Left Behind?
The Effect of *No Child Left Behind* on Academic Achievement Gaps

Sean F. Reardon
Erica Greenberg
Demetra Kalogrides
Kenneth A. Shores
Rachel A. Valentino

Stanford University

NCLB was intended, in part, to close racial achievement gaps:

"The purpose of this title is to ensure that all children have a fair, equal, and significant opportunity to obtain a high-quality education and reach, at a minimum, proficiency on challenging State academic achievement standards and state academic assessments. This purpose can be accomplished by…closing the achievement gap between high- and low- performing children, especially the achievement gaps between minority and nonminority students, and between disadvantaged children and their more advantaged peers (115 Stat. 1439-40)."

NCLB was intended, in part, to close racial achievement gaps:

"The purpose of this title is to ensure that all children have a fair, equal, and significant opportunity to obtain a high-quality education and reach, at a minimum, proficiency on challenging State academic achievement standards and state academic assessments. This purpose can be accomplished by…closing the achievement gap between high- and low- performing children, especially the achievement gaps between minority and nonminority students, and between disadvantaged children and their more advantaged peers (115 Stat. 1439-40)."

Has it?

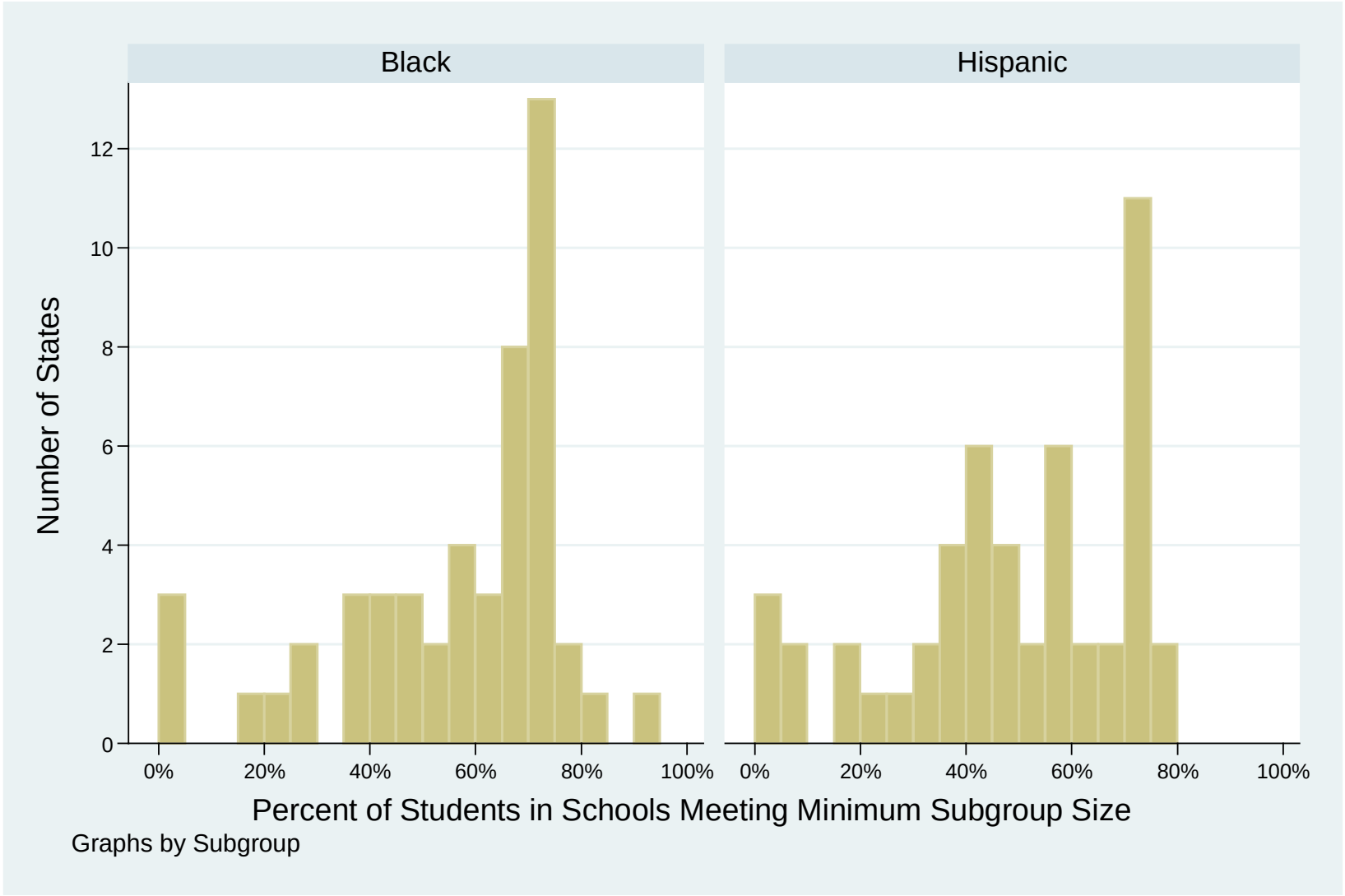
Why/how might NCLB have affected achievement gaps?

- "informational aspects" of NCLB
  - testing of all students in grades 3-8;
  - reporting by subgroup
- subgroup accountability targets
  - requirement to make "adequate yearly progress" for all subgroups
- highly qualified teacher provision
- increased support for supplemental services for children in underperforming schools

Potential variation among states in effects of NCLB

- test scores are not reported separately within schools in which a subgroup contained too few students to yield reliable information regarding the subgroup's performance
  - states set the minimum subgroup size threshold
  - ranged from 5 (MD) to 100 (CA); most states set it at 30-40
- states vary widely in the proportion of black/Hispanic students who are in schools where they meet this minimum subgroup size, because of variation among states in
  - racial composition
  - between-school racial segregation
  - average school size
  - minimum subgroup size
- If NCLB operates through informational aspects and/or through subgroup-specific accountability pressure, NCLB may put more upward pressure on black/Hispanic students' scores in states where most are in schools where their scores were reported separately.

Distribution of Proportions of Black and Hispanic Students in Schools Meeting Minimum Subgroup Reporting Size



We fit precision-weighted random-coefficients models (pooling data across subjects and data sources in some cases):

$$\hat{G}_{csg} = (\lambda + u_{\lambda s}) + (\gamma + u_{\gamma s})(coh_c^*) + (\alpha + u_{\alpha s})(gr_g) + \beta(gr_g \cdot coh_c^*) + \eta(E_g) + \mathbf{X}_{cs}\mathbf{A} + \mathbf{W}_{csg}\mathbf{B} + (\delta + u_{\delta s})(exp_{cg}) + e'_{csg} + \epsilon_{csg}$$

$$e'_{csg} \sim N[0, \sigma^2]$$
$$\epsilon_{csg} \sim N[0, \omega^2_{csg}] = N[0, var(\hat{G}_{csg})]$$
$$\begin{bmatrix} u_{\lambda s} \\ u_{\gamma s} \\ u_{\alpha s} \\ u_{\delta s} \end{bmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_\lambda & \tau_{\lambda\gamma} & \tau_{\lambda\alpha} & \tau_{\lambda\delta} \\ \tau_{\gamma\lambda} & \tau_\gamma & \tau_{\gamma\alpha} & \tau_{\gamma\delta} \\ \tau_{\alpha\lambda} & \tau_{\alpha\gamma} & \tau_\alpha & \tau_{\alpha\delta} \\ \tau_{\delta\lambda} & \tau_{\delta\gamma} & \tau_{\delta\alpha} & \tau_\delta \end{pmatrix}\right].$$

where exp_{cg} is the number of years cohort c has been exposed to NCLB by grade g.

The parameters of interest are δ, the average (across states) effect of a year's exposure to NCLB, and τ_δ, the variance of this effect across states.

Number of Years Exposed to NCLB by Spring of School Year, by Calendar Year and Grade

	Year (Spring of Academic Year)																					
Grade	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
K	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1
1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	2	2	2	2	2	2	2
2	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	3	3	3	3	3	3	3
3	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	3	4	4	4	4	4	4
4	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	3	4	5	5	5	5	5
5	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	3	4	5	6	6	6	6
6	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	3	4	5	6	7	7	7
7	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	3	4	5	6	7	8	8
8	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	3	4	5	6	7	8	9

Pre-2003 kindergarten cohort; not subject to NCLB in current year

Pre-2003 kindergarten cohort; subject to NCLB in current year

Post-2002 kindergarten cohort; subject to NCLB in current year

Number of Years Exposed to NCLB by Spring of School Year, by Calendar Year and Grade

	Year (Spring of Academic Year)																					
Grade	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
K	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1
1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	2	2	2	2	2	2	2
2	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	3	3	3	3	3	3	3
3	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	3	4	4	4	4	4	4
4	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	3	4	5	5	5	5	5
5	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	3	4	5	6	6	6	6
6	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	3	4	5	6	7	7	7
7	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	3	4	5	6	7	8	8
8	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	3	4	5	6	7	8	9

Pre-2003 kindergarten cohort; not subject to NCLB in current year
Pre-2003 kindergarten cohort; subject to NCLB in current year
Post-2002 kindergarten cohort; subject to NCLB in current year

Number of Years Exposed to NCLB by Spring of School Year, by Kindergarten Cohort and Grade

	Cohort (Fall of Kindergarten Entry Year)																					
Grade	….	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
K	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	2	2	2	2	2	2
2	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	3	3	3	3	3	3	3
3	0	0	0	0	0	0	0	0	0	0	0	0	1	2	3	4	4	4	4	4	4	4
4	0	0	0	0	0	0	0	0	0	0	0	1	2	3	4	5	5	5	5	5	5	5
5	0	0	0	0	0	0	0	0	0	0	1	2	3	4	5	6	6	6	6	6	6	6
6	0	0	0	0	0	0	0	0	0	1	2	3	4	5	6	7	7	7	7	7	7	7
7	0	0	0	0	0	0	0	0	1	2	3	4	5	6	7	8	8	8	8	8	8	8
8	0	0	0	0	0	0	0	1	2	3	4	5	6	7	8	9	9	9	9	9	9	9

Pre-2003 kindergarten cohort; not subject to NCLB in current year
Pre-2003 kindergarten cohort; subject to NCLB in current year
Post-2002 kindergarten cohort; subject to NCLB in current year

Data

- NAEP data (State NAEP) from 1996-2009
- State test data from 1997-2011 (most data from 2003-2010)
- From both, we estimate state-level achievement gaps (and their standard errors):
  - White-black and white-Hispanic achievement gaps
  - Math and reading gap estimates
  - Grades 2-8 (NAEP is grades 4 & 8)
- Three measures of achievement gap:
  - Standardized effect sizes (Cohen's d)
  - Matric-free quasi-effect sizes (V statistic)
  - Proficiency gaps (differences in proportions proficient)

Number of Available Achievement Gap Estimates from State Test Data, by Kindergarten Cohort and Grade

	Cohort (Fall of Kindergarten Entry Year)																				
Grade	….	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
K																						
1																						
2														4	8	8	12	12	12	12	3	3
3									3	8	25	50	60	83	119	174	190	201	197	178	4	
4						2	2	28	39	49	58	66	104	123	175	190	198	196	179	6		
5							3	10	29	53	62	83	121	178	193	201	198	177	7			
6						8	12	32	40	46	69	86	176	192	202	201	178	5				
7					6	6	18	30	32	64	102	180	193	202	195	174	7					
8					37	46	72	85	116	143	176	192	202	199	175	6						

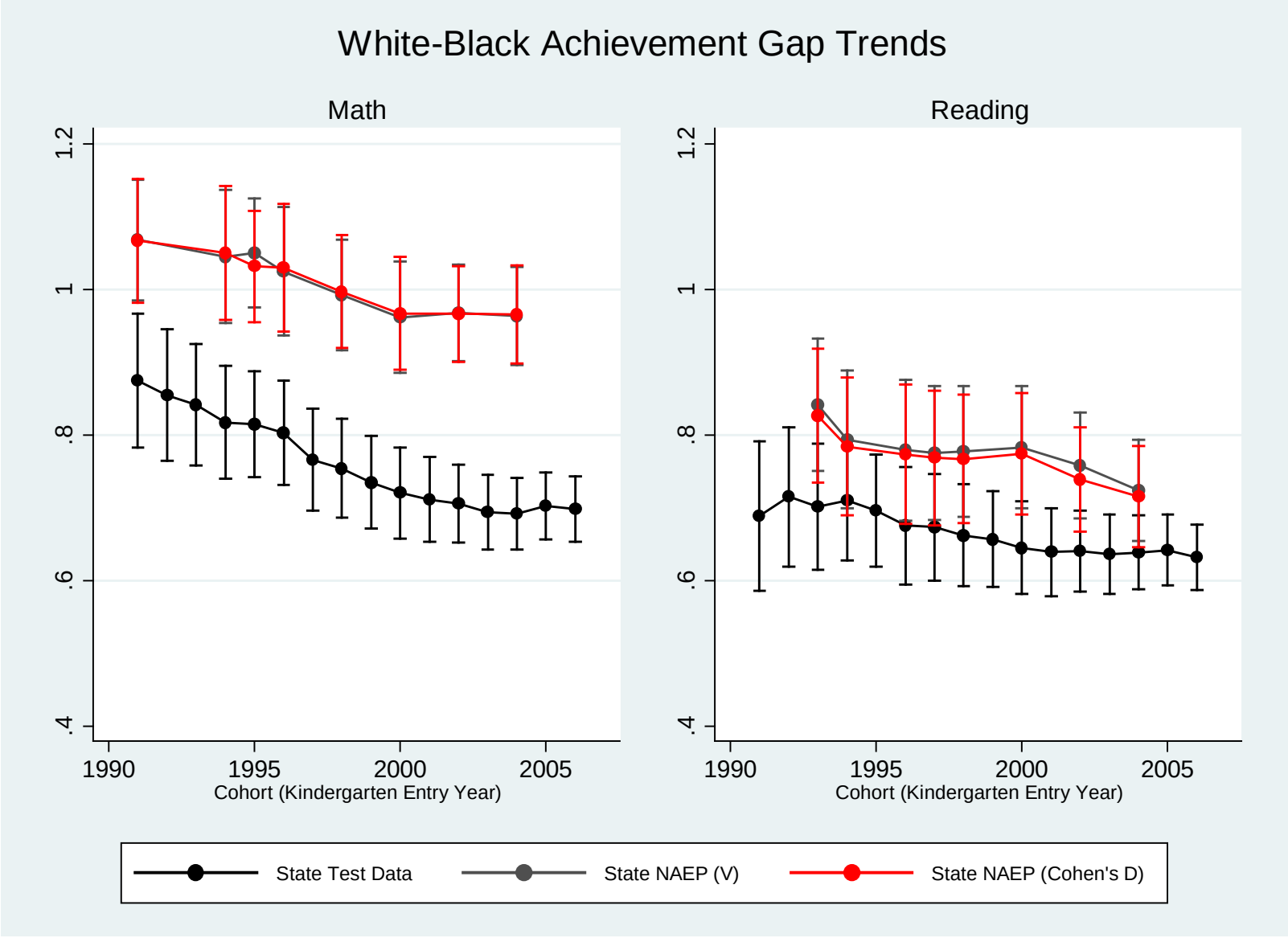
Note: counts indicate the total number of state-by-subject (math or reading)-by-group (white-black or white-Hispanic) available for a given cohort-grade cell. Maximum possible count is 204 (51 states x 2 subjects x 2 groups).

Number of Available Achievement Gap Estimates from NAEP Data, by Kindergarten Cohort and Grade

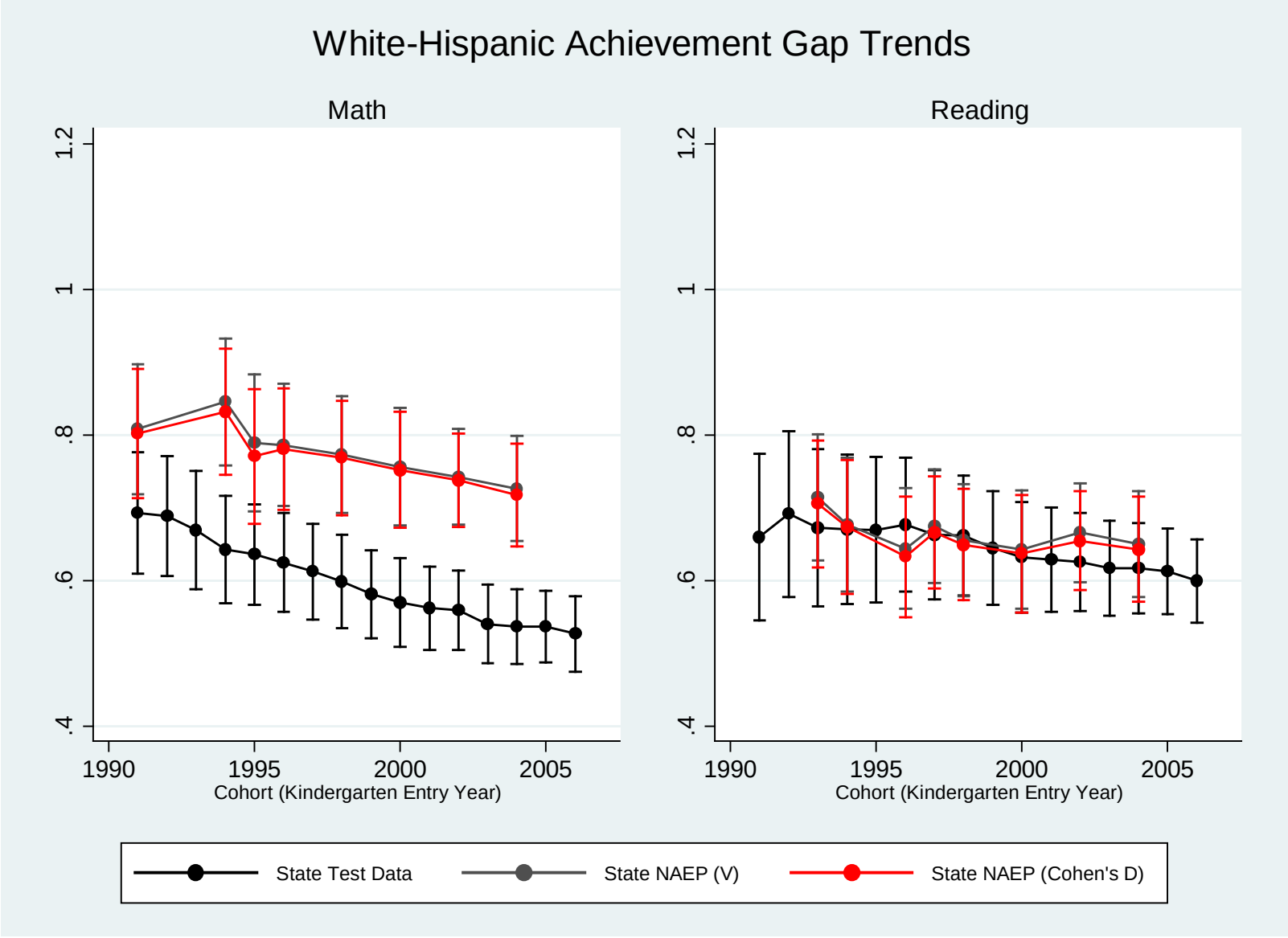
	Cohort (Fall of Kindergarten Entry Year)																				
Grade	….	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
K																						
1																						
2																						
3																						
4					88		82		84		95	204		204		204		204				
5																						
6																						
7																						
8					82		95	204		204		204		204								

Note: counts indicate the total number of state-by-subject (math or reading)-by-group (white-black or white-Hispanic) available for a given cohort-grade cell. Maximum possible count is 204 (51 states x 2 subjects x 2 groups).

White-Black Achievement Gap Trends, Math and Reading, 1991-2006 Cohorts



White-Hispanic Achievement Gap Trends, Math and Reading, 1991-2006 Cohorts



Estimated Annual Effect of NCLB on White-Black Achievement Gap

Math and Reading Gaps Pooled

	Zero/Partial NCLB Exposure	Full/Partial NCLB Exposure	All Observations
All data (V)	0.008	-0.004	-0.005
	(0.006)	(0.004)	(0.005)
NAEP data (V)	0.008 +		0.008 +
	(0.004)		(0.004)
State data (V)	0.011	-0.002	-0.006 *
	(0.007)	(0.004)	(0.003)
State data (proficiency gap)	-0.002	-0.678 *	-0.949 ***
	(0.454)	(0.282)	(0.234)

Each cell indicates the estimated annual effect of exposure to NCLB. Each coefficient is from a separate model. Robust standard errors are in parentheses. + p<.10; * p<.05; ** p<.01; *** p<.001.

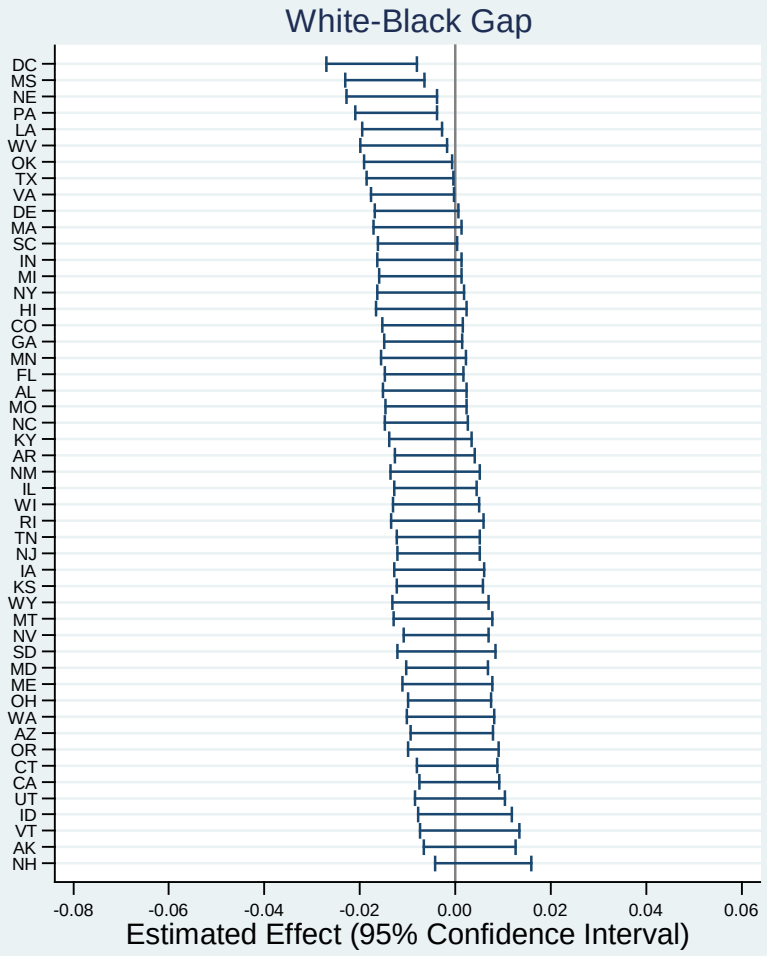
Estimated Annual Effect of NCLB on White-Hispanic Achievement Gap

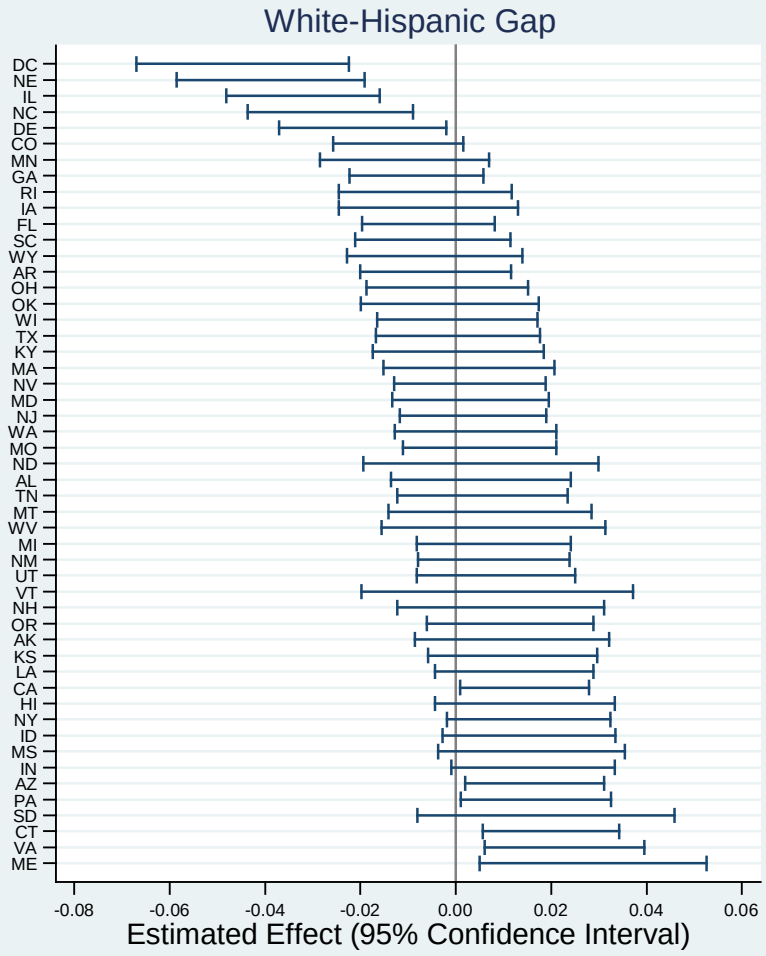
Math and Reading Gaps Pooled

	Zero/Partial NCLB Exposure	Full/Partial NCLB Exposure	All Observations
All data (V)	-0.005	0.010 +	0.003
	(0.007)	(0.005)	(0.004)
NAEP data (V)	-0.002		0.001
	(0.006)		(0.005)
State data (V)	-0.006	0.014 **	0.006
	(0.016)	(0.005)	(0.005)
State data (proficiency gap)	-0.750 +	-0.011	-0.375
	(0.415)	(0.310)	(0.268)

Each cell indicates the estimated annual effect of exposure to NCLB. Each coefficient is from a separate model. Robust standard errors are in parentheses. + p<.10; * p<.05; ** p<.01; *** p<.001.

Estimated Annual NCLB Effects, by State

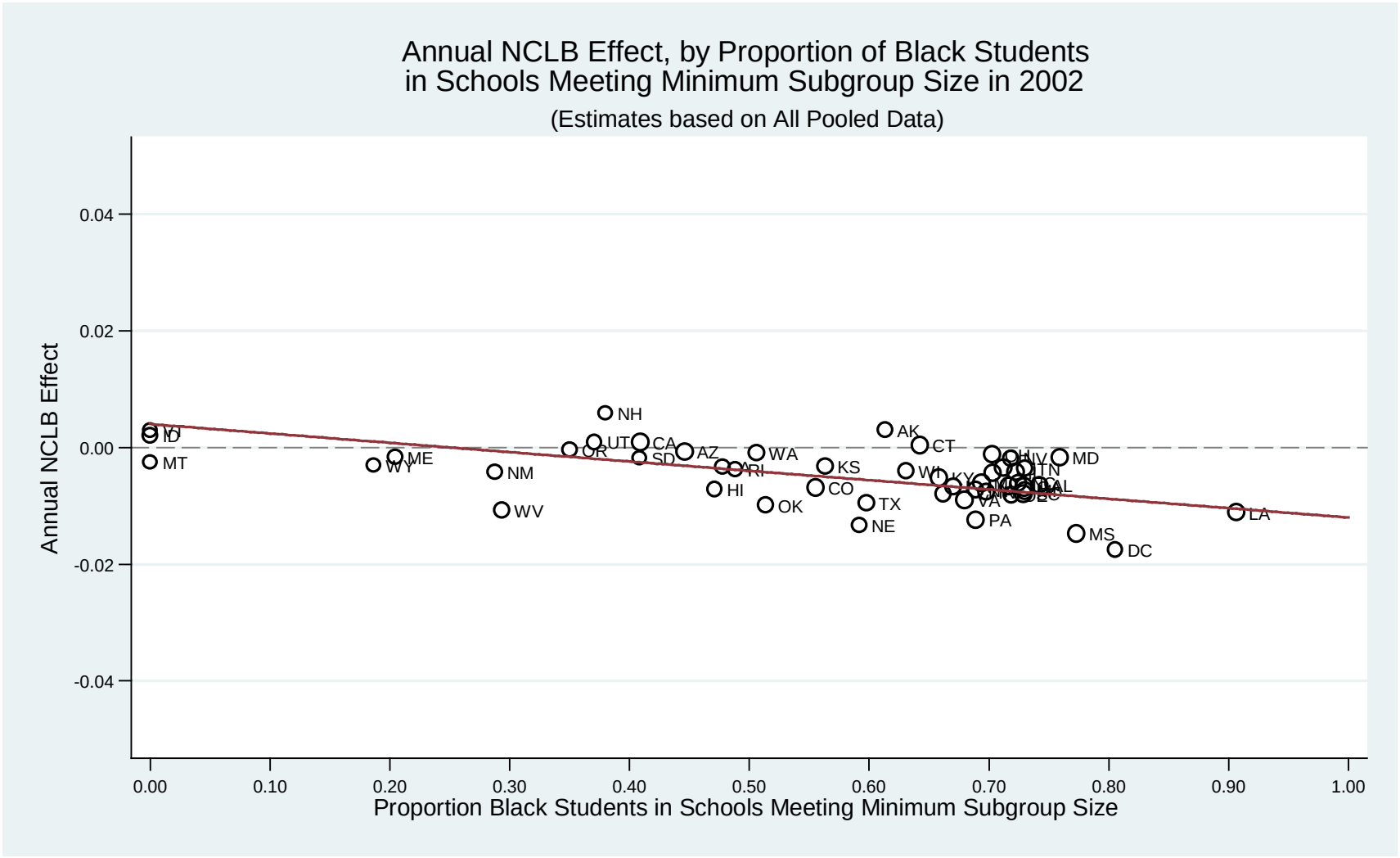


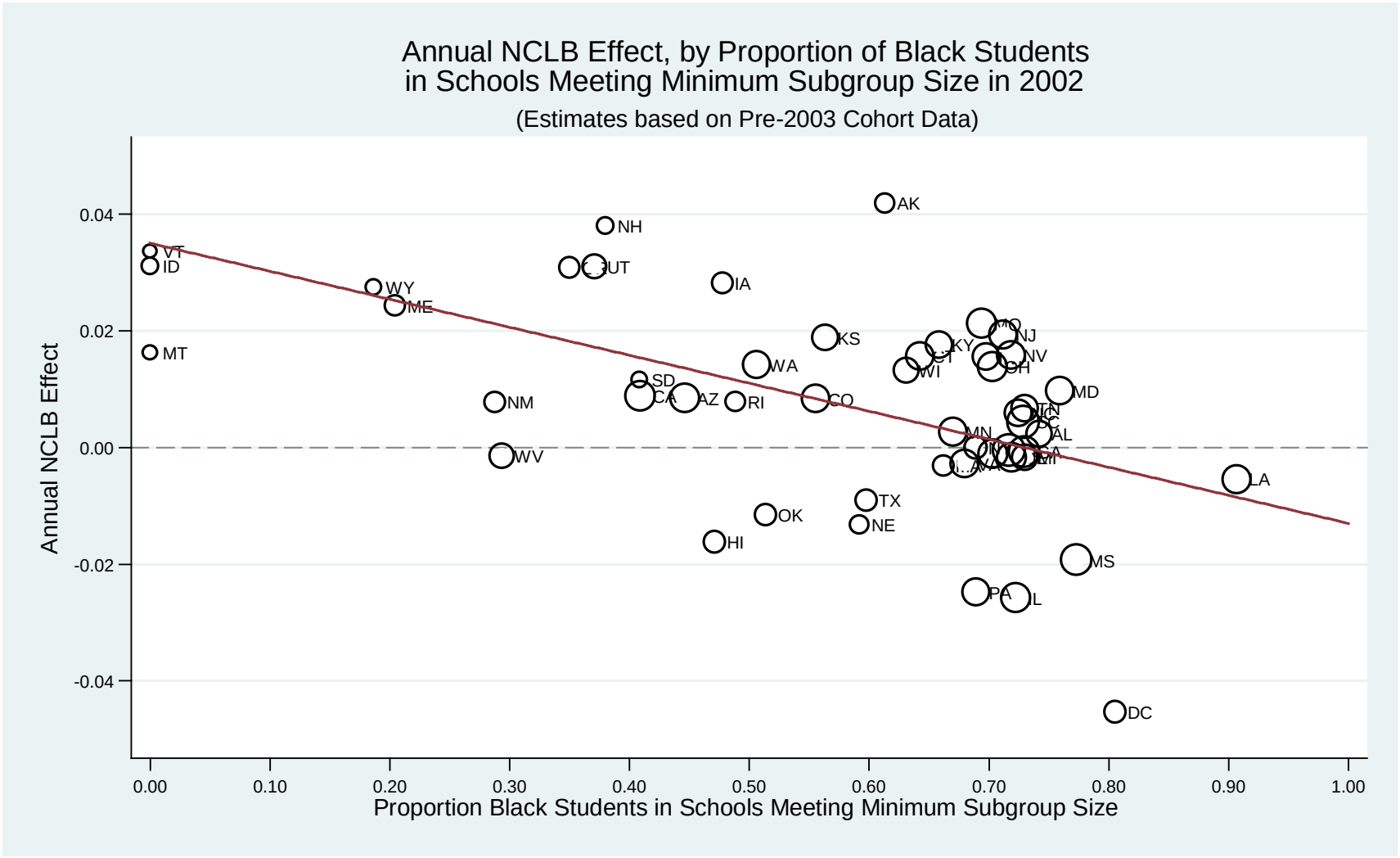


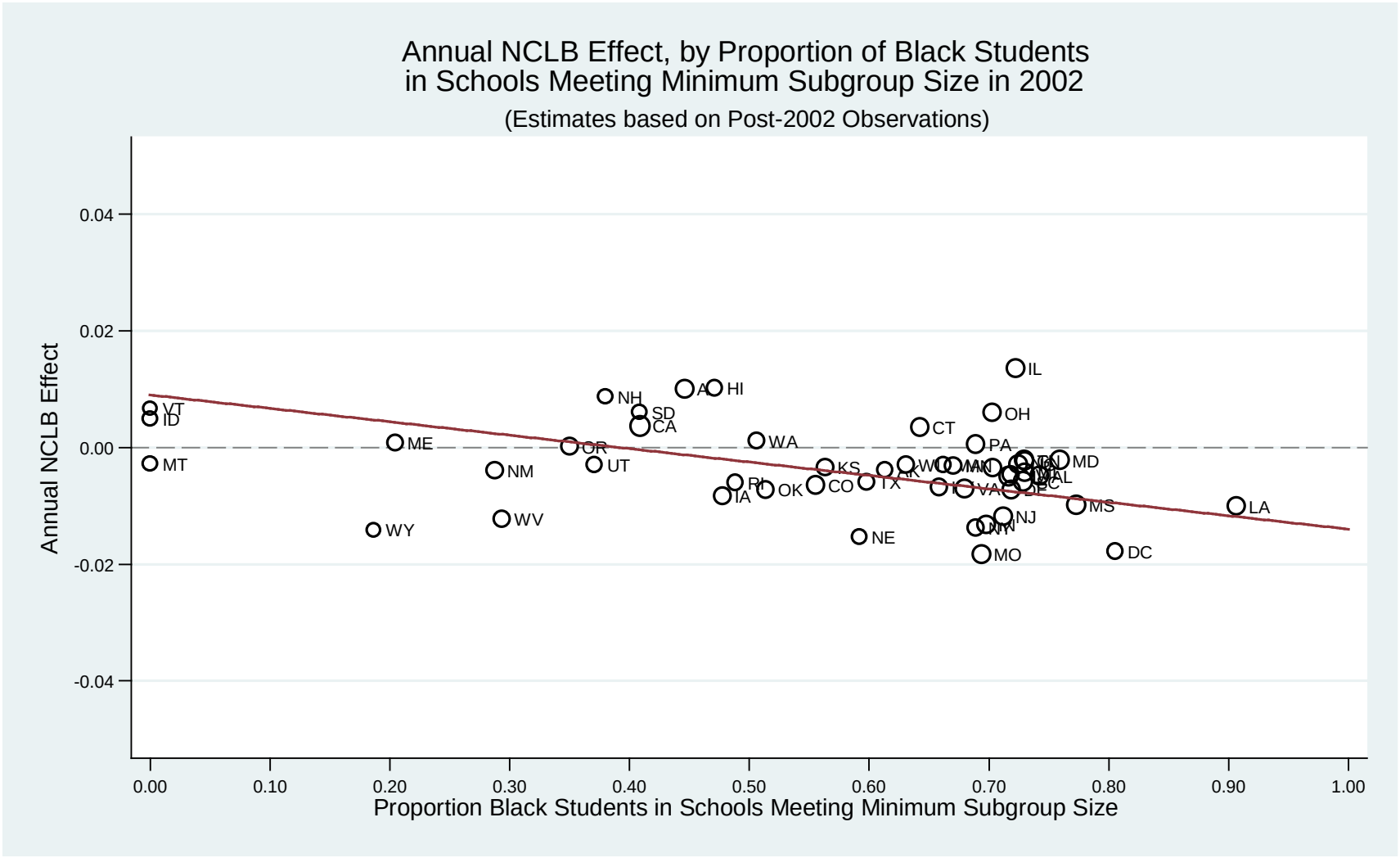
Estimated Annual Effect of NCLB on the Achievement Gaps, by the Proportion of Black or Hispanic Students in Schools Subject to Accountability

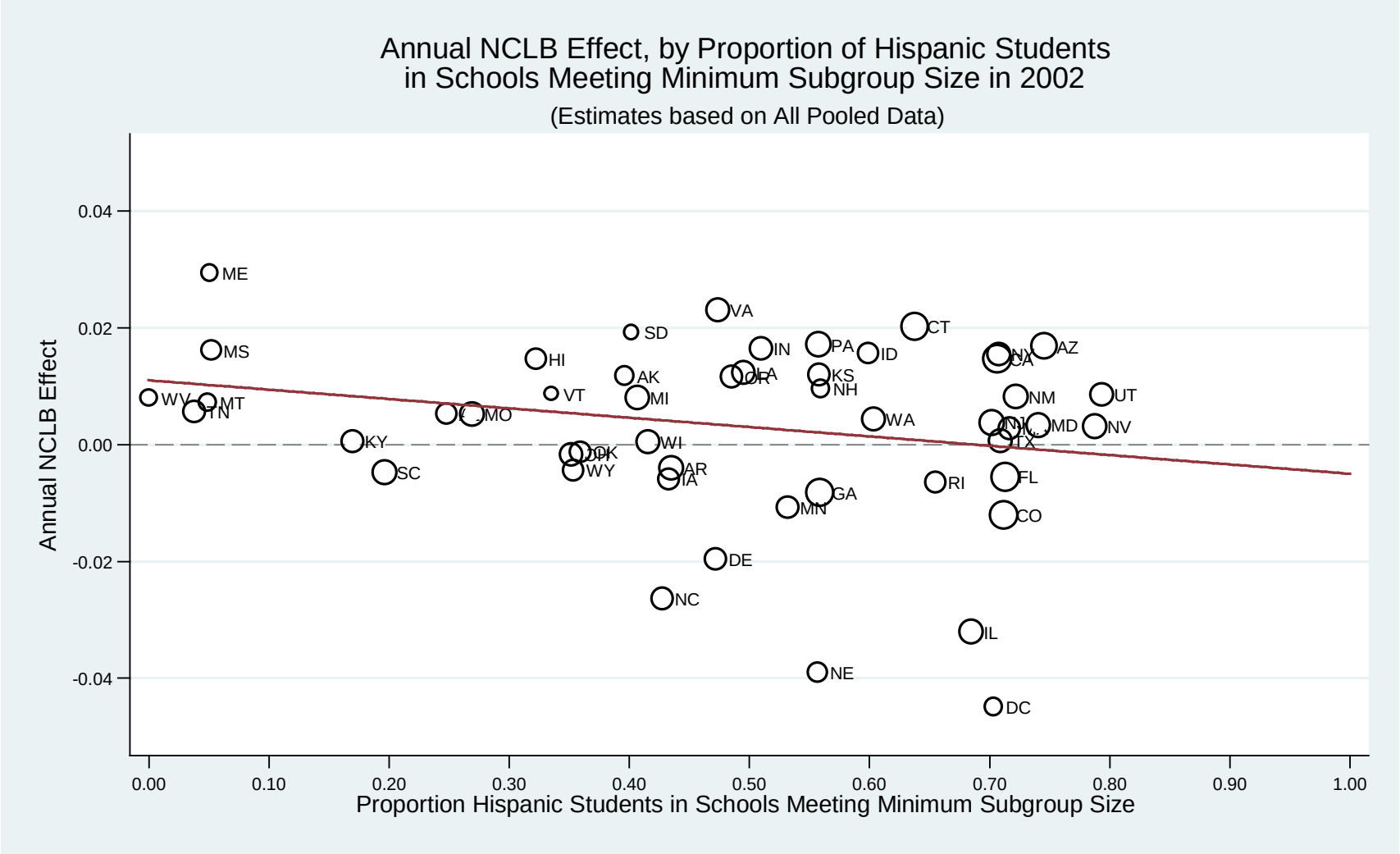
	Annual Effect of NCLB on White-Black Gap						Annual Effect of NCLB on White-Hispanic Gap					
	Zero/Partial NCLB Exposure Observations		Full/Partial NCLB Exposure Observations		All Observations		Zero/Partial NCLB Exposure Observations		Full/Partial NCLB Exposure Observations		All Observations	
All data (V)							**All data (V)**					
Exposure	0.035	***	0.009		0.004		0.000		0.017	**	0.011	+
	(0.008)		(0.007)		(0.006)		(0.010)		(0.006)		(0.006)	
Exposure x % Accountable	-0.048	***	-0.023	**	-0.016	*	-0.014		-0.015		-0.016	
	(0.012)		(0.009)		(0.008)		(0.017)		(0.012)		(0.013)	
State data (proficiency gap)							**State data (proficiency gap)**					
Exposure	0.468		-0.266		-0.550	+	-0.245		0.342		0.072	
	(0.538)		(0.335)		(0.319)		(0.519)		(0.378)		(0.342)	
Exposure x % Accountable	-1.115	*	-0.728	+	-0.737	+	-1.418		-0.796		-0.980	
	(0.521)		(0.439)		(0.427)		(0.892)		(0.637)		(0.628)	

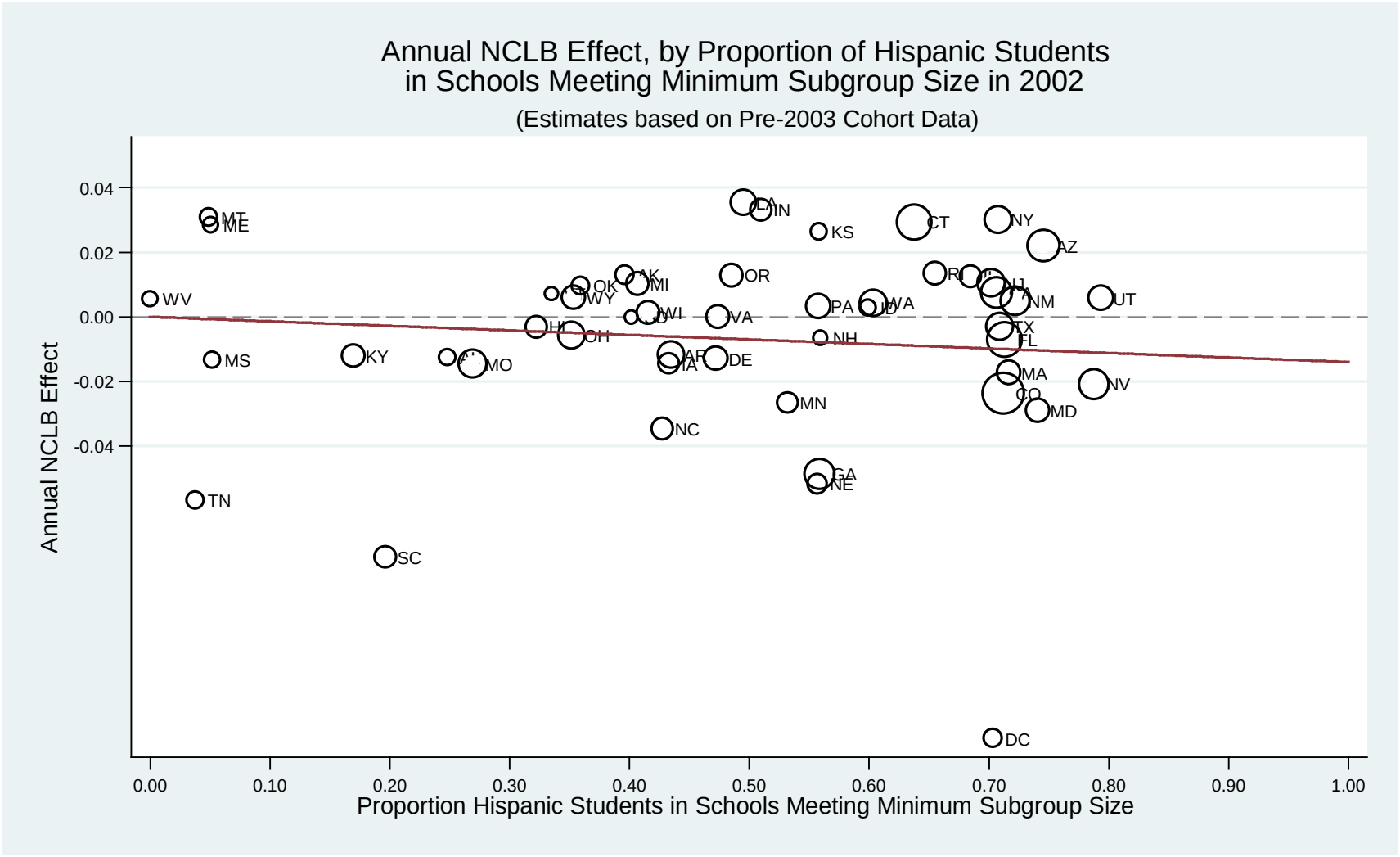
Note: All models include controls for grade, cohort, and time-varying economic and school composition and segregation covariates. Robust standard errors are in parentheses. + p <.10; * p <.05; ** p <.01; *** p <.001.

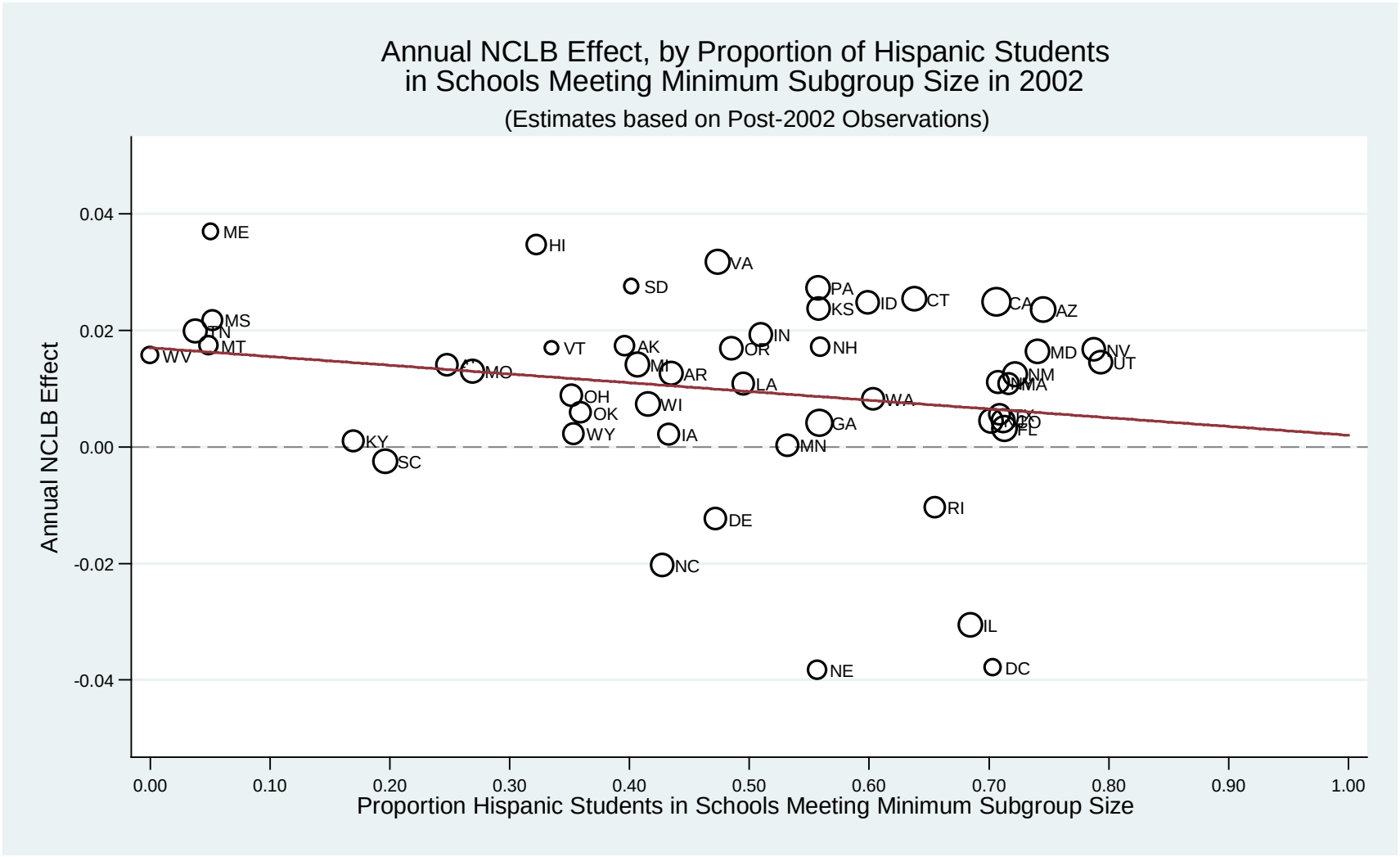

Annual NCLB Effect, by Proportion of Black Students
in Schools Meeting Minimum Subgroup Size in 2002
(Estimates based on All Pooled Data)
Annual NCLB Effect
0.04
0.02
0.00
-0.02
-0.04
0.00
0.10
0.20
0.30
0.40
0.50
0.60
0.70
0.80
0.90
1.00
Proportion Black Students in Schools Meeting Minimum Subgroup Size
NH
AK
UT
CA
CT
OR
SD
AZ
WA
MD
ID
MT
WY
ME
NM
KS
HI
CO
OK
TX
WV
NE
PA
MS
DC
LA


Annual NCLB Effect, by Proportion of Black Students
in Schools Meeting Minimum Subgroup Size in 2002
(Estimates based on Pre-2003 Cohort Data)
Annual NCLB Effect
0.04
0.02
0.00
-0.02
-0.04
0.00
0.10
0.20
0.30
0.40
0.50
0.60
0.70
0.80
0.90
1.00
Proportion Black Students in Schools Meeting Minimum Subgroup Size
AK
NH
VT
ID
UT
IA
WY
ME
KS
NJ
MT
WA
NV
SD
MD
NM
CA
AZ
RI
CO
AL
WV
LA
TX
OK
NE
HI
MS
PA
DC


Annual NCLB Effect, by Proportion of Black Students
in Schools Meeting Minimum Subgroup Size in 2002
(Estimates based on Post-2002 Observations)
Annual NCLB Effect
0.04
0.02
0.00
-0.02
-0.04
0.00
0.10
0.20
0.30
0.40
0.50
0.60
0.70
0.80
0.90
1.00
Proportion Black Students in Schools Meeting Minimum Subgroup Size
VT
ID
MT
ME
WY
NM
WV
OR
UT
NH
SD
CA
AZ
HI
WA
IA
OK
CO
KS
TX
IL
CT
OH
PA
MD
MS
LA
NJ
NE
MO
DC


Annual NCLB Effect, by Proportion of Hispanic Students
in Schools Meeting Minimum Subgroup Size in 2002
(Estimates based on All Pooled Data)
Annual NCLB Effect
0.04
0.02
0.00
-0.02
-0.04
ME
MS
WV
MT
TN
VA
SD
HI
AK
VT
MI
IN
PA
ID
CT
AZ
KS
NH
NM
UT
MO
WA
MD
NV
KY
WI
SC
WY
AR
IA
RI
FL
GA
MN
CO
DE
NC
IL
NE
DC
0.00
0.10
0.20
0.30
0.40
0.50
0.60
0.70
0.80
0.90
1.00
Proportion Hispanic Students in Schools Meeting Minimum Subgroup Size


Annual NCLB Effect, by Proportion of Hispanic Students
in Schools Meeting Minimum Subgroup Size in 2002
(Estimates based on Pre-2003 Cohort Data)
Annual NCLB Effect
0.04
0.02
0.00
-0.02
-0.04
MT
ME
WV
MS
TN
SC
KY
MO
OK
MI
WY
OH
OR
VA
KS
IN
CT
NY
AZ
PA
WA
NM
UT
NH
TX
FL
MA
NV
CO
MD
MN
NC
GA
NE
DE
DC
0.00
0.10
0.20
0.30
0.40
0.50
0.60
0.70
0.80
0.90
1.00
Proportion Hispanic Students in Schools Meeting Minimum Subgroup Size


Annual NCLB Effect, by Proportion of Hispanic Students
in Schools Meeting Minimum Subgroup Size in 2002
(Estimates based on Post-2002 Observations)
Annual NCLB Effect
0.04
0.02
0.00
-0.02
-0.04
ME
HI
VA
SD
PA
KS
ID
CT
CA
AZ
MS
TN
MT
WV
VT
AK
IN
OR
NH
MD
NV
UT
MO
MI
AR
LA
NM
MA
OH
OK
WI
WA
GA
WY
IA
MN
KY
SC
RI
DE
NC
IL
NE
DC
0.00
0.10
0.20
0.30
0.40
0.50
0.60
0.70
0.80
0.90
1.00
Proportion Hispanic Students in Schools Meeting Minimum Subgroup Size

Conclusions

- Overall, there is little evidence that NCLB systematically reduced achievement gaps
  - Some evidence that gaps widened
  - Estimates are very precise (we can rule out effects larger than ± 0.02 standard deviations/year)

- However, the effects vary among states
  - White-Black gaps narrowed the most as a result of NCLB in states where most black students were in schools where their scores were reported (states with large black populations, high levels of segregation, small schools, and low minimum subgroup reporting thresholds – e.g., LA, MS, DC)
  - White-Hispanic gaps show a similar pattern (but much less precisely estimated, and not significant).
  - This pattern is consistent with a theoretical model of informational effects and/or accountability pressure

Some remaining puzzles

- Why are the estimated effects somewhat different in the two identification strategies (or different when we use both identification strategies simultaneously that when we use either along)?
  - NCLB more effective in early grades than in late grades?
  - NCLB more effective in later years than early years (implementation delays)?
  - Model misspecified?
  - Bias due to missing factors that change contemporaneously with exposure to NCLB?
- Why are the estimated effects somewhat different for white-black and white-Hispanic gaps?
  - Overlap of Hispanic and EL populations (Hispanic EL students may be counted zero, one, or two times for AYP)?
  - Concentration of most of Hispanic population in relatively few states

Model Derivation

G_{csg} is the achievement gap in the spring of grade g for students in cohort c in state s

- G_{cs0} is the gap for cohort c in state s in the spring of their kindergarten year
- $G_{cs(-1)}$ is the gap when these children entered kindergarten

coh_c^* is the cohort's year of kindergarten entry, centered on 2002

Now let the gap at kindergarten entry follow a state-specific linear trend, plus some effect of the vector of time-varying state covariates $\mathbf{X}_{cs}$ and a mean-zero error term:

$$G_{cs(-1)} = \lambda_s + \gamma_s(coh_c^*) + \mathbf{X}_{cs}\mathbf{A} + \nu_{cs}.$$

Let Δ_{csg} be the change in the gap in state s in cohort c during grade g:

$$G_{csg} = G_{cs(g-1)} + \Delta_{csg}$$

Now write Δ_{csg} as a function of a state fixed effect (v_s), a linear cohort effect (β), a linear grade effect (η), an effect of some cohort-state-grade specific vector of covariates $\mathbf{w}_{csg}$, a state-specific effect of the presence of NCLB (δ_s), and a mean-zero error term (e_{csg}):

$$\Delta_{csg} = \alpha + v_s + \beta(coh_c^*) + \eta(g) + \mathbf{w}_{csg}\mathbf{B} + \delta_s T_{cg} + e_{csg},$$

where T_{cg} is a variable indicating whether NCLB was in effect when cohort c was in grade g.

We want to estimate δ_s, the annual effect of exposure to NCLB in state s.

Substituting [3] into [2] recursively, we get

$$G_{csg} = G_{cs(-1)} + \sum_{k=0}^{g} \Delta_{csk}$$

$$G_{csg} = G_{cs(-1)} + \sum_{k=0}^{g} \Delta_{csk}$$

$$= [\lambda_s + \gamma_s(coh_c^*) + \mathbf{X}_{cs}\mathbf{A} + \nu_{cs}]$$
$$+ \sum_{k=0}^{g} [\alpha + \upsilon_s + \beta(coh_c^*) + \eta(k) + \delta_s T_{ck} + \mathbf{w}_{csk}\mathbf{B} + e_{csk}]$$

$$= [\lambda_s + \gamma_s(coh_c^*) + \mathbf{X}_{cs}\mathbf{A} + \nu_{cs}] + (g+1)\left(\alpha + \upsilon_s + \beta(coh_c^*)\right)$$
$$+ \eta\left(\sum_{k=0}^{g} k\right) + \delta_s\left(\sum_{k=0}^{g} T_{ck}\right) + \left(\sum_{k=0}^{g} \mathbf{w}_{csk}\right)\mathbf{B} + \sum_{k=0}^{g} e_{csk}$$

$$= \lambda_s + \gamma_s(coh_c^*) + \alpha_s(gr_g) + \beta(gr_g \cdot coh_c^*) + \eta(E_g) + \delta_s(exp_{cg})$$
$$+ \mathbf{X}_{cs}\mathbf{A} + \mathbf{W}_{csg}\mathbf{B} + e'_{csg},$$

where $gr_g = g + 1$; $E_g = \sum_{k=0}^{g} k = \frac{1}{2}(gr_g^2 - gr_g)$; $exp_{cg} = \sum_{k=0}^{g} T_{ck}$;
$\mathbf{W}_{csg} = \sum_{k=0}^{g} \mathbf{w}_{csk}$; and $e'_{csg} = \nu_{cs} + \sum_{k=0}^{g} e_{csk}$.

Identification

The partial derivative of the model with respect to coh (holding grade and state constant) is:

$$\frac{\partial G}{\partial coh} = \begin{cases} \gamma_s + \beta \cdot gr & if \quad T = 0, coh \leq 2002 \\ \gamma_s + \beta \cdot gr + \delta_s & if \quad T = 1, coh \leq 2002 \\ \gamma_s + \beta \cdot gr & if \quad T = 1, coh > 2002 \end{cases}$$

The partial derivative with respect to gr (holding cohort and state constant) is:

$$\frac{\partial G}{\partial gr} = \begin{cases} \alpha_s + \beta coh + \eta(2gr + 1) & if \quad T = 0 \\ \alpha_s + \beta coh + \eta(2gr + 1) + \delta_s & if \quad T = 1 \end{cases}$$

So there are two sources of identification of δ: the change in within-cohort grade trends or within-grade cohort trends between the pre- and post NCLB years:

$$\delta_s = \left[\frac{\partial G}{\partial coh}\middle| s, gr, coh \leq 2002, T = 1\right] - \left[\frac{\partial G}{\partial coh}\middle| s, gr, coh \leq 2002, T = 0\right]$$

$$\delta_s = \left[\frac{\partial G}{\partial gr}\middle| s, coh, gr, T = 1\right] - \left[\frac{\partial G}{\partial gr}\middle| s, coh, gr, T = 0\right]$$

and the change in within-grade gap trends between the pre- and post 2002 cohorts:

$$\delta_s = \left[\frac{\partial G}{\partial coh}\middle| s, gr, coh > 2002, T = 1\right] - \left[\frac{\partial G}{\partial coh}\middle| s, gr, coh \leq 2002, T = 1\right]$$

NAEP vs State Test Data

	NAEP	State Tests
Gap measures		
Standardized effect size (Cohen's d)	x	
V (from micro data)	x	
V (from proficiency counts)		x
Proficiency gap		x
Cohorts		
Pre-NCLB (zero exposure)	x	sparse
Partial exposure	x	x
Full exposure		x
Grades	4,8	2-8
Years	biennial	annual
Stakes	low	high
Common Test Across States	yes	no
Common Test Across Time	yes	no
Sample Size	~2,500	~100,000
N (state-by-cohort-by-grade-by-subject-by-group)	2,158	8,001

What is V?

Computing a standardized effect size requires knowing the means and standard deviations of the two group's test score distributions.

But when we have only proficiency counts (as in the case of the state test data), we don't know means and standard deviations, so we cannot compute d, the standardized effect size.

If $P_{a>b}$ is the probability that a randomly chosen member of group a has a test score higher than a randomly chosen member of group b, we define

$$V = \sqrt{2}\Phi^{-1}(P_{a>b}).$$

V can be computed directly from micro data (because we can estimate $P_{a>b}$ very precisely from micro data, with no distributional assumptions).

V can be estimated from aggregate proficiency count data very reliably, under modest distributional assumptions (and is relatively insensitive to failures of those distributional assumptions).

What is V?

If $P_{a>b}$ is the probability that a randomly chosen member of group a has a test score higher than a randomly chosen member of group b, then we define

$$V = \sqrt{2}\Phi^{-1}(P_{a>b}).$$

If there is some monotonic transformation of the test score scale that renders the test score distributions of groups a and b normal (though not necessarily with equal variance), then in that test metric

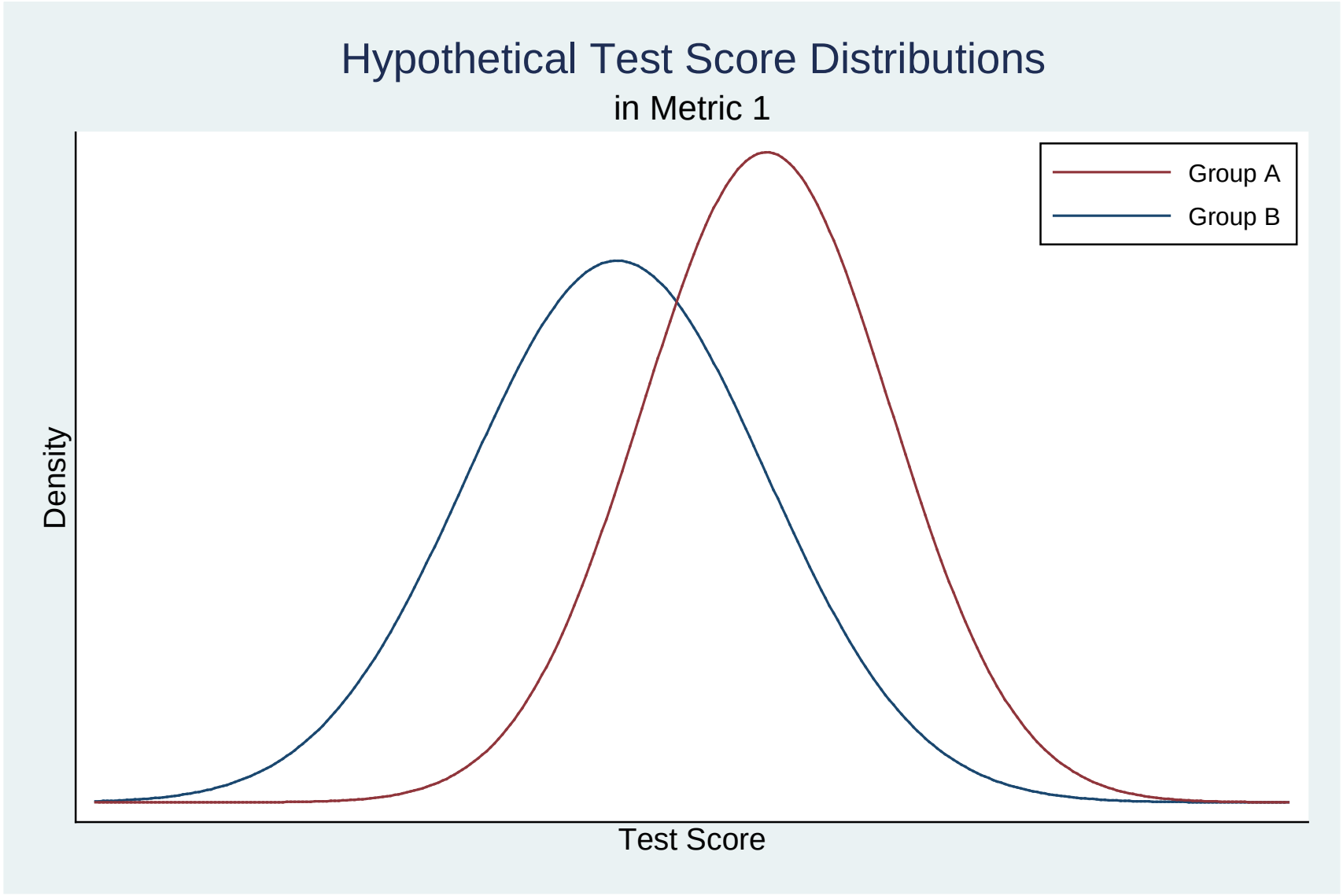
$$V = \frac{\bar{Y}_a - \bar{Y}_b}{\sqrt{\frac{1}{2}(\sigma_a^2 + \sigma_b^2)}} = d$$

where $\bar{Y}_x$ and σ_x^2 are the mean and variance of the test score distribution in group x in the test metric in which the distribution is normal.

Note that d will be affected by a nonlinear monotonic transformation of the test score scale, but V will not. Thus V is a scale-invariant quasi-effect size measure (Ho and Reardon 2012).

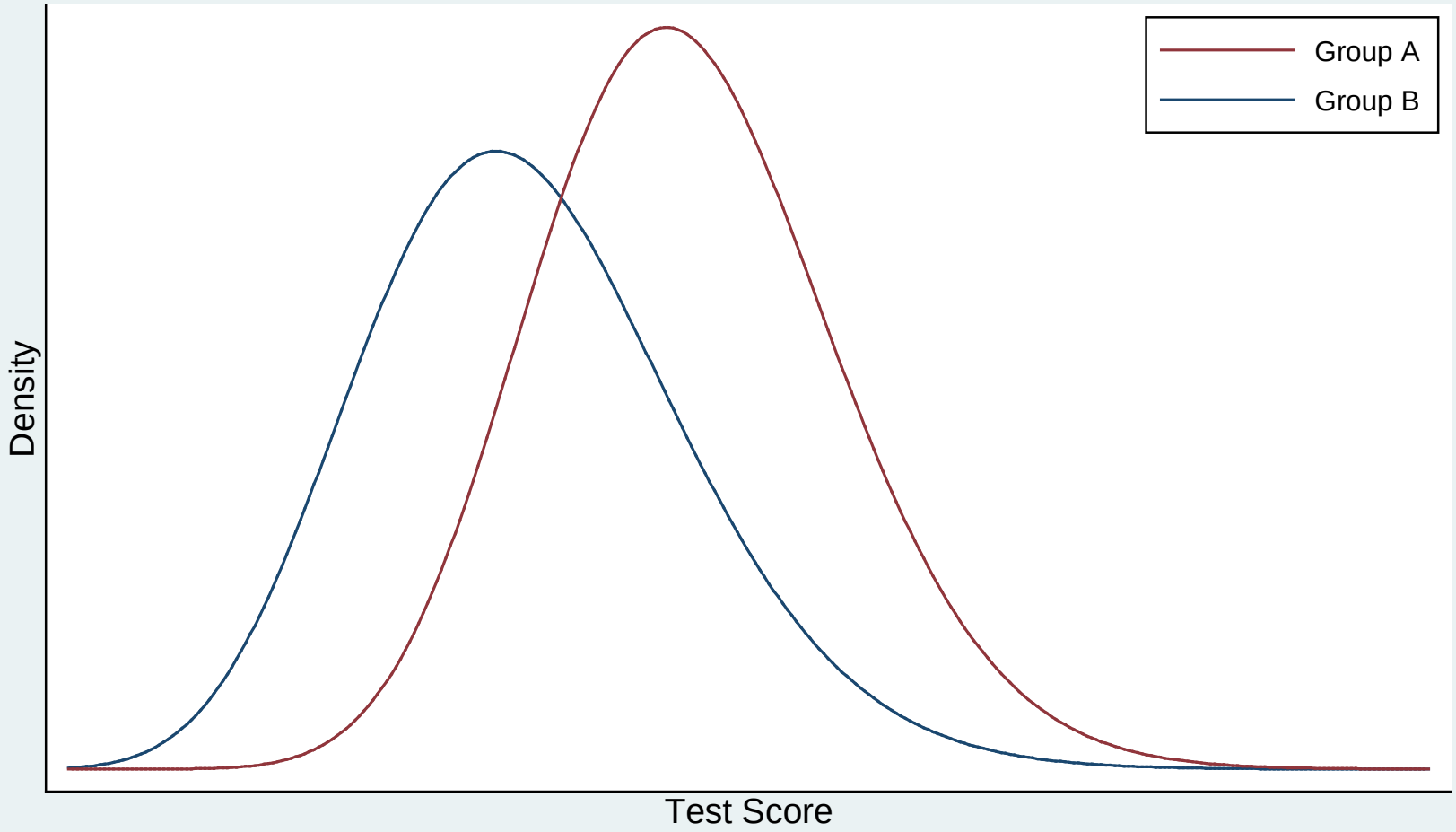
V has three very useful properties for our purposes:

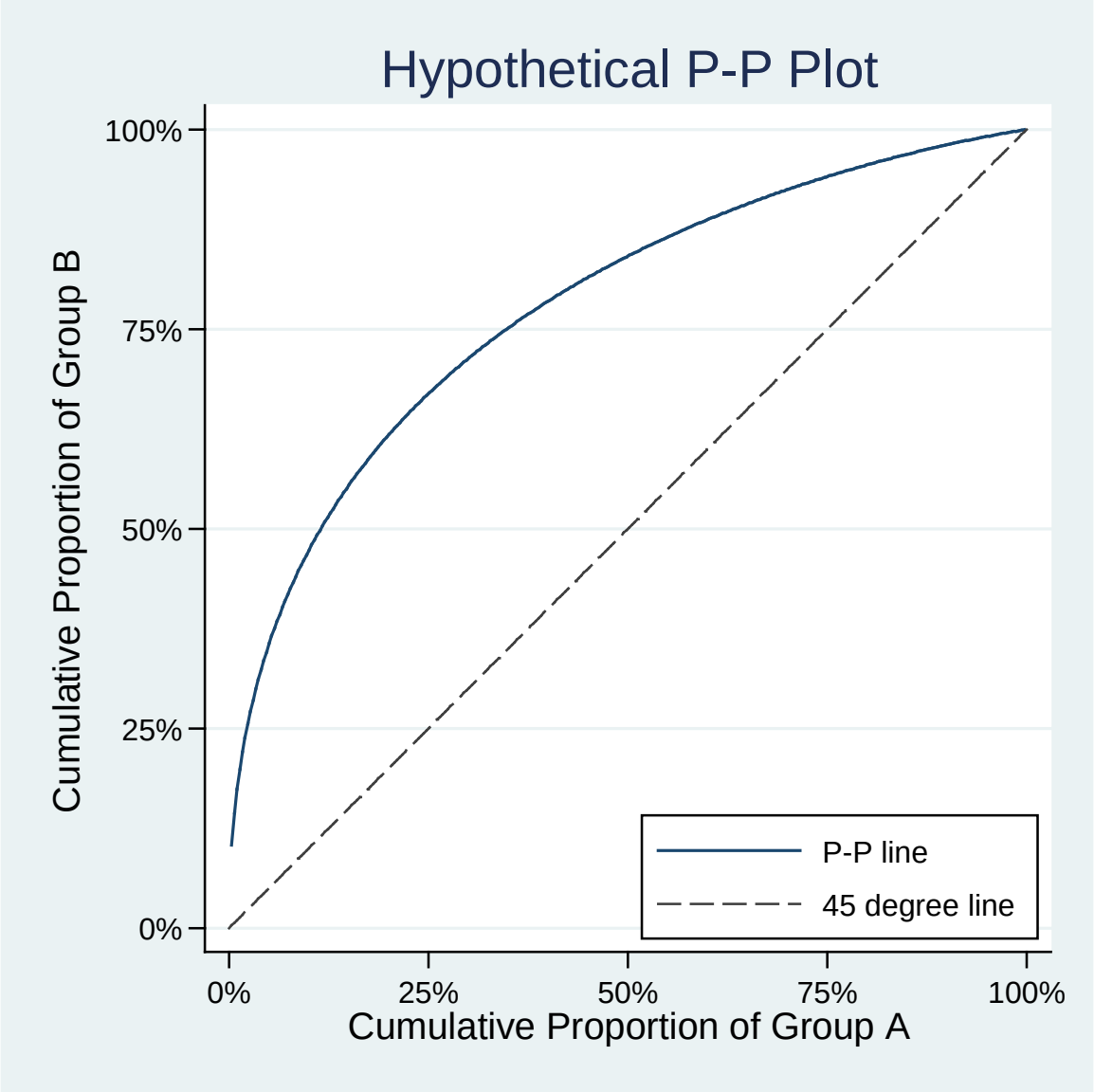
- V is interpretable as a standardized quasi-effect size (like Cohen's d). Indeed, if the two test score distributions are *respectively normal* (meaning there is some monotonic transformation of the test score metric that will render both distributions normal), then V will equal the standardized effect size (Cohen's d).
- V is insensitive to the test metric in which scores are reported. It depends only on the extent of overlap between the distributions. Thus, it allows us to compare gaps on tests that do not have the same test metric. As long as the two tests would yield the same amount of overlap between the two distributions, V will be the same.
- V can be readily estimated from proficiency count data (like the state test data typically reported under NCLB) (Ho and Reardon 2012).
  - When the (unobserved) underlying test score distributions are normal (or could be transformed to be normal), unbiased estimation of V is possible.
  - Even when the underlying distributions are not normal (and cannot be transformed into normal distributions), estimates of V have very little bias (bias is typically < 0.01 standard deviations).


Hypothetical Test Score Distributions
in Metric 1
Group A
Group B
Density
Test Score

Hypothetical Test Score Distributions

in Metric 2




Hypothetical P-P Plot
100%
75%
50%
25%
0%
Cumulative Proportion of Group B
0%
25%
50%
75%
100%
Cumulative Proportion of Group A
P-P line
45 degree line