Web Appendix

Chay, Kenneth Y.; McEwan, Patrick J. and Urquiola, Miguel. "The Central Role of Noise in Evaluating Interventions That Use Test Scores to Rank Schools." *American Economic Review*, 2005, *95*(4), pp. 1237-1258.

Table A1: P-900 effects from specifications with controls for 1988 scores and school sizes

	1988-1990 gain score				1988-1992 gain score			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: **Mathematics**								
P-900	-0.16	0.18	-0.14	0.28	1.79***	1.91***	1.80***	2.11***
	(0.51)	(0.46)	(0.51)	(0.53)	(0.56)	(0.50)	(0.56)	(0.60)
1988 score		-0.17***				-0.15***		
		(0.02)				(0.02)		
1988 school size (1/10)		0.12***				0.17***		
		(0.03)				(0.03)		
SES index, 1990				0.15***				0.18***
				(0.01)				(0.01)
Change in SES, 1990-1992								0.07***
								(0.01)
Cubic in '88 score	Y	N	N	N	Y	N	N	N
Flexible score, size	N	N	Y	Y	N	N	Y	Y
Region dummies	N	N	N	Y	N	N	N	Y
Adjusted-R^2	0.041	0.045	0.047	0.134	0.053	0.059	0.061	0.145
Sample Size	2,644	2,644	2,644	2,644	2,591	2,591	2,591	2,591
Panel B: **Language**								
P-900	-0.02	0.43	-0.00	0.57	1.67***	2.37***	1.69***	2.12***
	(0.48)	(0.43)	(0.48)	(0.49)	(0.48)	(0.43)	(0.48)	(0.52)
1988 score		-0.30***				-0.28***		
		(0.02)				(0.02)		
1988 school size (1/10)		0.12***				0.15***		
		(0.03)				(0.03)		
SES index, 1990				0.13***				0.16***
				(0.01)				(0.01)
Change in SES, 1990-1992								0.07***
								(0.01)
Cubic in '88 score	Y	N	N	N	Y	N	N	N
Flexible score, size	N	N	Y	Y	N	N	Y	Y
Region dummies	N	N	N	Y	N	N	N	Y
Adjusted-R^2	0.155	0.159	0.162	0.235	0.173	0.177	0.183	0.258
Sample Size	2,644	2,644	2,644	2,644	2,591	2,591	2,591	2,591

Notes: ***, **, and * indicate statistical significance at the 1, 5, and 10 percent level, respectively. The sample covers urban schools with 15 or more students in the fourth grade in 1988. Huber-White standard errors are in parentheses. The "Flexible score, size" specifications include a cubic in the 1988 score, a quadratic in the 1988 school size, and the interaction of 1988 score with 1988 school size. The results are similar if one adds controls for the cube of the 1988 school size and interactions of the 1988 score with the square of 1988 size, the square of 1988 score with 1988 size, and the square of 1988 score with the square of 1988 size.

Table A2: P-900 effects on 1988-1992 gain scores, in all regions except Region 13

	All schools		± 5 points		± 3 points		± 2 points	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: **Mathematics**								
P-900	4.47 ***	2.12 ***	1.63 **	1.40 *	1.75 **	1.58 *	2.42 **	2.31 **
	(0.51)	(0.68)	(0.70)	(0.77)	(0.84)	(0.89)	(0.94)	(0.96)
SES index, 1990		0.16 ***		0.15 ***		0.11 ***		0.11 ***
		(0.02)		(0.03)		(0.03)		(0.04)
Change in SES,		0.08 ***		0.10 ***		0.11 ***		0.07 **
1990-1992		(0.02)		(0.02)		(0.03)		(0.03)
Cubic in 1988 score	N	Y	N	Y	N	Y	N	Y
R^2	0.051	0.146	0.010	0.075	0.013	0.080	0.028	0.098
Sample Size	1,640	1,640	569	569	345	345	233	233
Panel B: **Language**								
P-900	6.68 ***	2.20 ***	3.17 ***	1.88 ***	2.48 ***	1.77 **	2.89 ***	2.33 ***
	(0.45)	(0.59)	(0.63)	(0.69)	(0.79)	(0.82)	(0.93)	(0.89)
SES index, 1990		0.15 ***		0.14 ***		0.09 ***		0.09 **
		(0.02)		(0.03)		(0.03)		(0.04)
Change in SES,		0.07 ***		0.09 ***		0.11 ***		0.07 **
1990-1992		(0.01)		(0.02)		(0.02)		(0.03)
Cubic in 1988 score	N	Y	N	Y	N	Y	N	Y
R^2	0.124	0.265	0.044	0.130	0.028	0.115	0.040	0.120
Sample Size	1,640	1,640	569	569	345	345	233	233

Notes: ***, **, and * indicate statistical significance at the 1, 5, and 10 percent level, respectively. The sample covers urban schools with 15 or more students in the fourth grade in 1988 for all regions except Santiago (region 13). Huber-White standard errors are in parentheses.

Table A3: SES of treated and eligible schools close to the cutoff

	All Regions						All Regions but region 13					
	All schools		± 5 points of cutoff		± 2 points of cutoff		All schools		± 5 points of cutoff		± 2 points of cutoff	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: **1990 SES Index on P-900**												
P-900	-17.01***	-4.20***	-7.48***	-4.98***	-6.35***	-4.32***	-14.59***	-0.30	-5.53***	-1.49	-4.54**	-2.76
	(0.81)	(1.05)	(1.08)	(1.16)	(1.51)	(1.59)	(0.94)	(1.25)	(1.26)	(1.40)	(1.79)	(1.80)
Cubic in '88 score	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y
Sample size	2,644	2,644	896	896	366	366	1,671	1,671	578	578	236	236
Panel B: **1992 SES Index on P-900**												
P-900	-20.85***	-5.15***	-8.32***	-4.52***	-5.84***	-4.86***	-19.59***	-1.73	-6.71***	-1.16	-3.58*	-2.28
	(0.95)	(1.23)	(1.22)	(1.21)	(1.70)	(1.69)	(1.08)	(1.38)	(1.40)	(1.38)	(1.96)	(1.92)
Cubic in '88 score	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y
Sample size	2,591	2,591	883	883	363	363	1,640	1,640	569	569	233	233
Panel C: **1990 SES Index on Eligibility**												
Eligible	-15.29***	1.93*	-3.96***	0.54	-1.29	0.60	-14.01***	2.61**	-4.23***	0.75	-2.61	-0.82
	(0.76)	(1.13)	(1.02)	(1.28)	(1.53)	(1.60)	(0.95)	(1.30)	(1.28)	(1.48)	(1.85)	(2.01)
Cubic in '88 score	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y
Sample size	2,644	2,644	896	896	366	366	1,671	1,671	578	578	236	236
Panel C: **1992 SES Index on Eligibility**												
Eligible	-19.70***	-0.22	-7.25***	-0.65	-2.71	-0.68	-18.70***	2.08	-5.36***	1.31	-1.43	0.12
	(0.88)	(1.21)	(1.14)	(1.31)	(1.71)	(1.75)	(1.07)	(1.32)	(1.40)	(1.42)	(2.00)	(2.19)
Cubic in '88 score	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y
Sample size	2,591	2,591	883	883	363	363	1,640	1,640	569	569	233	233

Notes: Each cell reports a coefficient from a regression where the dependent variable is the 1990 SES index (higher values indicate higher SES). P900 indicates whether a school was treated. Eligible indicates whether a school's 1988 average score fell below its respective regional cutoff. ***, **, and * indicate statistical significance at the 1, 5, and 10 percent level, respectively. The sample covers urban schools with 15 or more students in the fourth grade in 1988. Huber-White standard errors are in parentheses.

Table A4: P-900 effects on 1988-1992 gain scores, alternate identification strategy

	Mathematics			Language		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: 1988 average score>42.4 and ≤43.4						
P-900	3.00	3.06	3.69*	2.73	2.77	2.96
	(1.91)	(2.03)	(2.00)	(1.75)	(1.84)	(2.01)
Cubic in 1988 score	N	Y	Y	N	Y	Y
SES controls	N	N	Y	N	N	Y
Panel B: 1988 average score>43.4 and ≤46.4						
P-900	-0.44	-0.54	-0.15	-0.13	-0.21	0.25
	(1.10)	(1.11)	(1.10)	(0.95)	(0.94)	(0.95)
Cubic in 1988 score	N	Y	Y	N	Y	Y
SES controls	N	N	Y	N	N	Y
Panel C: 1988 average score>46.4 and ≤47.4						
P-900	4.84***	4.98***	5.30***	4.18***	4.31***	4.85***
	(1.63)	(1.67)	(1.72)	(1.44)	(1.40)	(1.42)
Cubic in 1988 score	N	Y	Y	N	Y	Y
SES controls	N	N	Y	N	N	Y
Panel D: 1988 average score>47.4 and ≤49.4						
P-900	4.86***	4.85***	5.15***	3.29***	3.34***	3.65***
	(1.25)	(1.20)	(1.23)	(1.11)	(1.07)	(1.07)
Cubic in 1988 score	N	Y	Y	N	Y	Y
SES controls	N	N	Y	N	N	Y
Panel E: 1988 average score>49.4 and ≤51.4						
P-900	-0.42	-0.47	0.19	0.85	0.80	1.46
	(1.64)	(1.62)	(1.64)	(1.27)	(1.27)	(1.30)
Cubic in 1988 score	N	Y	Y	N	Y	Y
SES controls	N	N	Y	N	N	Y
Panel F: All test score intervals						
P-900	1.89***	1.88***	2.41***	1.71***	1.70***	2.20***
	(0.65)	(0.65)	(0.65)	(0.56)	(0.56)	(0.56)
Cubic in 1988 score	N	Y	Y	N	Y	Y
SES controls	N	N	Y	N	N	Y
Interval dummies	Y	Y	Y	Y	Y	Y

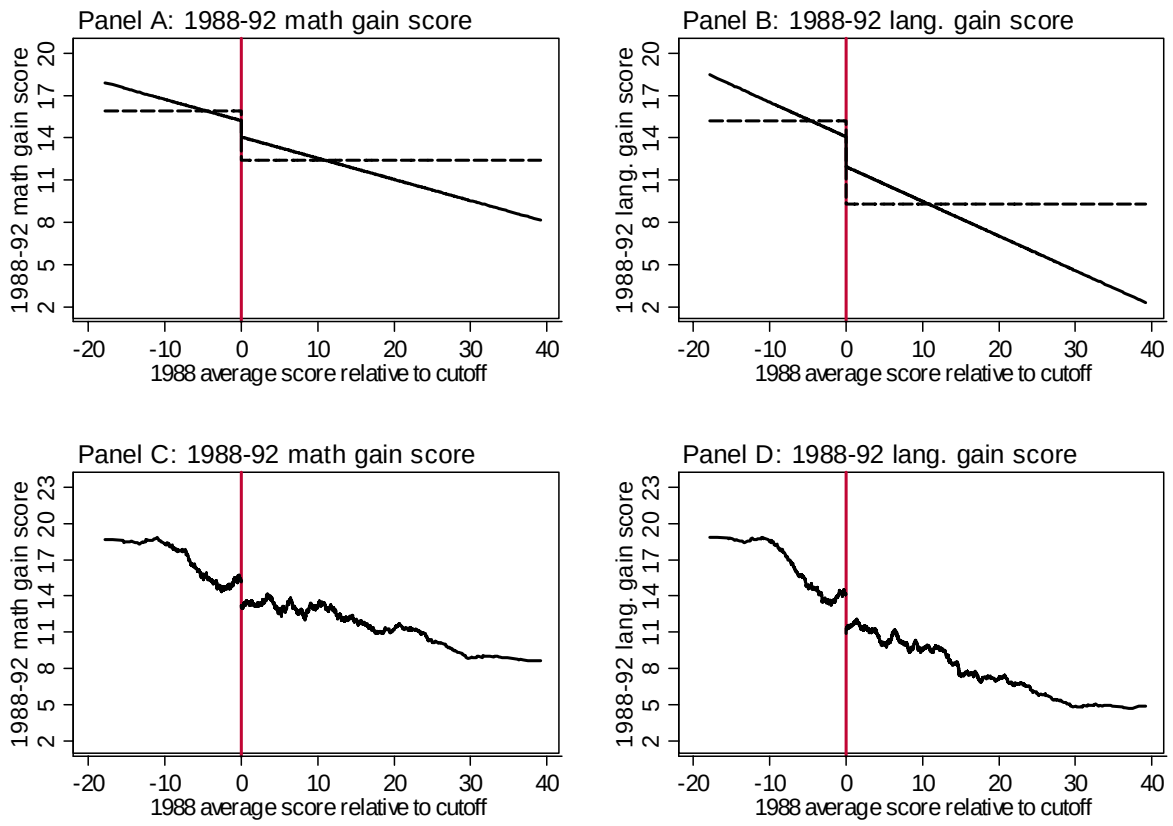
Notes: The "experiments" are as follows. For schools with test scores greater than 42.4 and less than or equal to 43.4, those in Regions 1-6 and 8-13 are eligible; Regions 7 is ineligible (N=70, of which 37 are treated). For schools with test scores greater than 43.4 and less than or equal to 46.4, those in Regions 1-5 and 9-13 are eligible; Regions 6-8 are ineligible (N=236, of which 88 are treated). For schools with test scores greater than 46.4 and less than or equal to 47.4, those in Regions 1-5, and 9-12 are eligible; Regions 6-8 and 13 are ineligible (N=95, of which 32 are treated). For schools with test scores greater than 47.4 and less than or equal to 49.4, those in Regions 1-5 and 10-12 are eligible; Regions 6-9 and 13 are ineligible (N=214, of which 37 are treated). For schools with test scores greater than 49.4 and less than or equal to 51.4, those in Regions 1, 3, 4, 11, and 12 are eligible; Regions 2, 5-10, and 13 are ineligible (N=196, of which 21 are treated). We have excluded a sixth potential "experiment" – schools with test scores between 51.4 and 52.4 – because it included only 5 treated schools. Panel F uses the pooled sample from Panels A-E and controls for dummy variables indicating test score intervals.

Table A5: Fictitious program effects, 1988-1992 language gain scores

	(1)	(2)	(3)
Panel A: "P-450"			
"P-450"	7.54 ***	-0.82	-1.23
	(0.55)	(0.87)	(1.73)
Cubic in 1988 score	N	Y	Y
SES controls	N	Y	Y
Within ± 5 points	N	N	Y
R^2	0.063	0.208	0.017
Sample Size	3,929	3,929	769
Panel B: "Reverse P-900"			
"Reverse P-900"	-6.26***	0.06	-1.04
	(0.31)	(0.52)	(1.04)
Cubic in 1988 score	N	Y	Y
SES controls	N	Y	Y
Within ± 5 points	N	N	Y
R^2	0.078	0.208	0.121
Sample Size	3,929	3,929	977

Notes: ***, **, and * indicate statistical significance at the 1, 5, and 10 percent level, respectively. Huber-White standard errors are in parentheses. "P-450" is the fictitious program obtained by selecting the lowest 10 percent of schools in 1988, using average scores. "Reverse P-900" selects the *top* 20 percent of performers in 1988. All specifications include the full sample of schools.

Figure A1: Reduced-form estimates



Notes: The sample includes urban schools with fourth grade enrollments of 15 or more. Panels A and B plot the OLS predictions from reduced-form regressions in column 1 (dotted line) and column 2 (solid lines) of Table 7. Panels C and D plot nonparametric predictions from an unweighted uniform kernel smoother. The bandwidths are 0.3 for eligible schools and 0.1 for ineligible schools.