sociological science

Supplement to:

Thurston Domina, Andrew M. Penner, and Emily K. Penner. 2016. "'Membership Has Its Privileges': Status Incentives and Categorical Inequality in Education." Sociological Science 3: 264-295.

Color-coded ID card program design and implementation

Program timeline

Informal interviews with officials at Live Oak and Mann as well as Sudden Valley Unified indicate that principals at the two high schools designed the color-coded ID program together in the spring of 2010. The principals report that while they discussed the program with students and teachers that spring, they did not finalize the prize system's rules and fully roll the program out until the students returned for classes in the fall of 2010. The principals also reported that the program was popular with teachers, students, and parents in its first year. We are unable to assess this claim empirically. On one hand, some teachers report voicing qualms about the program from the beginning. On the other, we note that we find no mention of the program in either schools' student newspapers in the 2010-11 school year, and the student newspaper coverage of the program in later school years was largely positive.

Nonetheless, controversy erupted over the use of color-coded ID cards at Live Oak and Mann in the fall of 2011 when a local newspaper published an article that described the program and quoted students who reported experiencing stigma associated with having a white ID card. Administrators at Live Oak and Mann defended the program as a motivational tool. However, community groups raised concerns about racial inequalities in access to the program's incentives, students complained that the program fostered "condescending attitudes," and the State Department of Education ruled that the program violated regulations protecting student privacy. Live Oak and Mann eliminated the color-coded IDs in the wake of this controversy, issuing standard white IDs to all students, but allowing students to keep (and to continue displaying) their gold or platinum ID cards. School newspaper reports and school officials further indicate that both Live Oak and Mann kept the incentive structure associated with the program intact throughout this second program year.

As noted in the paper, California students take CSTs each spring from second grade through eleventh grade. High school students take tests in up to four subjects based on the math, English language arts, history, and science classes they were enrolled in (except seniors, who do not take CSTs). The scores are reported in a series of five performance bands: advanced, proficient, basic, below basic, and far below basic. Students could earn a platinum ID card by scoring "advanced" on each of their CSTs, and a gold card by scoring "proficient" or "advanced" on all CSTs, without having dropped down from "advanced" in any subject. In addition, the ID card program at both schools included a "growth" clause that allowed students to earn a gold card by improving their performance band. To qualify for a gold card via growth, students had to move up a performance band on two or more CSTs, again without dropping a performance band on any subject. All other students received white cards. In order to create strong incentives for students on the CSTs, teachers were encouraged to offer to change students' grades based on their CST performance, so that any student who scored proficient or higher on the CST would receive a passing grade in the class, regardless of his or her performance in the class otherwise.

Differences in access to gold and platinum ID cards

The public controversy surrounding the Live Oak and Mann student ID card programs focused on racial inequalities in access to high-status gold and platinum ID cards. As Table S1 indicates, Asian students are substantially over-represented among gold and platinum ID card holders, while Hispanic students are substantially under-represented. These disturbing racial inequalities would be particularly troubling if they resulted from discrimination within the program, rather than inequalities in student achievement and achievement growth trajectories. The first analyses in this online supplement therefore consider the extent to which racial and other inequalities in gold and platinum ID receipt persist after controlling for student achievement.

Table S1 reports a series of multinomial logistic regression models that investigate inequalities in card placement in a multivariate context. Model 1 indicates that student race, gender, free/reduced lunch status, English-language status, grade, and high school are associated with students' odds of earning a gold or platinum card. Furthermore, model 2 indicates that many of these associations persist even after controlling for prior math and ELA achievement. The prior achievement controls included in this model are student scores on the ELA and math CST from the test prior to the test upon which ID card assignment is based. The model also includes quadratics and dummies for student proficiency levels on these twice-lagged exams to address potential nonlinearities in the relationship between prior achievement and card placement. The inclusion of these prior achievement controls substantially improves these models' predictive power, increasing the pseudo r-squared from 0.08 to 0.34. However, race and gender inequalities in ID placement odds persist even after the inclusion of these controls. Net of prior achievement, Asians are significantly more likely to earn gold and platinum cards than their Hispanic peers. Holding all else at the sample mean, the predicted rate of gold card receipt is 49 percent for Asians versus 38 percent for Hispanics. Boys are also more likely than girls to earn gold or platinum cards net of prior achievement controls. Indeed, this conditional gender gap is more pronounced than the gender gap in model 1.

These findings would be particularly troubling if they pointed to racial or gender discrimination in the ID card placement system that operated independently of test scores. However, the multivariate models reported in Table S2 point to a less sinister interpretation. The dependent variable in these models is a proxy for student card placement based on student CST performance band level. For Live Oak and Mann students, this variable closely corresponds to actual card placement. Approximately 87 percent of students receive the same value on this proxy as on the actual card placement variable.[1] The results reported in Table S2's first model

Table S1: Multinomial logistic regression – odds of receiving a gold or platinum card (2012 Live Oak and Mann High School students)

	Model 1		Model 2	
	Gold	Platinum	Gold	Platinum
Asian	2.78***	7.49***	1.53***	2.26***
	(0.26)	(1.12)	(0.18)	(0.48)
Black/other	0.98	0.69	0.94	0.65
	(0.17)	(0.27)	(0.20)	(0.31)
White	1.78***	2.26***	1.40**	1.58*
	(0.17)	(0.38)	(0.17)	(0.36)
Male	1.07	1.30**	1.35***	1.80***
	(0.07)	(0.12)	(0.11)	(0.24)
Free/reduced lunch	0.72***	0.61***	1	1.16
	(0.06)	(0.07)	(0.10)	(0.19)
Non-native English	1.41***	1.61***	1.28*	1.26
	(0.12)	(0.18)	(0.14)	(0.22)
10^{th} grade	1.80***	2.10***	1.66	1.47
	(0.17)	(0.25)	(0.52)	(0.54)
11^{th} grade	1.58***	1.03	0.53	0.09***
	(0.15)	(0.14)	(0.17)	(0.04)
12^{th} grade	1.39***	0.69**	0.6	0.12***
	(0.13)	(0.10)	(0.20)	(0.05)
Mann HS	1.23**	0.69***	1.48***	0.88
	(0.08)	(0.07)	(0.13)	(0.12)
Math CST			3.90***	23.27***
			(0.94)	(11.38)
ELA CST			5.87***	42.86***
			(1.63)	(25.47)
Math CST^2			0.91	0.75*
			(0.09)	(0.11)
ELA CST^2			0.74*	0.53**
			(0.09)	(0.10)
Math proficiency level			X	X
ELA proficiency level			X	X
Math test dummies			X	X
R-squared	0.08		0.34	
N	4,966		4,250	

Note : All models use school-clustered standard errors to account for school clustering.
* p<0.05; ** p<0.01; *** p<0.001

Table S2: Multinomial logistic regression – odds of meeting gold or platinum card proxy (all 2012 Sudden Valley Unified School District high school students)

	Model 1: Card proxy (Live Oak & Mann)		Model 2: Card Proxy (Other SVUSD HS)	
	Gold	Platinum	Gold	Platinum
Asian	1.55***	2.17***	1.27**	2.02***
	(0.19)	(0.42)	(0.11)	(0.31)
Black/other	1.14	0.77	0.94	0.84
	(0.24)	(0.32)	(0.19)	(0.42)
White	1.37*	1.53*	0.98	1.21
	(0.17)	(0.31)	(0.11)	(0.27)
Male	0.98	1.33*	1.20**	1.35**
	(0.09)	(0.17)	(0.07)	(0.16)
Free/reduced lunch	0.95	1.14	0.86*	0.94
	(0.10)	(0.17)	(0.07)	(0.13)
Non-native English	1.40**	1.61**	0.96	1.12
	(0.16)	(0.26)	(0.08)	(0.18)
10^{th} grade	2.10*	1.85	0.66*	0.44**
	(0.72)	(0.72)	(0.11)	(0.12)
11^{th} grade	0.77	0.21***	0.25***	0.12***
	(0.27)	(0.09)	(0.05)	(0.04)
12^{th} grade	0.44*	0.09***	0.15***	0.06***
	(0.16)	(0.04)	(0.03)	(0.02)
Math CST	3.26***	10.29***	4.45***	4.23***
	(0.86)	(4.56)	(0.77)	(1.56)
ELA CST	3.83***	10.21***	4.73***	1.34
	(1.17)	(5.37)	(0.86)	(0.39)
Math CST^2	1.12	1.09	0.92	1.25
	(0.16)	(0.18)	(0.08)	(0.15)
ELA CST^2	1.55***	2.17***	1.27**	2.02***
	(0.19)	(0.42)	(0.11)	(0.31)
Math proficiency level	X	X	X	X
ELA proficiency level	X	X	X	X
Math test dummies	X	X	X	X
School dummies	X	X	X	X
R-squared	0.38		0.45	
N	4,250		13,977	

Note : All models use school-clustered standard errors to account for school clustering.

* p<0.05; ** p<0.01; *** p<0.001

indicate that the inequalities observed on the actual placement are also present on the placement proxy. This finding suggests that disparities between reported ID card placement criteria and actual ID card placements are not responsible for racial and gender inequalities in ID card placement; that is, the differences across race and gender that we observe do not appear to be a function of bias in whether the ID card rules were followed. Furthermore, Table S2's second model indicates that similar inequalities are apparent in the placement proxy in the other district high schools that do not have a color-coded ID system. This finding indicates racial inequalities in ID card placement occur because Asians are more likely than Hispanics to score more highly than one would expect even after accounting for their prior achievement. Similarly, the findings in Table S2 indicate that boys are more likely to earn high-status IDs than girls, in large part because they are more likely to outperform their prior achievement.

Tables S3 and S4 provide the full model results corresponding to Tables 4 and 5 in the paper. Table S5 presents additional significance tests for the coefficients presented in Table 5.

Regression discontinuity analyses to estimate the effects of ID card assignment

The process of assigning color-coded ID cards at Live Oak and Mann High School utilized arbitrary cut-points on the continuous California Standards Tests scores to split students into discrete, ordered social categories. Our analyses take advantage of this formulaic assignment process to estimate the effect of receiving a low-status white ID card. The CSTs, like all standardized tests, measure student skills in relevant content areas with error. Assuming that this error is randomly distributed at the ID card assignment threshold, whether a student's lowest score is above or below the threshold can be viewed as a matter of chance for students near the threshold. To the extent that this assumption holds, our regression discontinuity analyses provide unbiased estimates of the causal effect of white card assignment. In this online supplement, we describe a series of sensitivity analyses that we have conducted to test this assumption and assess the validity of the estimates of the effects of white card assignment that we report in Table 6 in the paper.

Tests for CST score manipulation

Perhaps the most important threat to validity in many RD analyses is the possibility that the continuous measure upon which treatment assignment is made may be manipulated. As an example, consider Niu and Tienda's (2010) RD analyses of the effect of a law that guaranteed college admission to students who finished in the top 10% of their high school class. In this setting, the possibility that school officials might manipulate some students' class ranks in order to boost their college chances

Table S3: Full difference-in-difference estimates of average effects of ID card program implementation at Live Oak and Mann High Schools

	ELA CST	Math CST	ELA CAHSEE	Math CAHSEE	ELA Grades	Math Grades
ID School	0.34**	0.27	0.25***	0.42**	0.01	0.08
	(0.06)	(0.22)	(0.03)	(0.08)	(0.22)	(0.07)
2010	0.03	0.08*	0.04*	0.02	-0.20***	0.02
	(0.02)	(0.02)	(0.01)	(0.04)	(0.06)	(0.04)
2011	0.01	0.17**	0.12**	0.02	-0.16*	0.10
	(0.03)	(0.04)	(0.03)	(0.05)	(0.07)	(0.09)
2012	0.07*	0.21***	0.07*	0.10*	-0.09	0.15*
	(0.02)	(0.02)	(0.03)	(0.04)	(0.07)	(0.07)
ID*2010	0.00	0.08	0.00	-0.16*	0.11	-0.12
	(0.02)	(0.07)	(0.04)	(0.06)	(0.06)	(0.07)
ID*2011	0.18***	0.31*	0.14**	0.01	0.21**	-0.01
	(0.03)	(0.10)	(0.03)	(0.05)	(0.07)	(0.10)
ID*2012	0.19***	0.33***	0.06	-0.09	0.21**	-0.02
	(0.03)	(0.02)	(0.03)	(0.04)	(0.07)	(0.11)
Hispanic	-0.14***	-0.14**	-0.16***	-0.21***	-0.36***	-0.26***
	(0.02)	(0.03)	(0.01)	(0.03)	(0.05)	(0.02)
Asian	0.27***	0.41***	0.25***	0.57***	0.50***	0.57***
	(0.02)	(0.03)	(0.03)	(0.03)	(0.04)	(0.05)
Pacific Islander	0.09**	0.06	0.09*	0.15***	0.14*	0.21***
	(0.02)	(0.05)	(0.03)	(0.02)	(0.05)	(0.05)
African American	-0.23***	-0.24***	-0.36***	-0.42***	-0.35***	-0.28***
	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)
Special Education	-1.00***	-0.28**	-0.91***	-0.90***	0.18	0.10
	(0.14)	(0.06)	(0.08)	(0.08)	(0.11)	(0.10)
Male	-0.07***	0.11***	-0.13***	0.15***	-0.35***	-0.10***
	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)	(0.01)
ELL	-0.89***	-0.28***	-1.08***	-0.73***	-0.46***	-0.35***
	(0.03)	(0.03)	(0.03)	(0.02)	(0.06)	(0.03)
Free/reduced lunch	0.03	0.08*	-0.01	0.12**	0.18***	0.15***
	(0.03)	(0.02)	(0.03)	(0.02)	(0.04)	(0.03)
Geometry CST		-0.06				
		(0.05)				
Summative math CST		0.45***				
		(0.06)				
Constant	0.20**	-0.30**	0.31***	0.06	0.38**	0.07
	(0.04)	(0.08)	(0.04)	(0.05)	(0.13)	(0.07)
N	19,567	18,763	19,504	19,466	19,686	19,722

Note : All models use school-clustered standard errors to account for school clustering. ELA and Math grade analyses use tobit estimation to correct for floor effects. A summarized version of this table is found in Table 4.

* p<0.05; ** p<0.01; *** p<0.001

Table S4: Full difference-in-difference estimates of average effects of ID card program implementation at Live Oak and Mann High Schools by student characteristics

	ELA CST	Math CST	ELA CST	Math CST	ELA CST	Math CST
ID school	0.71***	0.76**	0.36***	0.40*	0.70***	0.74**
	(0.07)	(0.18)	(0.04)	(0.15)	(0.06)	(0.17)
2010	0.47***	0.53***	0.00	0.09	0.44**	0.52***
	(0.06)	(0.06)	(0.07)	(0.06)	(0.08)	(0.07)
2011	0.34**	0.59***	0.03	0.19**	0.33**	0.58***
	(0.06)	(0.03)	(0.05)	(0.04)	(0.08)	(0.04)
2012	0.50***	0.60***	0.07	0.29***	0.48***	0.64***
	(0.07)	(0.04)	(0.04)	(0.04)	(-0.07)	(0.06)
ID*2010	-0.28**	-0.24*	0.04	0.06	-0.23*	-0.16
	(0.07)	(0.08)	(0.08)	(0.06)	(0.08)	(0.09)
ID*2011	-0.02	0.02	0.24**	0.22*	0.09	0.03
	(0.07)	(0.14)	(0.05)	(0.07)	(0.07)	(0.13)
ID*2012	-0.09	0.00	0.21*	0.25**	-0.06	-0.02
	(0.10)	(0.11)	(0.07)	(0.07)	(0.12)	(0.10)
Hispanic*ID*2010			-0.12	-0.09	-0.03	-0.11
			(0.11)	(0.08)	(0.07)	(0.07)
Asian*ID*2010			-0.04	-0.06	-0.08	0.02
			(0.11)	(0.17)	(0.10)	(0.09)
Pacific Is*ID*2010			0.16	0.23*	0.03	-0.09
			(0.10)	(0.08)	(0.08)	(0.07)
Black*ID*2010			-0.15*	-0.04	-0.01	-0.05
			(0.05)	(0.04)	(0.07)	(0.11)
Hispanic*ID*2011			-0.20	0.02	-0.16*	-0.08
			(0.10)	(0.04)	(0.05)	(0.04)
Asian*ID*2011			0.10	0.23	-0.04	0.19
			(0.10)	(0.20)	(0.07)	(0.10)
Pacific Is*ID*2011			-0.06	0.17	-0.21*	-0.17*
			(0.11)	(0.13)	(0.06)	(0.07)
Black*ID*2011			-0.30**	0.03	-0.06	0.14**
			(0.07)	(0.08)	(0.09)	(0.03)
Hispanic*ID*2011			-0.21	0.02	-0.08	0.02
			(0.11)	(0.05)	(0.05)	(0.05)
Asian*ID*2011			0.07	0.08	-0.04	0.09
			(0.07)	(0.12)	(0.07)	(0.11)
Pacific Is*ID*2011			0.13	0.36**	-0.02	0.03
			(0.09)	(0.09)	(0.07)	(0.07)
Black*ID*2011			-0.10	-0.21	0.12	-0.13
			(0.06)	(0.11)	(0.08)	(0.20)

Table S4, continued

	ELA CST	Math CST	ELA CST	Math CST	ELA CST	Math CST
Close*ID*2010	0.36**	0.29**			0.36**	0.29**
	(0.09)	(0.06)			(0.09)	(0.07)
Under*ID*2010	0.35**	0.37**			0.34**	0.37**
	(0.09)	(0.08)			(0.09)	(0.08)
Close*ID*2011	0.27**	0.26*			0.26*	0.27*
	(0.08)	(0.09)			(0.08)	(0.08)
Under*ID*2011	0.09	0.21*			0.09	0.24*
	(0.11)	(0.08)			(0.11)	(0.08)
Close*ID*2012	0.34***	0.36**			0.34**	0.36**
	(0.05)	(0.10)			(0.06)	(0.09)
Under*ID*2012	0.17	0.25*			0.18	0.26*
	(0.14)	(0.10)			(0.16)	(0.09)
Hispanic	-0.09**	-0.10*	-0.10*	-0.08	-0.09*	-0.10*
	(0.02)	(0.03)	(0.04)	(0.04)	(0.03)	(0.04)
Asian	0.14***	0.39***	0.26***	0.48***	0.18**	0.45
	(0.01)	(0.04)	(0.04)	(0.07)	(0.03)	(0.08)
Pacific Is	0.07*	0.10*	0.12*	0.26*	0.09	0.16
	(0.02)	(0.04)	(0.05)	(0.09)	(0.04)	(0.09)
Black	-0.17**	-0.18**	-0.22**	-0.23***	-0.17**	-0.21***
	(0.03)	(0.04)	(0.05)	(0.03)	(0.04)	(0.02)
Special Ed	-0.91***	-0.58**	-0.99***	-0.65**	-0.85***	-0.54**
	(0.13)	(0.11)	(0.14)	(0.12)	(0.12)	(0.11)
Male	-0.08***	0.07**	-0.07***	0.08**	-0.07***	0.09**
	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.02)
ELL	-0.76***	-0.24***	-0.90***	-0.39***	-0.67***	-0.21***
	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)
Free/reduced lunch	0.03	0.12**	0.02	0.10**	0.00	0.08*
	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)
Hispanic*ID			0.01	-0.13*	0.01	-0.13*
			(0.08)	(0.05)	(0.08)	(0.05)
Asian*ID			-0.02	0.12	-0.02	0.12
			(0.06)	(0.15)	(0.06)	(0.15)
Pacific Is*ID			-0.11	-0.28*	-0.11	-0.28*
			(0.06)	(0.12)	(0.06)	(0.12)
Black*ID			0.15*	-0.03	0.15*	-0.03
			(0.06)	(0.05)	(0.06)	(0.05)
Hispanic*2010			0.03	-0.01	0.05	0.04
			(0.07)	(0.07)	(0.07)	(0.08)
Hispanic*2011			-0.01	-0.03	0.03	0.05
			(0.06)	(0.03)	(0.04)	(0.04)

Table S4, continued

	ELA CST	Math CST	ELA CST	Math CST	ELA CST	Math CST
Hispanic*2012			-0.01	-0.09*	0.01	-0.04
			(0.06)	(0.04)	(0.06)	(0.05)
Asian*2010			0.07	-0.02	0.03	-0.14
			(0.10)	(0.14)	(0.09)	(0.12)
Asian*2011			-0.12	-0.04	-0.11	-0.13
			(0.10)	(0.15)	(0.08)	(0.13)
Asian*2012			0.04	-0.03	0.01	-0.13
			(0.04)	(0.12)	(0.05)	(0.11)
Pacific Is*2010			0.03	-0.05	0.02	-0.01
			(0.09)	(0.07)	(0.10)	(0.10)
Pacific Is*2011			-0.03	-0.19*	0.00	-0.09
			(0.09)	(0.07)	(0.08)	(0.09)
Pacific Is*2012			-0.01	-0.10	-0.02	-0.05
			(0.08)	(0.07)	(0.07)	(0.10)
Close	0.40***	0.25**			0.40***	0.25**
	(0.04)	(0.04)			(0.04)	(0.04)
Under	-0.06	-0.23**			-0.06	-0.22**
	(0.06)	(0.04)			(0.06)	(0.04)
Close*2010	-0.51**	-0.46***			-0.50**	-0.46***
	(0.08)	(0.07)			(0.08)	(0.07)
Close*2011	-0.41**	-0.48***			-0.41***	-0.48***
	(0.07)	(0.02)			(0.07)	(0.03)
Close*2012	-0.61***	-0.55***			-0.61***	-0.55***
	(0.04)	(0.04)			(0.04)	(0.04)
Under*2010	-0.52***	-0.50***			-0.53***	-0.51***
	(0.08)	(0.06)			(0.08)	(0.06)
Under*2011	-0.39**	-0.52***			-0.41**	-0.54***
	(0.10)	(0.06)			(0.10)	(0.06)
Under*2012	-0.52***	-0.44***			-0.51**	-0.44***
	(0.09)	(0.05)			(0.10)	(0.05)
Close*ID	-0.63***	-0.67***			-0.62***	-0.66***
	(0.06)	(0.06)			(0.06)	(0.05)
Under*ID	-0.52***	-0.64***			-0.52***	-0.63***
	(0.08)	(0.08)			(0.08)	(0.08)
Constant	0.10	-0.19*	0.18***	-0.24*	0.10*	-0.19*
	(0.05)	(0.07)	(0.02)	(0.08)	(0.04)	(0.07)
N	19,567	19,283	19,567	19,283	19,567	19,283

Note : All models use school-clustered standard errors to account for school clustering. A summarized version of this table is found in Table 5.

* p<0.05; ** p<0.01; *** p<0.001

Table S5: Supplemental tests of statistical significance for difference-in-difference estimates of average effects of ID card program implementation at Live Oak and Mann High Schools by student

Models 1 and 2:		
Under*ID*2011+ ID*2011=0	**ELA:** p=0.279	**Math:** p=0.082
Under*ID*2012+ ID*2012=0	**ELA:** p=0.170	**Math:** p=0.000
Close*ID*2011+ ID*2011=0	**ELA:** p=0.000	**Math:** p=0.002
Close*ID*2012+ ID*2012=0	**ELA:** p=0.004	**Math:** p=0.000
Under*ID*2011= Close*ID*2011	**ELA:** p=0.036	**Math:** p=0.585
Under*ID*2012= Close*ID*2012	**ELA:** p=0.120	**Math:** p=0.136
Models 3 and 4:		
Hispanic*ID*2011+ ID*2011=0	**ELA:** p=0.638	**Math:** p=0.038
Hispanic*ID*2012+ ID*2012=0	**ELA:** p=0.964	**Math:** p=0.000
Asian*ID*2011+ ID*2011=0	**ELA:** p=0.002	**Math:** p=0.064
Asian*ID*2012+ ID*2012=0	**ELA:** p=0.002	**Math:** p=0.031
Hispanic*ID*2011= Asian*ID*2011	**ELA:** p=0.064	**Math:** p=0.262
Hispanic*ID*2012= Asian*ID*2012	**ELA:** p=0.011	**Math:** p=0.596
Models 5 and 6:		
Under*ID*2011= Close*ID*2011	**ELA:** p=0.031	**Math:** p=0.726
Under*ID*2012= Close*ID*2012	**ELA:** p=0.158	**Math:** p=0.161
Hispanic*ID*2011= Asian*ID*2011	**ELA:** p=0.234	**Math:** p=0.003
Hispanic*ID*2012= Asian*ID*2012	**ELA:** p=0.531	**Math:** p=0.629

is a threat to validity, particularly if unmeasured selection processes are associated with this manipulation.

In our setting, however, the possibility for direct test-score manipulation seems relatively remote. CSTs are multiple choice exams that are machine-graded at a location that is remote from the Sudden Valley Unified school sites. The state then reports scores simultaneously to students, school officials, and state educational authorities. It is therefore unlikely that students or educators can manipulate scores at the proficiency margin to influence students' ID card assignments.

It is, however, possible that cheating (either by students or by educators) could bias student CST scores. We know of no evidence to suggest that California educators engaged in the sort of wide-spread standardized test score manipulation that occurred in the Atlanta Public Schools. However, in 2012, reports surfaced of a handful of incidents in which students elsewhere in California posted CST test questions or completed answer sheets on social media (Blume 2012). We cannot assess the extent to which cheating biases CST scores at Live Oak or Mann. However, Figure S1 indicates that the distribution of test scores for Live Oak and Mann students is smooth across the threshold for white card assignment. If students or teachers were cheating in an attempt to score above the threshold for proficiency, we might expect to find fewer than expected students just under the threshold and more than expected students just above the threshold. Thus, while it is of course possible that there was some cheating that might bias test scores, we have no reason

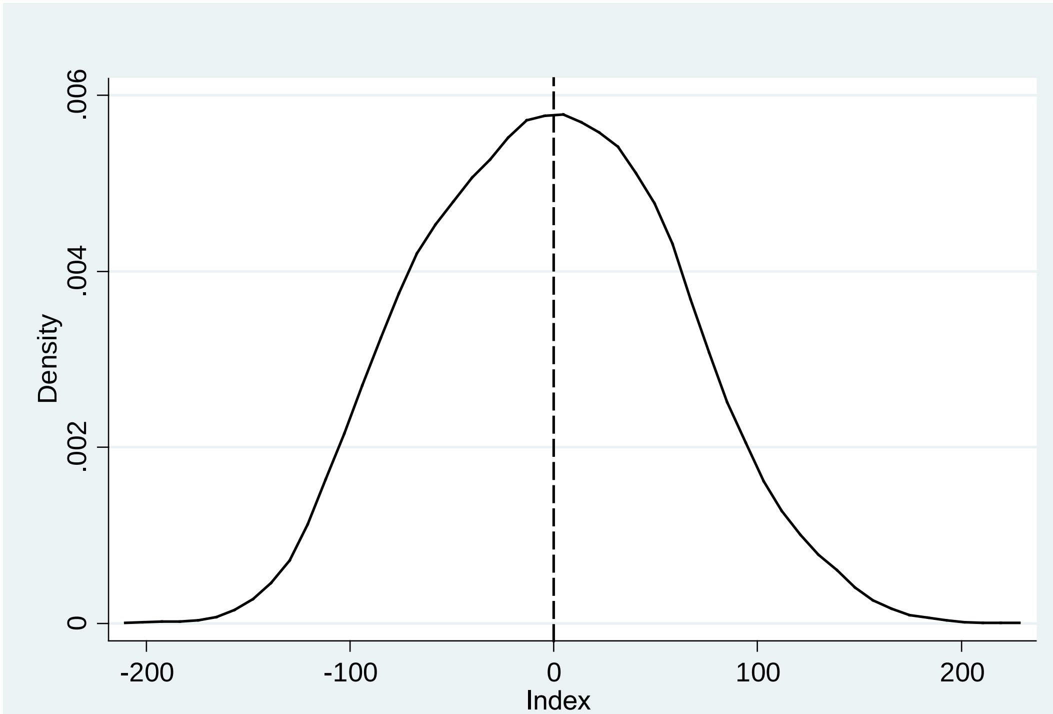

Figure S1: Distribution of lowest recentered 2011 ELA and Mathematics CST scores ("Index"), 2011-12 Live Oak and Mann students

to believe that there was cheating, and find no evidence that this differentially affected gold card eligibility.

The horizontal access of this distribution plot is the lowest of students' recentered 2011 math and ELA test scores. Our regression discontinuity analyses consider whether there are differences in students' later outcome at the threshold (Index=0). The fact that there is no apparent clustering of student test scores on either side of the threshold is consistent with the assumption that for students close to the threshold, assignment to a white or gold ID is as good as random; a formal density test (McCrary 2008) indicates that the distribution of observations is smooth at the threshold. We find further evidence to support the validity of the assumption that there was no systematic sorting at the threshold in supplementary analyses suggesting that students on the two sides of the threshold do not differ appreciably on observable characteristics (i.e. prior achievement, race/ethnicity, English language proficiency status, and other demographic characteristics). The sole exception is gender, where we find a higher than expected number of boys above the threshold. However, supplemental analyses (not reported) find that our results are robust to the inclusion of gender as a control variable, and we find no evidence to suggest that the effects of white card receipt vary by gender.

Tests for the fidelity of ID card assignment

Another potential source of bias in our RD estimates of the effect of white card assignment are disparities between our operationalization of the ID card policy and its actual implementation. Figure 3 in the manuscript indicates that there is a sizable discontinuity in students' probabilities of receiving a white card versus a higher status gold or platinum card at the threshold on our index variable. However, this discontinuity is not perfect. Virtually no students just below the proficiency threshold on either ELA or mathematics earn a gold or platinum card. However, a small proportion of students whose lowest ELA and mathematics score fell below the assignment threshold received a high status card. At the same time nearly 40 percent of students just above the threshold received a white card. Sudden Valley administrator reports suggest that the ID assignment was an automated process with few opportunities for intervention or manipulation. Based on these reports, together with the supplementary analyses described below, we suspect that these reports are accurate and that discrepancies between students' observed ID card assignment and their index scores are due primarily to limitations in our data which make it difficult for us to fully account for the assignment rules at Live Oak and Mann.

As we note elsewhere in the manuscript, although the ID cards are assigned based on students' performance on up to four CST tests (math, ELA, history, and science), we only have complete student performance data on two tests (math and ELA). The most important way in which this data limitation influences our analyses is that it makes it impossible to identify students who scored above the proficiency threshold on both mathematics and ELA CSTs but received a white card due to below-threshold scores in science or history. Supplementary analyses conducted with a subset of students for whom we have access to all test scores (but not data on ID card assignment) indicates that nearly 30 percent of Live Oak and Mann students who score just above the proficiency threshold on the math and ELA CST score below the proficiency threshold on the science or history CST. These "misassignment" rates vary somewhat with student grade. Among Live Oak and Mann students who score above the proficiency threshold on both math and ELA CSTs, ninth and tenth grade students are less likely to score below the proficiency threshold on science or history than eleventh or twelfth grade students. We suspect that this may be due to the combination of two factors: First, a larger proportion of eleventh and twelfth grade students opt out of math classes, thus avoiding the CST that is often most challenging for students. Second, upper-level science and history tests may be more difficult than their eighth and ninth grade equivalents. With complete test score data, the discontinuity in ID card assignment at the proficiency threshold might approach 90 percent.

We believe that the final ten percent is explained by the clause in the assignment rules stipulating that students whose score on any CST declines by one or more performance band cannot receive a gold card. Since the assignment threshold is not binding for students who dropped from advanced to proficient, we excluded the 597 students whose mathematics or ELA CST scores declined from advanced to proficient in the year prior to ID assignment from our RD analyses. However, we are unable to identify students who dropped a performance band in science or history. Our supplemental analyses suggest that such students account for the remaining discrepancy in students who score above the threshold but do not receive a gold card.

We believe the small percentage of students who receive gold cards but score below the cutoff is explained by the mechanism that allows students to earn gold cards via test score growth. While our analyses attempt to account for these criteria, we are unable to completely do so because we lack data on all students' science and history CSTs. Since the assignment threshold is not binding for students who qualify via growth criteria, we exclude the 349 students who improved by one or more performance bands in both ELA and math from our RD analyses. Doing so reduces the number of gold card recipients who scored below the threshold from 101 to 50. While we suspect that these 50 are students who met growth criteria using science and history, we cannot definitively confirm this. This is, however, consistent with the fact that in the data with all four test scores, but which is missing card assignment, we find that approximately twice as many students meet growth criteria based on math and ELA only, compared with all four scores.

In sum, these supplementary analyses lend credibility to the assertion that white card assignment is as good as random for Live Oak and Mann students near the proficiency threshold, and that the assignment process was followed faithfully. As such, it seems unlikely that unmeasured selection biases confound our estimates of the effects of white card assignment. Rather, since these analyses suggest that missing data on history and science CSTs lead us to include students who actually received white cards in the pool of gold card recipients, it seems likely that our analyses understate the actual effects of white card assignment. One approach to addressing this missing data problem is to treat our models as "fuzzy" regression discontinuities, and to accordingly inflate RD effect-size estimates by the inverse of the treatment discontinuity (Imbens and Lemieux 2008). In our case, this approach would involve dividing each of the effect-size estimates reported in Table 6 by the first stage to generate treatment on the treated (TOT) estimates of the effects of white card receipt. By this method, for example, the estimate for the effect of white card receipt on students' ELA CST scores is -34.10 points (-20.42/0.6) or approximately -0.60 standard deviations. Looking at the results of analogous models with a variety of specifications reported in Table S6, our scaled up standardized estimate for ELA CAHSEE is approximately 1, ELA grades is 1.45 (based on a standardized ITT effect of roughly .8), and math grades is 1.15 (based on a standardized ITT effect of roughly 0.6). While these scaled estimates are large, given that Lovaglia et al. (1998) find that randomly assigning college students to high or low status identities in a laboratory produces effects of roughly half a standard deviation on IQ, these scaled up estimates are perhaps not unreasonably large.

However, these scaled up estimates hinge on the assumption that the effects of ID card assignment are identical for students at the threshold of assignment on the math and ELA CSTs we observe as for students on the threshold of assignment on the science and history CSTs that we do not observe. Since this assumption is not

Table S6: Alternate specifications for regression discontinuity estimates of the effects of white vs. gold ID assignment on student achievement and behavior, 2011-12 Live Oak and Mann students

	Optimal bandwidth	50% optimal bandwidth	200% optimal bandwidth	Rectangular kernel	Bandwidth= 10	Donut RD	All 4 test scores
ELA CST	**-20.42****	**-39.48*****	**-14.70***	**-24.99****	**-27.35*****	-19.24†	-13.22†
	(-2.67)	(-3.86)	(-2.59)	(-2.94)	(-3.29)	(-1.76)	(-1.85)
Math CST	-7.57	-19.17	-1.22	-9.3	-10.77	-8.16	-12.55
	(-0.58)	(-0.76)	(-0.14)	(-0.70)	(-0.76)	-0.58	(-1.04)
ELA exit exam	**-22.17***	**-50.26****	-11.94†	**-17.02***	**-20.55***	**-20.55***	**-20.77***
	(-2.44)	(-2.86)	(-1.69)	(-1.96)	(-2.28)	(-2.28)	(-2.17)
Math exit exam	5.2	21.06	3.1	12.21	5.14	5.14	4.55
	(0.53)	(0.91)	(0.49)	(1.34)	(0.53)	(0.53)	(0.61)
ELA grades	**-1.56****	**-0.85***	**-0.89***	**-0.85***	**-0.44***	-0.1	---
	(-2.82)	(-2.33)	(-2.72)	(-2.33)	(-2.21)	(-0.48)	---
Math grades	**-0.84***	-0.25	**-0.63***	-0.66	**-.46***	-0.39†	---
	(-2.09)	(-0.77)	(-2.41)	(-1.20)	(-2.45)	(-1.82)	---
Suspensions	2.96	0.98	1.57	3.08	1.78	2.07	---
	(1.43)	(0.24)	(1.03)	(1.27)	(1.10)	(0.85)	---

Note : CST data are available for students in grades 9-11; exit exam data available only for 10^{th} graders; grades and suspensions available for all 9^{th}-12^{th} graders. The models estimated using the rectangular kernel used the optimal bandwidth, which differed slightly the optimal bandwidth of the triangular kernel models. The donut RD excluded values ±1 from the threshold, and used a bandwidth of 10. As in manuscript Table 6, the first stage varies across models, ranging from .5 to .7 (with the exception of the model for math grades using a rectangular kernel, which had a first stage of .42); because we do not have information on card placement in the dataset that contains all four scores, we cannot estimate the first stage for these models, which we believe is close to 1. As noted in the online supplement text, the models using all four test use a slightly different sample, and do not contain information on grades or suspensions. Z-statistics from robust standard errors reported in parentheses.

† p<0.10; * p<0.05; ** p<0.01; *** p<0.001

testable, we also estimated analogous RD models on additional data that contain all four CST scores used to make card assignments. These models have less statistical power, as the data only include the subset of students who were in the district in eighth grade (on average a slightly advantaged subpopulation who are more residentially stable), and we can only examine effects on the CST and CAHSEE (not grades and disciplinary information). Most importantly, these data do not contain information on ID card assignment, so that we cannot estimate the first stage. Thus, while we believe that the first stage here is close to one, we cannot ultimately test this. The results of these models are presented in the final column in Table S6. In general, these models provide estimates that are similar in magnitude, and if anything they are slightly smaller than the ITT estimates based only ELA and math CSTs presented in Table S6. Given this, we believe that the ITT estimates in Table S6 are likely to be closer to the estimates of card receipt than the scaled up estimates. As such, we discuss the ITT estimates as estimates of the effect of receiving a low status ID card.

External validity in the regression discontinuity setting

The supplemental analyses above suggest our RD estimates of the effects of white card assignment have a high degree of internal validity. They do not, however, speak to the external validity of these findings. We argue that these models provide reliable – if arguably conservative – estimates of the causal effects of white card placement *for students who score near the proficiency threshold*. What is less clear is the extent to which these findings apply to students who score far above or far below the proficiency threshold.

One way to address this issue is to consider the extent to which the RD models are sensitive to model specification. Our preferred RD analyses utilize local linear regression models with a triangular kernel to estimate the magnitude both of the discontinuity in students' white card placement probability at the proficiency threshold and to take advantage of this discontinuity to estimate the effect of white card placement on subsequent student outcomes. This approach focuses analytic attention efficiently on cases closely to the assignment threshold. As such, its results are sensitive to bandwidth, or the number of cases utilized in local averaging (Nichols 2007). The Stata "rd" command, which we use to estimate our main regression discontinuity models, selects bandwidths that minimize mean squared error, which is equivalent to squared bias plus variance (Imbens and Kalyanaraman 2012). Smaller bandwidths produce effect size estimates that are more local to the assignment threshold but less efficient; larger bandwidths produce effect size estimates that are more efficient but less local to the threshold. As such, estimating RD models at different bandwidths provides an important test of model sensitivity and the extent to which the effects of discontinuous assignment are local to the assignment threshold.

Figures S2, S3, and S4 illustrate our RD analysis of the effects of gold card assignment on ELA CST scores using the optimal bandwidth (cf. Imbens and Kalyanaraman, 2012), half the optimal bandwidth, and twice the optimal bandwidth, respectively. A close inspection of Figures S2-S4 highlights how our estimates are particularly local to the threshold. Each of these figures suggests that students whose lowest CST score was just below the white card assignment threshold are on average achieving at lower levels than their peers both just above the threshold and their peers further below the threshold. Although this basic pattern holds in all three figures, it is more apparent in Figure S3 (where the bandwidth is half the preferred) than in Figure S4 (where the bandwidth is twice the preferred). This pattern suggests that the negative effects of white card placement are most pronounced for students who just miss a higher-status gold card than for students who scored further from the assignment threshold. Figures S5-S7 present the RD estimates (using only the preferred bandwidths) for the other outcomes with significant differences in manuscript Table 6, showing a similar pattern. Taken together, these figures suggest that just missing the cutoff may particularly hinder students' later achievement across a range of outcomes.

Table S6 reports results for all of the different outcomes we examine from models using a variety bandwidths and a different kernel, showing that our results are robust to alternative model specifications. In addition to using the optimal bandwidth, half of the optimal bandwidth, and twice the optimal bandwidth, we also estimated models using the same bandwidth of 10 points (just over 0.10 standard deviations) across all outcomes, and models using a rectangular kernel. Overall, we find that our results are robust to a variety of different estimation strategies. ELA CST scores, CAHSEE scores, and grades are significant or marginally significant in all specifications. For ELA CST and CAHSEE scores we see that the results of our preferred approach (reported in Table 6 and reproduced in Model 1 of Table S6) are well within the range of the estimates produced by the other estimation strategies. By contrast, the result for ELA grades reported in Table 6 is substantially larger than the results from other approaches, which suggest that an effect in the range of 0.8 is perhaps more reasonable. Our result for math grades in Table 6 also appears somewhat higher than the results provided by alternate specifications (two of which are not significant), suggesting that an effect of around 0.5 is perhaps a more reasonable estimate.

Finally, we also estimated models in which we omitted students who scored right on the cutoff, a so-called donut regression discontinuity. This is often done when there are concerns regarding systematic manipulations across the threshold. While we found no evidence of heaping or score manipulation that would indicate that there were systematic biases across the threshold, the donut regression discontinuity allows us to observe the degree to which the differences we find are driven by individuals who are just above and below the threshold. Given that Figures S2-S7 suggest students who just miss the cutoff perform particularly poorly the next year, these models provide an indication of how much our results are driven by this phenomenon. In order to ensure sufficient sample sizes, we again use a bandwidth of ten across all of these models. Comparing these models to the models with the bandwidth of ten that do not drop individuals on the threshold, we find that our results for ELA grades appear to be particularly driven by individuals at the threshold. In the other outcomes, we see a more gradual attenuation of the effects, accompanied by a loss of statistical power that reduces two coefficients to marginal significance (the coefficients for the ELA exit exam models are identical, as this analysis looks only at the tenth grade students taking the CAHSEE, and there were no tenth grade students with the exact scores omitted).

The sensitivity of the ELA grade estimates is partially due to a large spike in students' odds of failing ELA courses below the ID card assignment threshold. Nearly 30 percent of students who scored just one point below the ID card threshold failed an ELA course the next year, compared to 6.5 percent of other students who scored within 10 points of the threshold (including 5.7 percent of those above the threshold and 7.2 percent of those below). Excluding failing students reduces our estimates for ELA grades by nearly half (in all models except for the model using a rectangular kernel, where the reduction is considerably smaller), suggesting that course failure accounts for a large portion of the ELA grade effect that we observe, and explaining why omitting students close to the threshold changes the ELA grade effects so much. As noted above, on balance Table S6 suggests that the effect of being below the threshold results in a 0.8 decrease in ELA grades; we thus see that students who do not fail ELA on balance receive grades 0.4 grade points lower, and the rest of the difference that we observe is due to increased rates of ELA course failure. More broadly, it is perhaps not surprising that we see somewhat larger effects on grades than on other outcomes, given that grades are affected not only by students' effort, but also by teachers' perceptions. Given that we found no evidence to believe that there are systematic differences across the threshold, we interpret these differences as causal effects of scoring below the threshold.

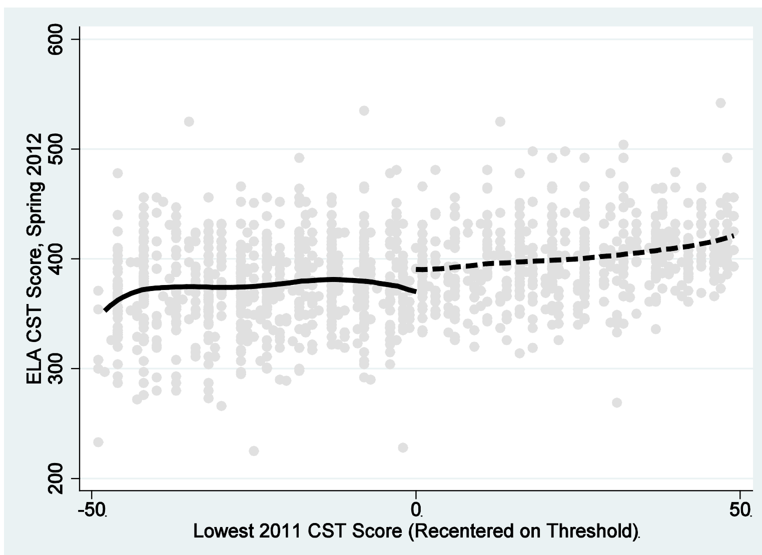

Figure S2: Regression discontinuity analysis of the effect of white card placement on Spring 2012 ELA CST scores, for 2011-12 Live Oak and Mann students (preferred bandwidth: 13.3)

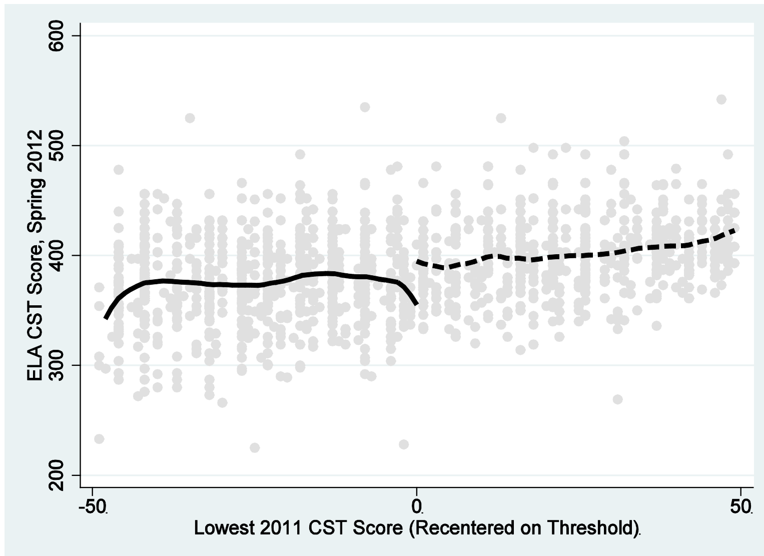

Figure S3: Regression discontinuity analysis of the effect of white card placement on Spring 2012 ELA CST scores, for 2011-12 Live Oak and Mann students (50% of preferred bandwidth: 6.6)

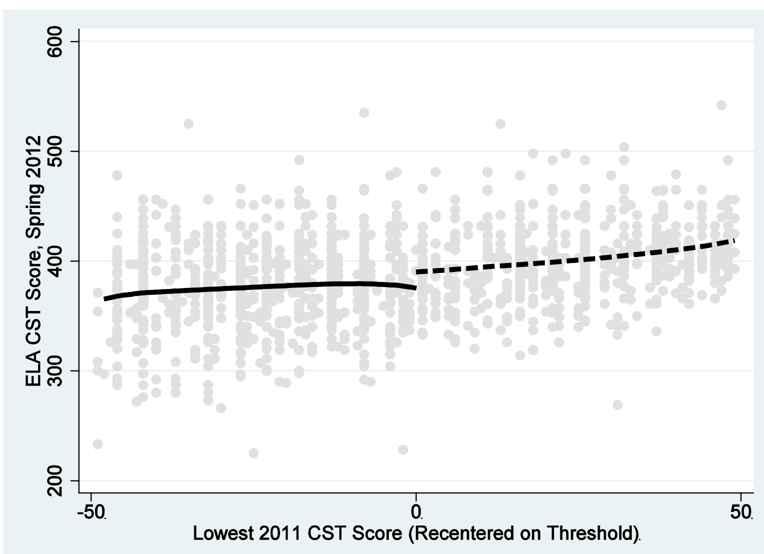

Figure S4: Regression discontinuity analysis of the effect of white card placement on Spring 2012 ELA CST scores, for 2011-12 Live Oak and Mann students (200% of preferred bandwidth: 26.6)

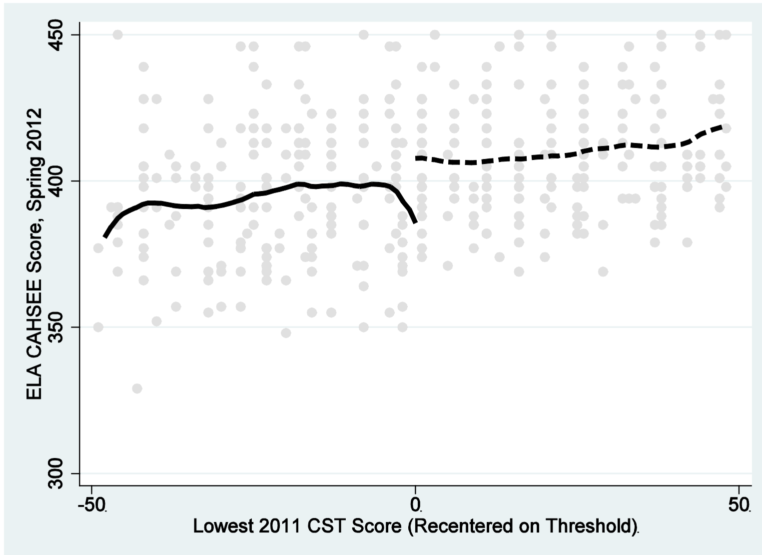

Figure S5: Regression discontinuity analysis of the effect of white card placement on Spring 2012 ELA CAHSEE scores, for 2011-12 Live Oak and Mann students (preferred bandwidth: 9.3)

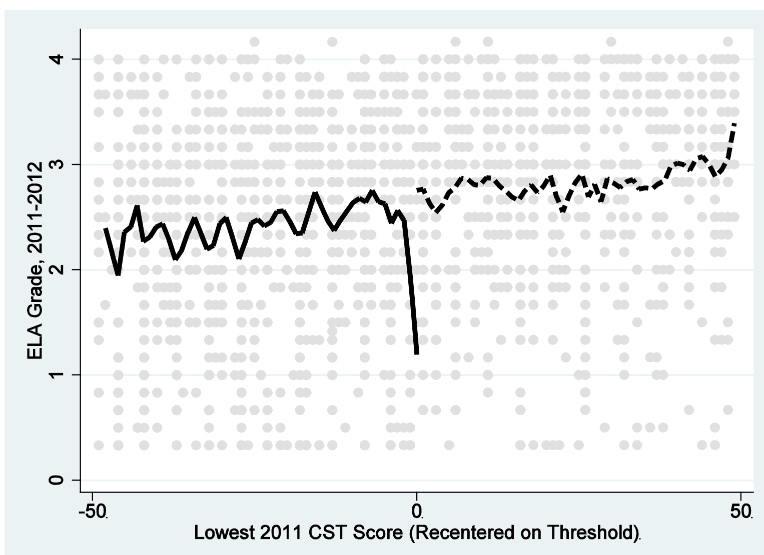

Figure S6: Regression discontinuity analysis of the effect of white card placement on Spring 2012 ELA grades, for 2011-12 Live Oak and Mann students (preferred bandwidth: 2.4)

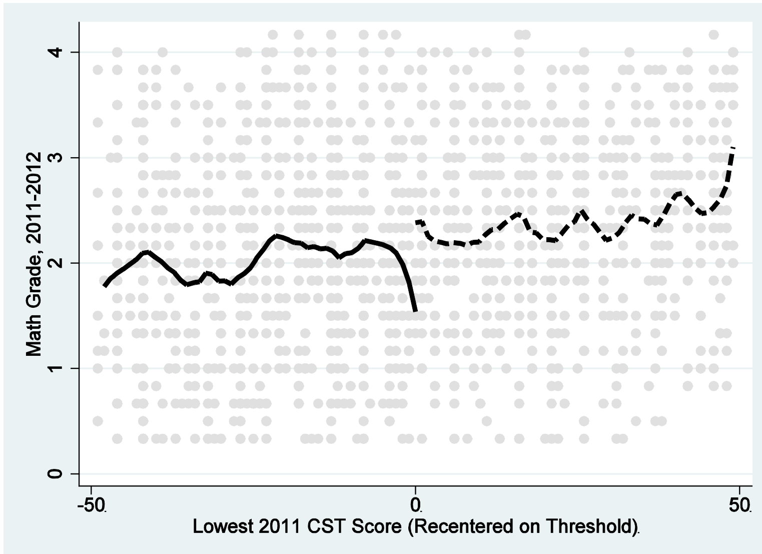

Figure S7: Regression discontinuity analysis of the effect of white card placement on Spring 2012 math grades, for 2011-12 Live Oak and Mann students (preferred bandwidth: 3.5)

Notes

1 There are three possible reasons why a student's actual card may not correspond with the proxy: (1) in addition to placing students based on their CST achievement, the placement system in both schools allows students to earn gold cards by improving their performance band; (2) while the proxy we use here is built on the two test scores made available to us, the test actual score placement systems use up to four CST scores (math, ELA, science, and history); (3) actual ID card placements might not correspond with the reported criteria. In our data, among 2012 Live Oak and Mann students who scored advanced on both math and ELA, 83 percent received platinum cards, 17 percent gold, and 0 percent white; of students whose lowest score was basic but also had an advanced or a proficient score, 0 percent received platinum cards, 5 percent gold, and 95 percent white; of students who scored basic or lower in both math and ELA, 0 percent received platinum cards, 1 percent gold, and 99 percent white.