“Going down with the ship? The effect of school accountability on the distribution of teacher experience in California.”

David P. Sims
Economics Department,
Brigham Young University.

NCLB Emerging Findings Research Conference

August 2009

NATIONAL CENTER ON
Performance Incentives

CALDER

The issue: How does high-stakes failure affect schools?

Logistically:

- Administrative Punishments
- Stigma (Ladd and Glennie 2001)
- School Choice (Greene 2001, Figlio and Rouse 2006).
- More resources – and more requirements (SES)

Most systems provide escalating punishment for repeat (or serial) failures

How does failure translate to change for?

- Students
- Teachers
- Administration

Can we separate failure effects from threat effects?

This study: What effect does school failure have on teacher experience?

Underlying idea – failure to make AYP might mean:

- Less pleasant working conditions
- More work
- Morale/stigma issues
- New idea of responsibility

Response of teachers?

Bigger implications?

- Teacher experience linked to student outcomes (Hanushek 1997, etc).

Subgroup rules

All “numerically” significant subgroups must meet the standard.

50 students and 15% of school population
or 100 students

Race/socioeconomics disadvantaged

Additional subgroups may raise failure probabilities (Kane and Staiger 2003).

A Standard Empirical Approach

Suppose we want to know the relationship between failure and subsequent teacher experience:

Mathematically:

$$E_{it} = \varphi F_{it-1} + X'_{it}\gamma + \mu_{it}$$

i = school

t = year

E = teacher experience measure

F failure indicator

X other school characteristics

Intuitively

if more experienced teachers leave $\varphi<0$

If more resources available to failing schools perhaps $\varphi>0$

Is that convincing?

Schools that fail might differ from other schools in ways that could drive down teacher experience (e.g. parental support)

Schools that fail one year are more likely to have had bad luck and thus improve the next year.

In some sense what we wish to measure is randomly assigned failure – of course that is not possible – but we can actually get close

Subgroup cutoffs imply a natural experiment

Schools with similar test scores or other characteristics have different probabilities of failure built in – not due to their own actions

e.g. Susan Anthony and Caroline Wenzel elementary schools are a few miles apart in suburbia south of Sacramento.

- Anthony had 49 Hispanics so no subgroup and they met AYP.
- Wenzel had 52 Hispanics so an extra subgroup and they failed to meet AYP.

Another example – L.A.

Ford Elementary overall state API score 632 with 52 African-American students – fails.

Abbott Elementary overall state API score 635 with 48 African-American students passes.

Thus subgroup rules create a sort or randomized failure

I compare the teacher experience response in a set of similar schools of those schools with a particular extra subgroup to those without

Identification Strategy –Instrumental Variables

The formal analog to this intuitive idea is to add another equation to our posited relationship:

$$F_{it-1} = \delta Z_{t-1} + X'_{it}\theta + \omega_{it-1}$$

This says that Subgroup status is related to failure probability.

If true we can use subgroup status as an excluded instrument for failure to make AYP.

Assume subgroups only affect experience through failure given achievement levels and smooth race controls

CA Data:

- AYP data (participation, subgroup status, meeting standards or not).
- School level characteristics (racial composition, meal eligibility, ell status, IEP status, etc).
- Data on every teacher compiled in the fall of each year (PAIF) includes experience, credential, education. I aggregate this to the school level.
- I also restrict data to comparable school windows.

Data matching:

Timing of hiring versus accountability announcements

- Many staffing decisions made before test results announced

So I match failure status to teacher experience 1.5 years later

- e.g. Spring 2003 (failure status)

- Fall 2004 (teacher experience)

AYP – failure years (03,04,05)

Teacher years – (04,05,06)

Descriptive Statistics: Teacher information

Sample:	All	10-20% Hispanic	10-20% Black
Teacher Experience	12.983 (3.453)	13.964 (3.528)	12.558 (3.306)
District Experience	10.782 (3.199)	11.408 (3.419)	10.616 (3.201)
Fraction Novice	0.162 (0.118)	0.142 (0.111)	0.169 (0.113)
Fraction emergency credential	0.028 (0.053)	0.021 (0.045)	0.032 (0.052)
N=	21009	3660	3011

Descriptive Statistics: School information

Sample:	All	10-20% Hispanic	10-20% Black
School failed at time (t-1)	0.363	0.153	0.410
	(0.481)	(0.360)	(0.492)
Fraction Hispanic Students	0.433	0.147	0.444
	(0.291)	(0.032)	(0.212)
Fraction Black Students	0.078	0.080	0.142
	(0.114)	(0.134)	(0.032)
Fraction Disadvantaged Students	0.530	0.266	0.613
	(0.313)	(0.209)	(0.259)
Fraction Free Lunch	0.503	0.256	0.584
	(0.308)	(0.207)	(0.260)
Fraction English Learner	0.254	0.088	0.258
	(0.219)	(0.089)	(0.181)
Mobility	0.175	0.169	0.207
	(0.113)	(0.122)	(0.118)
Significant subgroups	2.637	2.368	3.177
	(0.940)	(1.111)	(0.969)
Hispanic Subgroup Present	0.759	0.384	0.894
	(0.428)	(0.486)	(0.308)
Black Subgroup Present	0.164	0.138	0.396
	(0.370)	(0.345)	(0.489)
n	21009	3660	3011

The argument – what I will show

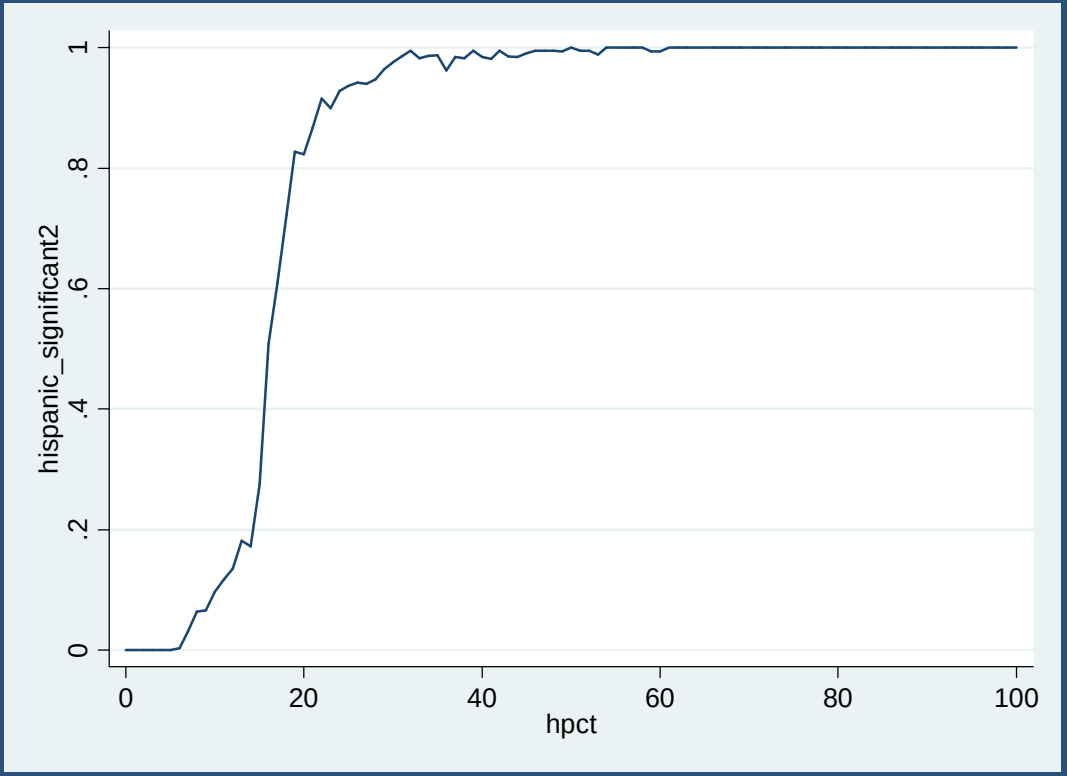
1. For a given Subgroup, that status phases in in a narrow range (of student body composition)

    I restrict my sample to that range

2. Subgroup status in that range leads to higher failure probability

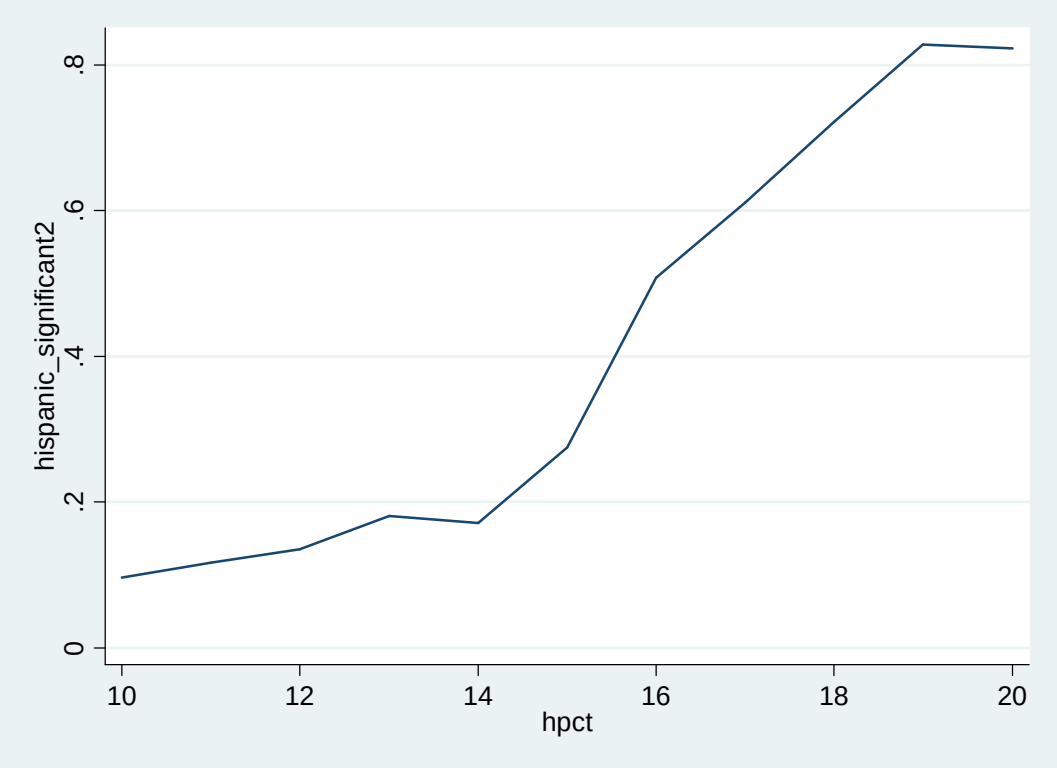
3. Subgroups status in that range also leads to lower teacher experience

1. Where subgroups phase-in

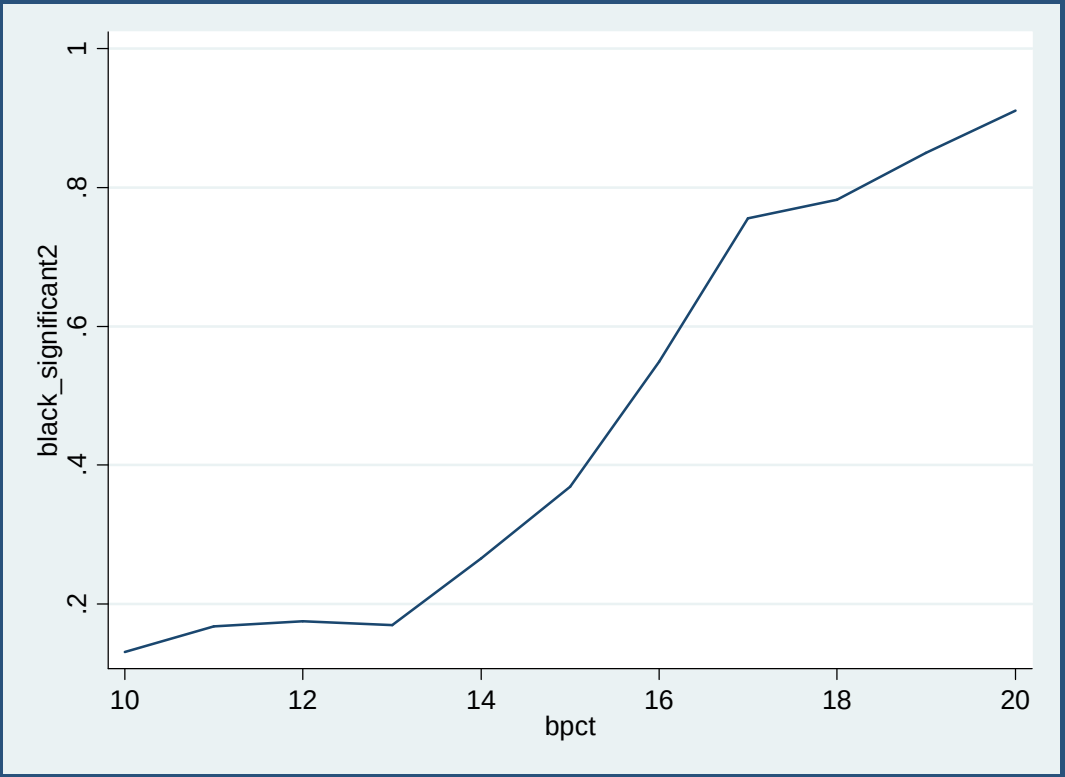


Relationship between Hispanic student percentage and Hispanic subgroup status

A smaller range captures most of the variation

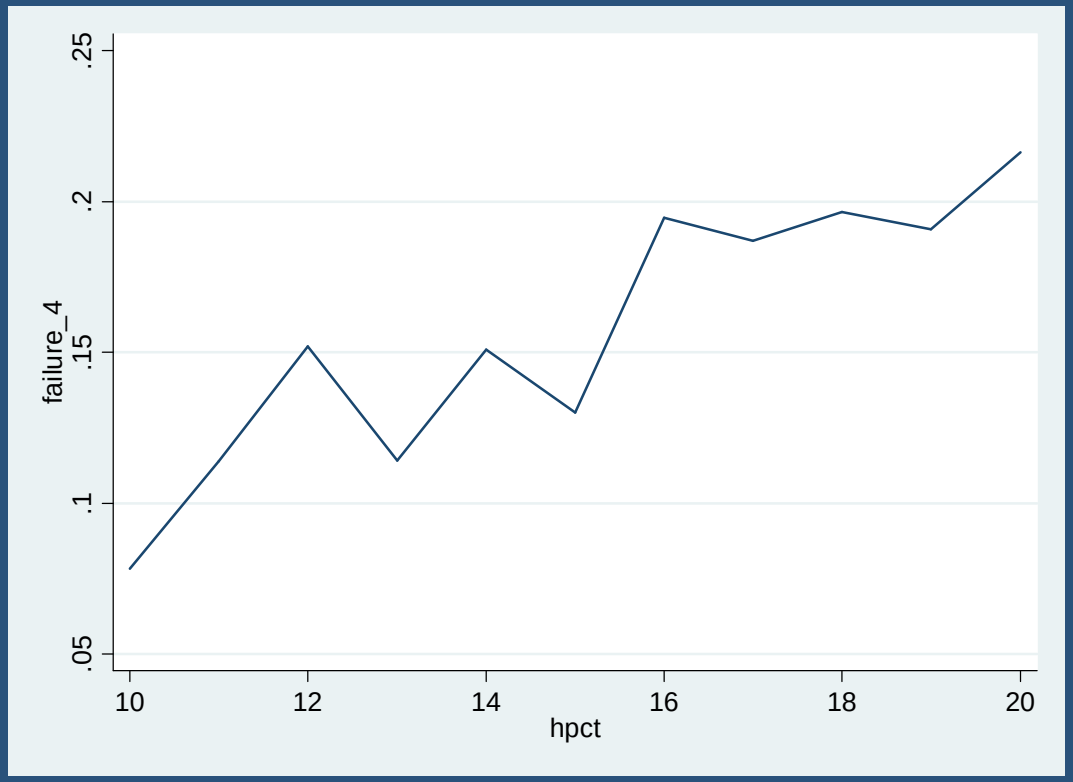


Relationship between Hispanic student percentage and Hispanic subgroup status – smaller range

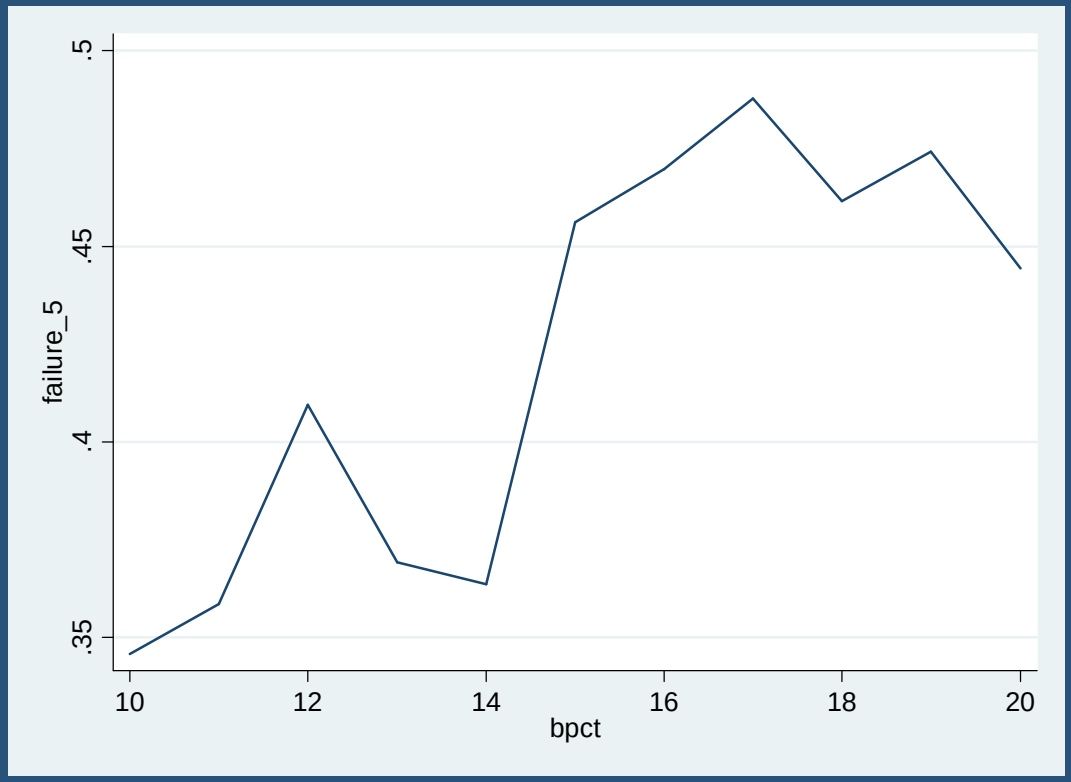


Relationship between Black student percentage and Black subgroup status

2. Failure and subgroups



Relationship between Hispanic student percentage and failing NCLB standard



Relationship between Black student percentage and failing NCLB standard

Failure and subgroups mathematically

A. Common subgroups	
Hispanic= subgroup	0.090*
	(0.016)
Black = subgroup	0.139*
	(0.022)
F-stat	39.58
B. Subgroup Cutoffs (at 15%)	
Hispanic cutoff	0.038*
	(0.018)
Black cutoff	0.051*
	(0.025)
F-stat	3.94
N=	21,009

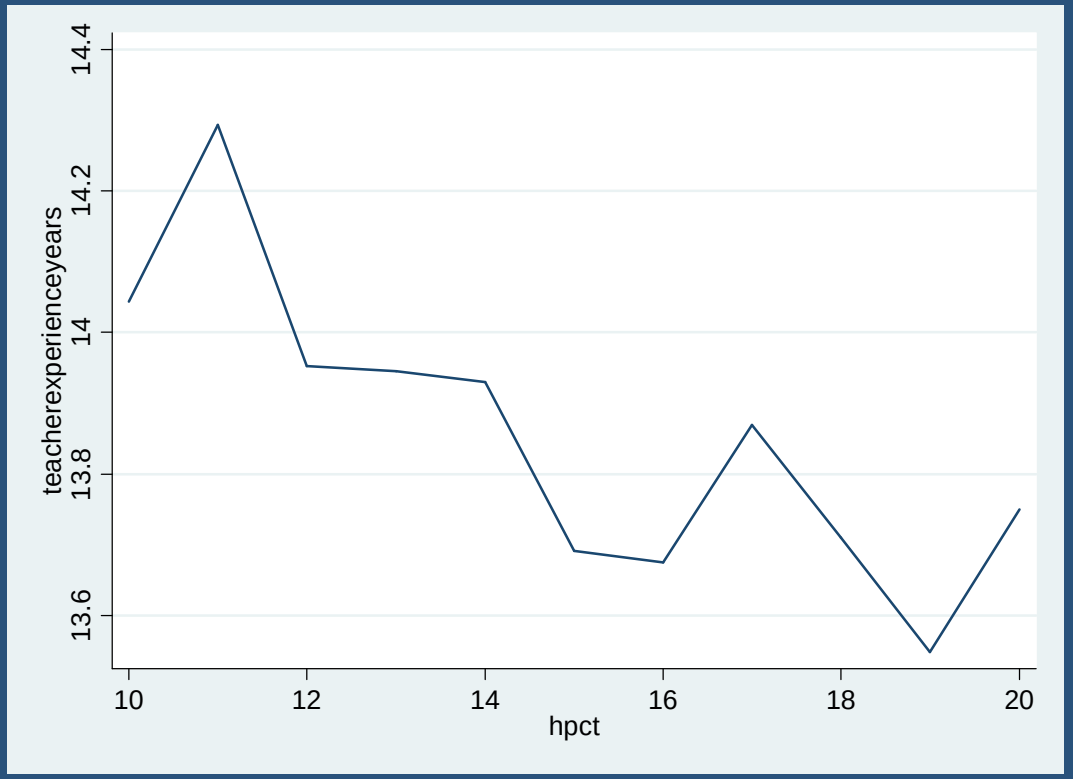
An unequal starting line for evaluation

A school with a Hispanic subgroup, all else equal, has a 9 percent greater chance of failing

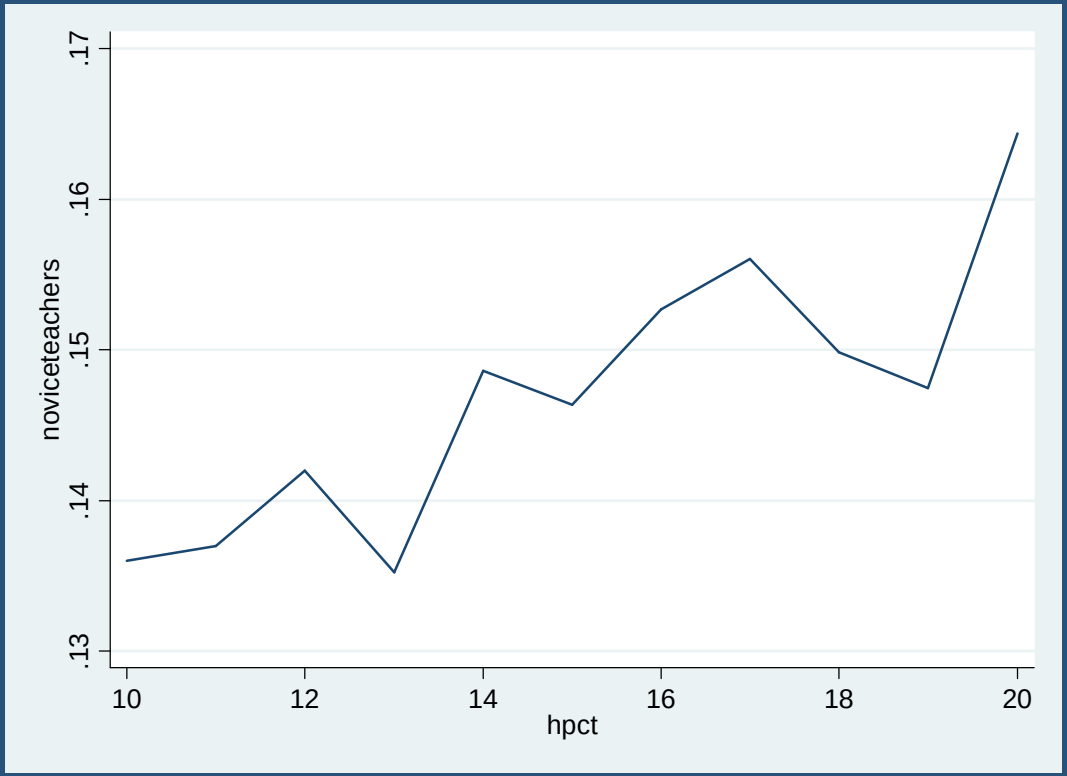
A school with a Black subgroup, all else equal, has a 14 percent greater chance of failing

These results take into account other possible school differences

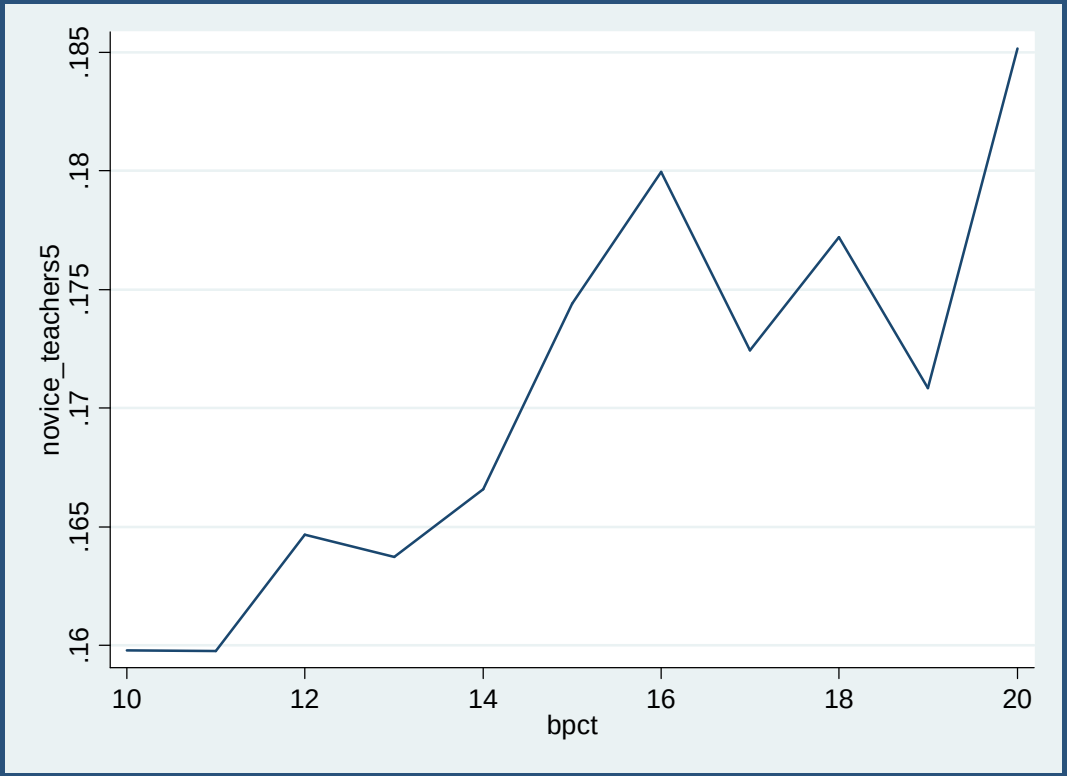
3. Effect of subgroups on teacher experience



Relationship between Hispanic student percentage and average teacher experience



Relationship between Hispanic student percentage and novice teachers



Relationship between Black student percentage and novice teachers

Putting it together – failure and teacher experience

Least Squares estimates of effect of AYP failure on school level teacher characteristics

Teacher Characteristic:	Teacher Experience	District Experience	Fraction Novice	Fraction emergency credential
	(1)	(2)	(3)	(4)
A. Single failure lag				
School failed at time (t-1)	-0.384* (0.064)	-0.551* (0.060)	0.027* (0.002)	0.017* (0.001)
B. Multiple failure lags				
School failed at time (t-1)	-0.303* (0.069)	-0.356* (0.063)	0.018* (0.003)	0.011* (0.002)
School failed at time (t-2)	-0.380* (0.068)	-0.534* (0.063)	0.024* (0.002)	0.016* (0.002)

Hispanic 10%-20% sample

Teacher Characteristic:	Teacher Experience	District Experience	Fraction Novice	Fraction Emergency credential
Method	(1)	(2)	(3)	(4)
A. OLS on this sample	-0.359* (0.143)	-0.833* (0.137)	0.027* (0.005)	0.017* (0.002)
B. **Instrumental Variables - does school have a Hispanic subgroup?**	**-2.749* (1.254)**	**-2.664* (1.203)**	**0.104* (0.039)**	**0.056* (0.016)**
C. + demographic controls	-2.629* (1.261)	-3.111* (1.228)	0.116* (0.040)	0.065* (0.017)
D. +higher order race terms	-3.395* (1.607)	-3.557* (1.552)	0.099* (0.049)	0.074* (0.021)
E. Instrumental variables – Indicator for >15%	-2.387 (1.859)	-1.576 (1.769)	0.086 (0.058)	0.024 (0.023)

Other samples

Teacher Characteristic:		Teacher Experience	District Experience	Fraction Novice	Fraction emergency credential
Sample	Instrument	(1)	(2)	(3)	(4)
Black 10-20% sample (n=3011)	**Black subgroup**	**-1.846* (0.677)**	**-2.783* (0.671)**	**0.109* (0.024)**	**0.084* (0.012)**
	Black students >15%	-4.917* (1.383)	-6.120* (1.470)	0.141* (0.044)	0.039* (0.019)
Schools with 40-60 Hispanic students (n=1587)	Hispanic subgroup	-3.903 (2.520)	-2.312 (2.331)	0.332* (0.101)	0.070* (0.035)
Schools with 40-60 black students (n=1848)	Black subgroup	-2.154 (2.036)	-3.717 (2.103)	0.041 (0.066)	0.051* (0.025)

Failure leads to lower experience

An average school of 30 teachers loses about 60 years of aggregate experience

Gains 2-3 new novice teachers

Gains about 1 emergency credentialed teacher

Issues

Size of IV Estimates

- Effect on a particular group
- Maybe failure is not the right mechanism

Who chooses experience decrease

- Teacher choice
- School choice

Effects on schools

- Small magnitudes relative to other issues

Effects of one additional subgroup on teacher experience

Teacher Characteristic:	Teacher Experience	District Experience	Fraction Novice	Fraction emergency credential
Sample	(1)	(2)	(3)	(4)
A. Hispanic 10-20% sample (n=3660)	**-0.270*** **(0.120)**	**-0.262*** **(0.116)**	**0.010*** **(0.004)**	**0.005*** **(0.002)**
A. Black 10-20% sample (n=3011)	-0.331* (0.123)	-0.500* (0.119)	0.019* (0.004)	0.015* (0.002)
C. Schools with 40-60 Hispanic students. (n=1587)	-0.316 (0.202)	-0.186 (.187)	0.027* (0.006)	0.006* (0.002)

Policy Conclusions:

Subgroup rules lead to comparable schools being evaluated by different standards.

Schools that are thus given more stringent requirements have a relative loss of teacher experience (at least in the short run).

This may reflect teacher reaction to failure more broadly.

Implementing a mandate for a qualified teacher in every classroom as part of accountability requires careful thinking as other plan elements may give contrary incentives to well-qualified teachers.