient * Economically Disadvantaged	-0.0111	0.0050	-0.0150	-0.0082	0.0039	-0.0069
	(0.0131)	(0.0140)	(0.0082)	(0.0100)	(0.0081)	(0.0061)
P(Advanced + Interaction)	0.62	0.90	0.31	0.35	0.74	0.48
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Mean Outcome Not Economically Disadvantaged	0.74	0.60	0.42	0.30	0.12	0.19
Mean Outcome Economically Disadvantaged	0.62	0.51	0.29	0.15	0.07	0.07

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. N = 96,169. Bandwidth = 9 scale score points. Mean outcomes are for students in the Reading Not Proficient/Partially Proficient sample. All regressions include year fixed effects. Standard errors are clustered at the year level.

In Table 19, Table 20, and Table 21 I estimate the effect of receiving a Partially Proficient label on the reading exam for male and female students, white and black students, and students who are and are not economically disadvantaged. The estimated effects for all groups of students are small and statistically insignificant.

X. Alternative Specifications

In the Appendix, I estimate higher label treatment effects including year fixed effects and covariates in the regressions for all samples using bandwidths that are 0.5 times and 1.5 times the

bandwidths, rounded to the nearest whole number, used in the main body of the paper. I do this to see how sensitive my results are to my choice of bandwidth.

Like the results in the main body of the paper, most estimated treatment effects are small and statistically insignificant. Sometimes I estimate large treatment effects for the smaller bandwidth but this same treatment effect for a larger bandwidth is a much smaller magnitude. For example, using a bandwidth of 3 scale score points, I estimate that getting an Advanced label on their math exam causes black students to be 24 percentage points more likely to complete a bachelor's degree. Using a bandwidth of 8 scale score points, I estimate that effect to only be 6 percentage points.

The significance level of a treatment effect often changes when I use a different bandwidth. Table A.17 lists all the treatment effects that are significant at the 5% level. I estimate the treatment effect for 4 samples, 6 outcomes, and 7 groups of students giving me estimates for 168 treatment effects[9]. Out of those treatment effects, 34 of them are significant with at least one bandwidth, 8 are significant with at least 2 bandwidths and only one is significant with all 3 bandwidths. Assuming all real treatment effects are 0, I would expect about 8 treatment effects to be significant due to random chance for any given choice of bandwidth. Given the limited number of significant effects I find relative to what I would expect due to random chance and given how the significance of an effect often changes when I use a different bandwidth, it is possible that the best way to interpret my results is that I do not find strong evidence of a large effect of which performance label a student gets on their post-secondary outcomes.

XI. Discussion and Conclusion

In this paper I study how which label a student receives summarizing their performance on standardized exams affects their postsecondary outcomes. I do this by using a regression discontinuity research design to compare the outcomes of students close to cutoffs to receive different performance labels. I use data on students in Michigan public schools who took 11th grade math and reading exams from the 2007 – 2008 school year to the 2013 – 2014 school year. I look at the effect both of receiving the label associated with the highest range of exam scores

[9] The 4 samples are Math Proficient/Advanced Sample, Reading Proficient/Advanced Sample, Math Not Proficient/Partially Proficient Sample, and Reading Not Proficient/Partially Proficient Sample. The 6 outcomes are any postsecondary enrollment, 2-year enrollment, 4-year enrollment, any postsecondary degree, associate degree, and bachelor's degree. The 7 groups of students are all, male, female, white, black, not economically disadvantaged, and economically disadvantaged.

(Advanced) compared to the label for the second highest range of scores (Proficient) and the effect of receiving the label for the second lowest range of scores (Partially Proficient) compared to the label for receiving the lowest range of scores (Not Proficient). I look for effects for all students, for male students, for female students, for white students, for black students, for students who are not economically disadvantaged, and for students who are economically disadvantaged.

While I find some statistically significant treatment effects, it is possible that they are almost all due to random chance rather than the treatment effect being different from 0. Many of these effects are not significant when using another bandwidth, and I would expect to find some significant effects given the large number of estimates I get. Out of a total of 168 estimated treatment effects, 34 are significant using at least one bandwidth and only one is significant using all 3 bandwidths. Because of this, it is possible that the best interpretation of my results is that I do not find strong evidence that which performance label a student receives affects their postsecondary outcomes.

There are many different directions that future researchers could go in when studying test score labels. They could see if the labels change the quality of an institution a student goes to. They could see if the labels change the type of major a student chooses to study. They can look at the effect of labels in different states and for different grades. They could look at the effect of labels on K-12 academic outcomes such as scores on future exams, grades, high school graduation, and characteristics of the K-12 schools students attend.

REFERENCES

Avery, Christopher, and Joshua Goodman. 2021. "Ability Signals and Rigorous Coursework: Evidence from AP Calculus Participation." Working Paper.

Beuchert, Louise, Tine L. M. Eriksen, Morten V. Krægpøth. 2020. "The impact of standardized test feedback in math: Exploiting a natural experiment in 3rd grade." *Economics of Education Review*. Vol. 77

Bobba, Matteo, and Veronica Frisancho. 2019. "Perceived Ability and School Choice." TSE Working Paper 16-660

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik. 2014. "Robust Nonparamteric Confidence Intervals for Regression-Discontinuity Designs." *Econometrica*. Vol. 82(6) pp. 2295 – 2326

Donaldson, Morgaen L., **and John P. Papay.** 2015. "Teacher Evaluation for Accountability and Development." Pp. 174 – 193. In *Handbook of Research in Education Finance and Policy.* Edited by Helen F. Ladd and Margaret E. Goertz. Published by Taylor & Francis Group.

Figlio, David N., **and Helen F.** Ladd. 2015. "School Accountability and Student Achievement" Pp. 194 – 210. In *Handbook of Research in Education Finance and Policy.* Edited by Helen F. Ladd and Margaret E. Goertz. Published by Taylor & Francis Group.

Figlio, David, and Susanna Loeb. 2011. "School Accountability." Pp. 383 – 421 In *Handbook of the Economics of Education Volume 3*. Edited by Eric A. Hanushek, Stephen J. Machin, and Ludger Woessmann.

Foote, Andrew, Lisa Schulkind, and Teny M. Shapiro. 2015. "Missed Signals: The Effect of ACT College-Readiness Measures on Post-Secondary Decisions." *Economics of Education Review.* Vol. 46. Pp. 39 - 51

Gershenson, Seth. 2018. "Grade Inflation in High Schools (2005 – 2016)." Thomas B. Fordham Institute Report. https://fordhaminstitute.org/national/research/grade-inflation-high-schools-2005-2016

Papay, John P., Richard J. Murnane and John B. Willett. 2016. "The Impact of Test Score Labels on Human-Capital Investment Decisions." *The Journal of Human Resources*. Vol. 52(2) pp 357 – 388

Pattison, Evangeleen, Eric Grodsky, and Chandra Muller. 2013. "Is the Sky Falling? Grade Inflation and the Signaling Power of Grades." *Educational Researcher*. Vol. 42(5) pp. 259 – 265

Rotherham, Andrew J. 2006. "Making the Cut: How States Set Passing Scores on Standardized Tests." Education Sector. Published July 2006

Schleicher, Andreas. 2015. "Are American students overtested? Listen to what students themselves say." OECD Education and Sills Today. Published November 18, 2015.

APPENDIX

A.1 Robustness Check Different Bandwidths

A.1.1 Math Proficient/Advanced Sample

Table A.1 – Effect of Receiving a Higher Label Postsecondary Outcomes Math Proficient/Advanced Sample

	Any Postsecondary Enrollment	Any Postsecondary Enrollment	2-Year Enrollment	2-Year Enrollment	4-Year Enrollment	4-Year Enrollment
Advanced Label	-0.0018	-0.0003	0.0088	0.0017	-0.0004	0.0027
	(0.0050)	(0.0029)	(0.0109)	(0.0075)	(0.0083)	(0.0054)
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Race and Gender Controls	Y	Y	Y	Y	Y	Y
Bandwidth	3	8	3	8	3	8
Number of Observations	25,418	63,456	25,418	63,456	25,418	63,456
Mean Outcome	0.75	0.75	0.59	0.59	0.45	0.45
	Any Postsecondary Degree	Any Postsecondary Degree	Associate Degree	Associate Degree	Bachelor's Degree	Bachelor's Degree
Advanced Label	-0.0107*	-0.0075	-0.0049	-0.0061**	-0.0101	0.0004
	(0.0049)	(0.0047)	(0.0048)	(0.0018)	(0.0066)	(0.0049)
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Race and Gender Controls	Y	Y	Y	Y	Y	Y
Bandwidth	3	8	3	8	3	8
Number of Observations	25,418	63,456	25,418	63,456	25,418	63,456
Mean Outcome	0.30	0.30	0.11	0.11	0.19	0.19

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Bandwidth is measured in scale score points. Standard errors are clustered at the year level. Mean outcomes for the Math Proficient/Advanced Sample are shown.

Table A.2 – Effect of Receiving a Higher Label Postsecondary Outcomes Math Proficient/Advanced Sample Male and Female Students

	Any Postsecondary Enrollment	Any Postsecondary Enrollment	2-Year Enrollment	2-Year Enrollment	4-Year Enrollment	4-Year Enrollment
Advanced Label	0.0005	-0.0017	0.0235*	0.0096	0.0062	0.0045
	(0.0075)	(0.0043)	(0.0112)	(0.0050)	(0.0153)	(0.0086)
Advanced * Female	-0.0040	0.0036	-0.0352	-0.0182*	-0.0130	-0.0038
	(0.0086)	(0.0048)	(0.0338)	(0.0077)	(0.0174)	(0.0101)
P(Advanced + Interaction)	0.50	0.57	0.69	0.47	0.24	0.92
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Bandwidth	3	8	3	8	3	8
Number of Observations	25,418	63,456	25,418	63,456	25,418	63,456
Mean Outcome Males	0.94	0.94	0.52	0.52	0.83	0.83
Mean Outcome Females	0.97	0.97	0.53	0.53	0.90	0.90
	Any Postsecondary Degree	Any Postsecondary Degree	Associate Degree	Associate Degree	Bachelor's Degree	Bachelor's Degree
Advanced Label	0.0011	-0.0017	0.0010	-0.0018	-0.0002	0.0064
	(0.0138)	(0.0075)	(0.0075)	(0.0047)	(0.0141)	(0.0054)
Advanced * Female	-0.0239	-0.0127	-0.0146	-0.0102	-0.0192	-0.0134
	(0.0268)	(0.0110)	(0.0099)	(0.0092)	(0.0191)	(0.0087)
P(Advanced + Interaction)	0.19	0.12	0.05	0.05	0.04	0.47
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Bandwidth	3	8	3	8	3	8
Number of Observations	25,418	63,456	25,418	63,456	25,418	63,456
Mean Outcome Males	0.63	0.63	0.09	0.09	0.55	0.55
Mean Outcome Females	0.78	0.78	0.10	0.10	0.70	0.70

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Bandwidth is measured in scale score points. Mean outcomes are for students in the Math Proficient/Advanced Sample. All regressions include year fixed effects. Standard errors are clustered at the year level.

Table A.3 – Effect of Receiving a Higher Label Postsecondary Outcomes Math Proficient/Advanced Sample White and Black Students

	Any Postsecondary Enrollment	Any Postsecondary Enrollment	2-Year Enrollment	2-Year Enrollment	4-Year Enrollment	4-Year Enrollment
Advanced Label	-0.0015	-0.0010	0.0083	0.0001	0.0005	0.0016
	(0.0046)	(0.0025)	(0.0120)	(0.0091)	(0.0074)	(0.0056)
Advanced * Black	0.0648	0.0294*	0.1848*	-0.0256	0.1035**	0.0248
	(0.0401)	(0.0137)	(0.0934)	(0.0544)	(0.0398)	(0.0204)
P(Advanced + Interaction)	0.16	0.07	0.07	0.62	0.05	0.19
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Bandwidth	3	8	3	8	3	8
Number of Observations	23,093	57,599	23,093	57,599	23,093	57,599
Mean Outcome White Students	0.95	0.95	0.53	0.53	0.86	0.86
Mean Outcome Black Students	0.96	0.96	0.52	0.52	0.88	0.88
	Any Postsecondary Degree	Any Postsecondary Degree	Associate Degree	Associate Degree	Bachelor's Degree	Bachelor's Degree
Advanced Label	-0.0154*	-0.0080	-0.0091*	-	-0.0135	0.0010
	(0.0074)	(0.0072)	(0.0047)	0.0086***	(0.0080)	(0.0083)
				(0.0021)		
Advanced * Black	0.2608**	0.0927*	0.0328	0.0438*	0.2579***	0.0596
	(0.0924)	(0.0461)	(0.0402)	(0.0207)	(0.0637)	(0.0546)
P(Advanced + Interaction)	0.03	0.09	0.60	0.15	0.01	0.30
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Bandwidth	3	8	3	8	3	8
Number of Observations	23,093	57,599	23,093	57,599	23,093	57,599
Mean Outcome White Students	0.70	0.70	0.10	0.10	0.62	0.62
Mean Outcome Black Students	0.58	0.58	0.06	0.06	0.52	0.52

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Bandwidth is measured in scale score points. Mean outcomes are for students in the Math Proficient/Advanced Sample. All regressions include year fixed effects. Standard errors are clustered at the year level.

Table A.4 – Effect of Receiving a Higher Label Postsecondary Outcomes Math Proficient/Advanced Sample Not Economically Disadvantaged and Economically Disadvantaged Students

	Any Postsecondary Enrollment	Any Postsecondary Enrollment	2-Year Enrollment	2-Year Enrollment	4-Year Enrollment	4-Year Enrollment
Advanced Label	0.0004	0.0017	0.0084	0.0042	0.0038	0.0063
	(0.0050)	(0.0026)	(0.0114)	(0.0097)	(0.0089)	(0.0053)
Advanced * Economically Disadvantaged	-0.0189	-0.0157	0.0074	-0.0218	-0.0445	-0.0240
	(0.0180)	(0.0122)	(0.0548)	(0.0412)	(0.0344)	(0.0149)
P(Advanced + Interaction)	0.34	0.32	0.77	0.64	0.25	0.30
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Bandwidth	3	8	3	8	3	8
Number of Observations	25,418	63,456	25,418	63,456	25,418	63,456
Mean Outcome Not Economically Disadvantaged	0.96	0.96	0.52	0.52	0.88	0.88
Mean Outcome Economically Disadvantaged	0.91	0.91	0.57	0.57	0.74	0.74
	Any Postsecondary Degree	Any Postsecondary Degree	Associate Degree	Associate Degree	Bachelor's Degree	Bachelor's Degree
Advanced Label	-0.0014	-0.0022	0.0008	-0.0047	-0.0046	0.0063
	(0.0042)	(0.0040)	(0.0061)	(0.0030)	(0.0063)	(0.0054)
Advanced * Economically Disadvantaged	-0.1093***	-0.0360	-0.0502*	-0.0113	-0.0769**	-0.0408
	(0.0256)	(0.0252)	(0.0243)	(0.0174)	(0.0244)	(0.0269)
P(Advanced + Interaction)	0.01	0.18	0.05	0.32	0.01	0.22
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Bandwidth	3	8	3	8	3	8
Number of Observations	25,418	63,456	25,418	63,456	25,418	63,456
Mean Outcome Not Economically Disadvantaged	0.73	0.73	0.09	0.09	0.66	0.66
Mean Outcome Economically Disadvantaged	0.51	0.51	0.12	0.12	0.41	0.41

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Bandwidth is measured in scale score points. Mean outcomes are for students in the Math Proficient/Advanced Sample. All regressions include year fixed effects. Standard errors are clustered at the year level.

A.1.2 Reading Proficient/Advanced Sample

Table A.5 – Effect of Receiving a Higher Label Postsecondary Outcomes Reading Proficient/Advanced Sample

	Any Postsecondary Enrollment	Any Postsecondary Enrollment	2-Year Enrollment	2-Year Enrollment	4-Year Enrollment	4-Year Enrollment
Advanced Label	-0.0022	-0.0027**	-0.0049	-0.0041	0.0006	-0.0021
	(0.0032)	(0.0009)	(0.0070)	(0.0049)	(0.0063)	(0.0034)
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Covariates	Y	Y	Y	Y	Y	Y
Bandwidth	5	14	5	14	5	14
Number of Observations	72,167	182,594	72,167	182,594	72,167	182,594
Mean Outcome	0.90	0.90	0.58	0.58	0.73	0.73
	Any Postsecondary Degree	Any Postsecondary Degree	Associate Degree	Associate Degree	Bachelor's Degree	Bachelor's Degree
Advanced Label	-0.0088	-0.0032	0.0084*	0.0051	-0.0094	-0.0065
	(0.0058)	(0.0037)	(0.0042)	(0.0039)	(0.0076)	(0.0037)
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Covariates	Y	Y	Y	Y	Y	Y
Bandwidth	5	14	5	14	5	14
Number of Observations	72,167	182,594	72,167	182,594	72,167	182,594
Mean Outcome	0.56	0.56	0.11	0.11	0.46	0.46

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Bandwidth is measured in scale score points. Standard errors are clustered at the year level. Mean outcomes for the Reading Proficient/Advanced Sample are shown.

Table A.6 – Effect of Receiving a Higher Label Postsecondary Outcomes Reading Proficient/Advanced Sample Male and Female Students

	Any Postsecondary Enrollment	Any Postsecondary Enrollment	2-Year Enrollment	2-Year Enrollment	4-Year Enrollment	4-Year Enrollment
Advanced Label	0.0017	-0.0013	-0.0086	-0.0026	0.0105	0.0025
	(0.0047)	(0.0027)	(0.0085)	(0.0056)	(0.0061)	(0.0029)
Advanced * Female	-0.0066	-0.0025	0.0062	-0.0026	-0.0163	-0.0082
	(0.0069)	(0.0054)	(0.0098)	(0.0061)	(0.0100)	(0.0071)
P(Advanced + Interaction)	0.36	0.24	0.79	0.40	0.56	0.37
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Bandwidth	5	14	5	14	5	14
Number of Observations	72,167	182,594	72,167	182,594	72,167	182,594
Mean Outcome Males	0.88	0.88	0.55	0.55	0.69	0.69
Mean Outcome Females	0.92	0.92	0.60	0.60	0.77	0.77
	Any Postsecondary Degree	Any Postsecondary Degree	Associate Degree	Associate Degree	Bachelor's Degree	Bachelor's Degree
Advanced Label	0.0061	0.0014	0.0145**	0.0050	-0.0012	-0.0032
	(0.0033)	(0.0042)	(0.0055)	(0.0052)	(0.0074)	(0.0059)
Advanced * Female	-0.0248**	-0.0089	-0.0116	0.0002	-0.0120	-0.0065
	(0.0088)	(0.0086)	(0.0086)	(0.0053)	(0.0073)	(0.0094)
P(Advanced + Interaction)	0.11	0.26	0.67	0.27	0.25	0.14
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Bandwidth	5	14	5	14	5	14
Number of Observations	72,167	182,594	72,167	182,594	72,167	182,594
Mean Outcome Males	0.50	0.50	0.10	0.10	0.41	0.41
Mean Outcome Females	0.61	0.61	0.13	0.13	0.50	0.50

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Bandwidth is measured in scale score points. Mean outcomes are for students in the Reading Proficient/Advanced Sample. All regressions include year fixed effects. Standard errors are clustered at the year level.

Table A.7 – Effect of Receiving a Higher Label Postsecondary Outcomes Reading Proficient/Advanced Sample White and Black Students

	Any Postsecondary Enrollment	Any Postsecondary Enrollment	2-Year Enrollment	2-Year Enrollment	4-Year Enrollment	4-Year Enrollment
Advanced Label	-0.0002	-0.0025*	-0.0014	-0.0037	0.0011	-0.0024
	(0.0032)	(0.0011)	(0.0065)	(0.0050)	(0.0055)	(0.0027)
Advanced * Black	-0.0185	-0.0250	0.0195	-0.0143	-0.0240	-0.0374**
	(0.0247)	(0.0154)	(0.0337)	(0.0164)	(0.0233)	(0.0112)
P(Advanced + Interaction)	0.50	0.12	0.61	0.27	0.40	0.02
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Bandwidth	5	14	5	14	5	14
Number of Observations	66,775	168,735	66,775	168,735	66,775	168,735
Mean Outcome White Students	0.90	0.90	0.58	0.58	0.73	0.73
Mean Outcome Black Students	0.91	0.91	0.61	0.61	0.72	0.72
	Any Postsecondary Degree	Any Postsecondary Degree	Associate Degree	Associate Degree	Bachelor's Degree	Bachelor's Degree
Advanced Label	-0.0087	-0.0037	0.0098*	0.0069	-0.0086	-0.0074
	(0.0064)	(0.0033)	(0.0048)	(0.0037)	(0.0086)	(0.0040)
Advanced * Black	-0.0171	-0.0348**	-0.0203	-0.0293**	-0.0175	-0.0221
	(0.0280)	(0.0136)	(0.0139)	(0.0106)	(0.0327)	(0.0175)
P(Advanced + Interaction)	0.42	0.03	0.48	0.10	0.50	0.14
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Bandwidth	5	14	5	14	5	14
Number of Observations	66,775	168,735	66,775	168,735	66,775	168,735
Mean Outcome White Students	0.57	0.57	0.12	0.12	0.47	0.47
Mean Outcome Black Students	0.39	0.39	0.07	0.07	0.31	0.31

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Bandwidth is measured in scale score points. Mean outcomes are for students in the Reading Proficient/Advanced Sample. All regressions include year fixed effects. Standard errors are clustered at the year level.

Table A.8 – Effect of Receiving a Higher Label Postsecondary Outcomes Reading Proficient/Advanced Sample Not Economically Disadvantaged and Economically Disadvantaged Students

	Any Postsecondary Enrollment	Any Postsecondary Enrollment	2-Year Enrollment	2-Year Enrollment	4-Year Enrollment	4-Year Enrollment
Advanced Label	-0.0010	-0.0032*	-0.0023	-0.0037	0.0033	-0.0004
	(0.0036)	(0.0013)	(0.0040)	(0.0046)	(0.0064)	(0.0038)
Advanced * Economically Disadvantaged	-0.0066	0.0046	-0.0144	0.0008	-0.0152	-0.0091
	(0.0133)	(0.0067)	(0.0232)	(0.0098)	(0.0169)	(0.0106)
P(Advanced + Interaction)	0.55	0.81	0.54	0.80	0.49	0.36
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Bandwidth	5	14	5	14	5	14
Number of Observations	72,167	182,594	72,167	182,594	72,167	182,594
Mean Outcome Not Economically Disadvantaged	0.93	0.93	0.58	0.58	0.78	0.78
Mean Outcome Economically Disadvantaged	0.82	0.82	0.58	0.58	0.58	0.58
	Any Postsecondary Degree	Any Postsecondary Degree	Associate Degree	Associate Degree	Bachelor's Degree	Bachelor's Degree
Advanced Label	-0.0034	-0.0022	0.0089	0.0038	-0.0035	-0.0038
	(0.0075)	(0.0026)	(0.0053)	(0.0040)	(0.0096)	(0.0040)
Advanced * Economically Disadvantaged	-0.0289	-0.0082	-0.0020	0.0083	-0.0318*	-0.0181
	(0.0159)	(0.0143)	(0.0120)	(0.0095)	(0.0154)	(0.0109)
P(Advanced + Interaction)	0.04	0.50	0.47	0.21	0.02	0.08
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Bandwidth	5	14	5	14	5	14
Number of Observations	72,167	182,594	72,167	182,594	72,167	182,594
Mean Outcome Not Economically Disadvantaged	0.62	0.62	0.11	0.11	0.53	0.53
Mean Outcome Economically Disadvantaged	0.35	0.35	0.11	0.11	0.25	0.25

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Bandwidth is measured in scale score points. Mean outcomes are for students in the Reading Proficient/Advanced Sample. All regressions include year fixed effects. Standard errors are clustered at the year level.

A.1.3 Math Not Proficient/Partially Proficient Sample

Table A.9 – Effect of Receiving a Higher Label Postsecondary Outcomes Math Not Proficient/Partially Proficient Sample

	Any Postsecondary Enrollment	Any Postsecondary Enrollment	2-Year Enrollment	2-Year Enrollment	4-Year Enrollment	4-Year Enrollment
Partially Proficient Label	-0.0113*	-0.0010	-0.0140*	-0.0001	-0.0057	-0.0028
	(0.0057)	(0.0028)	(0.0062)	(0.0035)	(0.0065)	(0.0024)
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Covariates	Y	Y	Y	Y	Y	Y
Bandwidth	4	11	4	11	4	11
Number of Observations	88,885	222,002	88,885	222,002	88,885	222,002
Mean Outcome	0.75	0.75	0.59	0.59	0.45	0.45
	Any Postsecondary Degree	Any Postsecondary Degree	Associate Degree	Associate Degree	Bachelor's Degree	Bachelor's Degree
Partially Proficient Label	-0.0011	0.0045	-0.0001	0.0018	-0.0074*	-0.0004
	(0.0063)	(0.0034)	(0.0042)	(0.0029)	(0.0032)	(0.0030)
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Covariates	Y	Y	Y	Y	Y	Y
Bandwidth	4	11	4	11	4	11
Number of Observations	88,885	222,002	88,885	222,002	88,885	222,002
Mean Outcome	0.30	0.30	0.11	0.11	0.19	0.19

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Bandwidth is measured in scale score points. Standard errors are clustered at the year level. Mean outcomes for the Math Not Proficient/Partially Proficient Sample are shown.

Table A.10 – Effect of Receiving a Higher Label Postsecondary Outcomes Math Not Proficient/Partially Proficient Sample Male and Female Students

	Any Postsecondary Enrollment	Any Postsecondary Enrollment	2-Year Enrollment	2-Year Enrollment	4-Year Enrollment	4-Year Enrollment
Partially Proficient Label	-0.0143	-0.0022	-0.0072	-0.0013	-0.0125*	0.0008
	(0.0078)	(0.0073)	(0.0106)	(0.0069)	(0.0062)	(0.0051)
Partially Proficient * Female	0.0035	0.0031	-0.0140	0.0031	0.0097	-0.0065
	(0.0086)	(0.0097)	(0.0104)	(0.0084)	(0.0105)	(0.0111)
P(Advanced + Interaction)	0.15	0.80	0.01	0.64	0.77	0.43
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Bandwidth	4	11	4	11	4	11
Number of Observations	88,885	222,002	88,885	222,002	88,885	222,002
Mean Outcome Males	0.68	0.68	0.54	0.54	0.38	0.38
Mean Outcome Females	0.81	0.81	0.63	0.63	0.52	0.52
	Any Postsecondary Degree	Any Postsecondary Degree	Associate Degree	Associate Degree	Bachelor's Degree	Bachelor's Degree
Partially Proficient Label	0.0078	0.0156**	0.0084**	0.0098**	-0.0063	0.0022
	(0.0101)	(0.0052)	(0.0026)	(0.0031)	(0.0043)	(0.0030)
Partially Proficient * Female	-0.0184	-0.0201*	-0.0162**	-	-0.0036	-0.0045
	(0.0139)	(0.0093)	(0.0051)	0.0144***	(0.0084)	(0.0078)
				(0.0026)		
P(Advanced + Interaction)	0.26	0.46	0.28	0.15	0.14	0.72
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Bandwidth	4	11	4	11	4	11
Number of Observations	88,885	222,002	88,885	222,002	88,885	222,002
Mean Outcome Males	0.23	0.23	0.09	0.09	0.14	0.14
Mean Outcome Females	0.37	0.37	0.13	0.13	0.24	0.24

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Bandwidth is measured in scale score points. Mean outcomes are for students in the Math Not Proficient/Partially Proficient Sample. All regressions include year fixed effects. Standard errors are clustered at the year level.

Table A.11 – Effect of Receiving a Higher Label Postsecondary Outcomes Math Not Proficient/Partially Proficient Sample White and Black Students

	Any Postsecondary Enrollment	Any Postsecondary Enrollment	2-Year Enrollment	2-Year Enrollment	4-Year Enrollment	4-Year Enrollment
Partially Proficient Label	-0.0162**	-0.0022	-0.0204*	-0.0022	-0.0137***	-0.0060**
	(0.0064)	(0.0040)	(0.0088)	(0.0044)	(0.0034)	(0.0023)
Partially Proficient * Black	0.0245*	0.0050	0.0218	0.0060	0.0508**	0.0146*
	(0.0120)	(0.0073)	(0.0156)	(0.0105)	(0.0140)	(0.0068)
P(Advanced + Interaction)	0.43	0.62	0.93	0.73	0.05	0.35
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Bandwidth	4	11	4	11	4	11
Number of Observations	80,870	202,414	80,0.58870	202,414	80,870	202,414
Mean Outcome White Students	0.74	0.74	0.58	0.58	0.46	0.46
Mean Outcome Black Students	0.78	0.78	0.62	0.62	0.45	0.45
	Any Postsecondary Degree	Any Postsecondary Degree	Associate Degree	Associate Degree	Bachelor's Degree	Bachelor's Degree
Partially Proficient Label	-0.0059	0.0031	-0.0042	0.0003	-0.0098	-0.0006
	(0.0051)	(0.0025)	(0.0032)	(0.0022)	(0.0053)	(0.0039)
Partially Proficient * Black	0.0299**	0.0031	0.0155	0.0024	0.0216	0.0047
	(0.0088)	(0.0107)	(0.0084)	(0.0083)	(0.0116)	(0.0049)
P(Advanced + Interaction)	0.05	0.55	0.28	0.79	0.18	0.46
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Bandwidth	4	11	4	11	4	11
Number of Observations	80,870	202,414	80,870	202,414	80,870	202,414
Mean Outcome White Students	0.33	0.33	0.12	0.12	0.22	0.22
Mean Outcome Black Students	0.20	0.20	0.06	0.06	0.13	0.13

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Bandwidth is measured in scale score points. Mean outcomes are for students in the Math Not Proficient/Partially Proficient Sample. All regressions include year fixed effects. Standard errors are clustered at the year level.

Table A.12 – Effect of Receiving a Higher Label Postsecondary Outcomes Math Not Proficient/Partially Proficient Sample Not Economically Disadvantaged and Economically Disadvantaged Students

	Any Postsecondary Enrollment	Any Postsecondary Enrollment	2-Year Enrollment	2-Year Enrollment	4-Year Enrollment	4-Year Enrollment
Partially Proficient Label	-0.0085	-0.0010	-0.0157**	-0.0034	0.0003	0.0051
	(0.0069)	(0.0043)	(0.0061)	(0.0047)	(0.0071)	(0.0036)
Partially Proficient * Economically Disadvantaged	-0.0092	-0.0019	0.0031	0.0065	-0.0187	-0.0215***
	(0.0092)	(0.0046)	(0.0087)	(0.0056)	(0.0113)	(0.0057)
P(Advanced + Interaction)	0.04	0.29	0.22	0.50	0.05	0.00
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Bandwidth	4	11	4	11	4	11
Number of Observations	88,885	222,002	88,885	222,002	88,885	222,002
Mean Outcome Not Economically Disadvantaged	0.80	0.80	0.62	0.62	0.52	0.52
Mean Outcome Economically Disadvantaged	0.67	0.67	0.54	0.54	0.35	0.35
	Any Postsecondary Degree	Any Postsecondary Degree	Associate Degree	Associate Degree	Bachelor's Degree	Bachelor's Degree
Partially Proficient Label	0.0013	0.0067	-0.0013	0.0019	-0.0069	0.0007
	(0.0080)	(0.0048)	(0.0072)	(0.0033)	(0.0053)	(0.0042)
Partially Proficient * Economically Disadvantaged	-0.0077	-0.0072	0.0029	-0.0015	-0.0032	-0.0034
	(0.0092)	(0.0076)	(0.0089)	(0.0049)	(0.0096)	(0.0049)
P(Advanced + Interaction)	0.29	0.92	0.68	0.93	0.14	0.44
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Bandwidth	4	11	4	11	4	11
Number of Observations	88,885	222,002	88,885	222,002	88,885	222,002
Mean Outcome Not Economically Disadvantaged	0.38	0.38	0.13	0.13	0.26	0.26
Mean Outcome Economically Disadvantaged	0.19	0.19	0.08	0.08	0.10	0.10

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Bandwidth is measured in scale score points. Mean outcomes are for students in the Math Not Proficient/Partially Proficient Sample. All regressions include year fixed effects. Standard errors are clustered at the year level.

A.1.4 Reading Not Proficient/Partially Proficient Sample

Table A.13 – Effect of Receiving a Higher Label Postsecondary Outcomes Reading Not Proficient/Partially Proficient Sample

	Any Postsecondary Enrollment	Any Postsecondary Enrollment	2-Year Enrollment	2-Year Enrollment	4-Year Enrollment	4-Year Enrollment
Partially Proficient Label	-0.0046	-0.0004	-0.0044	0.0012	-0.0039	-0.0049
	(0.0112)	(0.0077)	(0.0121)	(0.0083)	(0.0100)	(0.0041)
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Covariates	Y	Y	Y	Y	Y	Y
Bandwidth	5	14	5	14	5	14
Number of Observations	55,190	148,864	55,190	148,864	55,190	148,864
Mean Outcome	0.68	0.68	0.56	0.56	0.36	0.36
	Any Postsecondary Degree	Any Postsecondary Degree	Associate Degree	Associate Degree	Bachelor's Degree	Bachelor's Degree
Partially Proficient Label	-0.0000	-0.0055	-0.0053	-0.0009	0.0082	-0.0016
	(0.0068)	(0.0036)	(0.0047)	(0.0035)	(0.0047)	(0.0026)
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Covariates	Y	Y	Y	Y	Y	Y
Bandwidth	5	14	5	14	5	14
Number of Observations	55,190	148,864	55,190	148,864	55,190	148,864
Mean Outcome	0.23	0.23	0.09	0.09	0.13	0.13

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Bandwidth is measured in scale score points. Standard errors are clustered at the year level. Mean outcomes for the Reading Not Proficient/Partially Proficient Sample are shown.

Table A.14 – Effect of Receiving a Higher Label Postsecondary Outcomes Reading Not Proficient/Partially Proficient Sample Male and Female Students

	Any Postsecondary Enrollment	Any Postsecondary Enrollment	2-Year Enrollment	2-Year Enrollment	4-Year Enrollment	4-Year Enrollment
Partially Proficient Label	-0.0071	-0.0065	-0.0022	-0.0016	-0.0101	-0.0143*
	(0.0125)	(0.0100)	(0.0151)	(0.0111)	(0.0109)	(0.0071)
Partially Proficient * Female	0.0015	0.0087	-0.0080	0.0020	0.0101	0.0173*
	(0.0080)	(0.0072)	(0.0096)	(0.0093)	(0.0143)	(0.0073)
P(Advanced + Interaction)	0.68	0.78	0.42	0.96	1.00	0.44
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Bandwidth	5	14	5	14	5	14
Number of Observations	55,190	148,864	55,190	148,864	55,190	148,864
Mean Outcome Males	0.63	0.63	0.51	0.51	0.32	0.32
Mean Outcome Females	0.74	0.74	0.61	0.61	0.41	0.41
	Any Postsecondary Degree	Any Postsecondary Degree	Associate Degree	Associate Degree	Bachelor's Degree	Bachelor's Degree
Partially Proficient Label	-0.0061	-0.0066	-0.0008	0.0015	-0.0000	-0.0034
	(0.0140)	(0.0046)	(0.0087)	(0.0033)	(0.0043)	(0.0042)
Partially Proficient * Female	0.0097	0.0019	-0.0108	-0.0048	0.0157	0.0030
	(0.0209)	(0.0079)	(0.0099)	(0.0049)	(0.0145)	(0.0060)
P(Advanced + Interaction)	0.72	0.49	0.02	0.54	0.23	0.93
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Bandwidth	5	14	5	14	5	14
Number of Observations	55,190	148,864	55,190	148,864	55,190	148,864
Mean Outcome Males	0.19	0.19	0.08	0.08	0.11	0.11
Mean Outcome Females	0.27	0.27	0.11	0.11	0.15	0.15

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Bandwidth is measured in scale score points. Mean outcomes are for students in the Reading Not Proficient/Partially Proficient Sample. All regressions include year fixed effects. Standard errors are clustered at the year level.

Table A.15 – Effect of Receiving a Higher Label Postsecondary Outcomes Reading Not Proficient/Partially Proficient Sample White and Black Students

	Any Postsecondary Enrollment	Any Postsecondary Enrollment	2-Year Enrollment	2-Year Enrollment	4-Year Enrollment	4-Year Enrollment
Partially Proficient Label	-0.0029	-0.0009	0.0001	0.0035	-0.0102	-0.0103**
	(0.0124)	(0.0076)	(0.0157)	(0.0081)	(0.0095)	(0.0041)
Partially Proficient * Black	-0.0120	0.0030	-0.0171	-0.0078	0.0167	0.0156
	(0.0173)	(0.0056)	(0.0230)	(0.0095)	(0.0149)	(0.0098)
P(Advanced + Interaction)	0.44	0.84	0.44	0.73	0.64	0.64
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Bandwidth	5	14	5	14	5	14
Number of Observations	49,821	134,602	49,821	134,602	49,821	134,602
Mean Outcome White Students	0.67	0.67	0.54	0.54	0.36	0.36
Mean Outcome Black Students	0.74	0.74	0.62	0.62	0.37	0.37
	Any Postsecondary Degree	Any Postsecondary Degree	Associate Degree	Associate Degree	Bachelor's Degree	Bachelor's Degree
Partially Proficient Label	0.0048	-0.0049	-0.0070	-0.0018	0.0092*	-0.0013
	(0.0087)	(0.0046)	(0.0075)	(0.0037)	(0.0047)	(0.0014)
Partially Proficient * Black	-0.0166	-0.0054	-0.0064	-0.0041	-0.0014	-0.0015
	(0.0095)	(0.0081)	(0.0095)	(0.0056)	(0.0107)	(0.0059)
P(Advanced + Interaction)	0.06	0.26	0.05	0.44	0.41	0.67
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Bandwidth	5	14	5	14	5	14
Number of Observations	49,821	134,602	49,821	134,602	49,821	134,602
Mean Outcome White Students	0.26	0.26	0.11	0.11	0.15	0.15
Mean Outcome Black Students	0.16	0.16	0.05	0.05	0.09	0.09

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Bandwidth is measured in scale score points. Mean outcomes are for students in the Reading Not Proficient/Partially Proficient Sample. All regressions include year fixed effects. Standard errors are clustered at the year level.

Table A.16 – Effect of Receiving a Higher Label Postsecondary Outcomes Reading Not Proficient/Partially Proficient Sample Not Economically Disadvantaged and Economically Disadvantaged Students

	Any Postsecondary Enrollment	Any Postsecondary Enrollment	2-Year Enrollment	2-Year Enrollment	4-Year Enrollment	4-Year Enrollment
Partially Proficient Label	-0.0041	0.0106	-0.0150	0.0077	0.0058	0.0001
	(0.0119)	(0.0072)	(0.0136)	(0.0082)	(0.0102)	(0.0068)
Partially Proficient * Economically Disadvantaged	-0.0015	-0.0238*	0.0204*	-0.0148	-0.0192	-0.0102
	(0.0115)	(0.0101)	(0.0103)	(0.0110)	(0.0148)	(0.0123)
P(Advanced + Interaction)	0.73	0.31	0.72	0.57	0.41	0.29
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Bandwidth	5	14	5	14	5	14
Number of Observations	55,190	148,864	55,190	148,864	55,190	148,864
Mean Outcome Not Economically Disadvantaged	0.74	0.74	0.60	0.60	0.42	0.42
Mean Outcome Economically Disadvantaged	0.62	0.62	0.51	0.51	0.29	0.29
	Any Postsecondary Degree	Any Postsecondary Degree	Associate Degree	Associate Degree	Bachelor's Degree	Bachelor's Degree
Partially Proficient Label	0.0013	0.0015	-0.0113	-0.0021	0.0087	0.0049
	(0.0085)	(0.0046)	(0.0064)	(0.0049)	(0.0063)	(0.0030)
Partially Proficient * Economically Disadvantaged	-0.0023	-0.0122	0.0123	0.0037	-0.0011	-0.0117*
	(0.0092)	(0.0091)	(0.0089)	(0.0066)	(0.0063)	(0.0053)
P(Advanced + Interaction)	0.91	0.19	0.89	0.75	0.16	0.16
Year Fixed Effects	Y	Y	Y	Y	Y	Y
Bandwidth	5	14	5	14	5	14
Number of Observations	55,190	148,864	55,190	148,864	55,190	148,864
Mean Outcome Not Economically Disadvantaged	0.30	0.30	0.12	0.12	0.19	0.19
Mean Outcome Economically Disadvantaged	0.15	0.15	0.07	0.07	0.07	0.07

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Bandwidth is measured in scale score points. Mean outcomes are for students in the Reading Not Proficient/Partially Proficient Sample. All regressions include year fixed effects. Standard errors are clustered at the year level.

A.2 Summary of Significant Treatment Effects

Table A.17 – List of Higher Label Treatment Effects Significant at the 5% Level

Sample	Group of Students	Outcome	Bandwidth
Math Proficient/Advanced	All	Associate Degree	8
Math Proficient/Advanced	Female	Any Postsecondary Degree	5
Math Proficient/Advanced	Female	Associate Degree	3, 8
Math Proficient/Advanced	Female	Bachelor's Degree	3, 5
Math Proficient/Advanced	White	Associate Degree	8
Math Proficient/Advanced	Black	4-Year Enrollment	3
Math Proficient/Advanced	Black	Any Postsecondary Degree	3
Math Proficient/Advanced	Black	Bachelor's Degree	3
Math Proficient/Advanced	Economically Disadvantaged	Any Postsecondary Degree	3
Math Proficient/Advanced	Economically Disadvantaged	Associate Degree	3
Math Proficient/Advanced	Economically Disadvantaged	Bachelor's Degree	3
Reading Proficient/Advanced	All	Any Postsecondary Enrollment	14
Reading Proficient/Advanced	All	Associate Degree	9
Reading Proficient/Advanced	Male	Associate Degree	5, 9
Reading Proficient/Advanced	White	Associate Degree	9
Reading Proficient/Advanced	Black	4-Year Enrollment	14
Reading Proficient/Advanced	Black	Any Postsecondary Degree	14
Reading Proficient/Advanced	Not Economically Disadvantaged	Associate Degree	9
Reading Proficient/Advanced	Economically Disadvantaged	Any Postsecondary Degree	5
Reading Proficient/Advanced	Economically Disadvantaged	Bachelor's Degree	5, 9
Math Not Proficient/Partially Proficient	Male	Any Postsecondary Degree	7, 11
Math Not Proficient/Partially Proficient	Male	Associate Degree	4, 7, 11
Math Not Proficient/Partially Proficient	Female	2-Year Enrollment	4
Math Not Proficient/Partially Proficient	Female	Associate Degree	7

Math Not Proficient/Partially Proficient	White	Any Postsecondary Enrollment	4
Math Not Proficient/Partially Proficient	White	4-Year Enrollment	4, 11
Math Not Proficient/Partially Proficient	Black	4-Year Enrollment	4
Math Not Proficient/Partially Proficient	Black	Any Postsecondary Degree	4
Math Not Proficient/Partially Proficient	Not Economically Disadvantaged	2-Year Enrollment	4
Math Not Proficient/Partially Proficient	Economically Disadvantaged	Any Postsecondary Enrollment	4
Math Not Proficient/Partially Proficient	Economically Disadvantaged	4-Year Enrollment	4, 11
Reading Not Proficient/Partially Proficient	Female	Associate Degree	5
Reading Not Proficient/Partially Proficient	White	4-Year Enrollment	14
Reading Not Proficient/Partially Proficient	Black	Associate Degree	5