Leveraging Lotteries for School Value-Added

Josh Angrist, Peter Hull, Parag Pathak, and Chris Walters

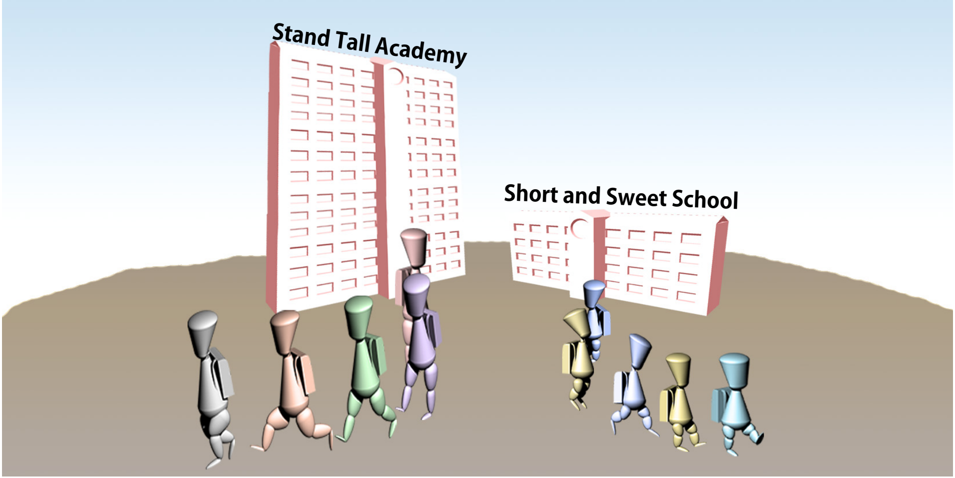
The MIT School Effectiveness and Inequality Initiative

Fall 2016

Accountability in Action

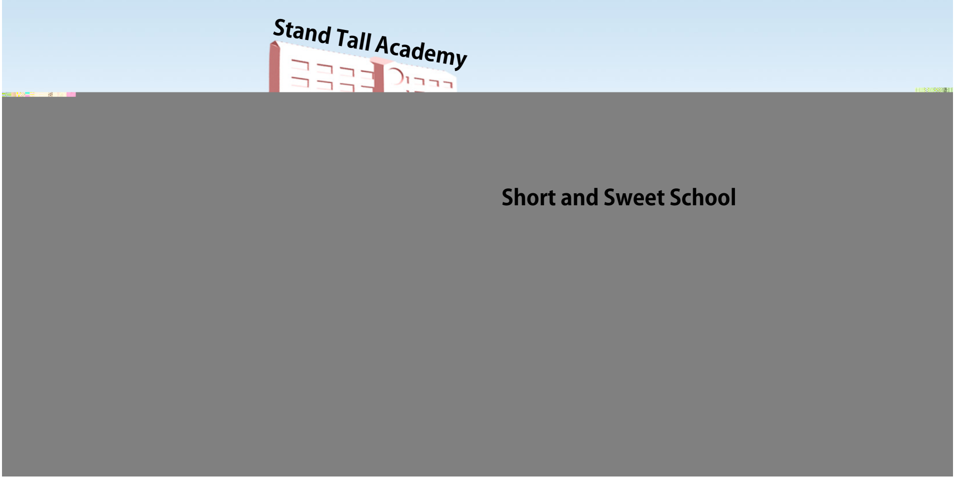
- States and districts increasingly use test scores to measure, compare, and report on school and teacher quality
  - Massachusetts "school report cards" compare MCAS levels and growth, along with graduation rates, grouping schools into 5 levels
- *Value-added models* (VAM) are a more sophisticated version of this, comparing scores across schools or classrooms, adjusting for student characteristics and past achievement
- VAM validity is high stakes
  - SEII and others have shown that schools and teachers that boost achievement boost longer-term outcomes such as post-secondary education and earnings
  - Teachers may be dismissed or promoted, schools or school networks scaled back or expanded, on the basis of VAM

The VAM Challenge



My kids gonna go Tall!

Simple Comparisons Mislead



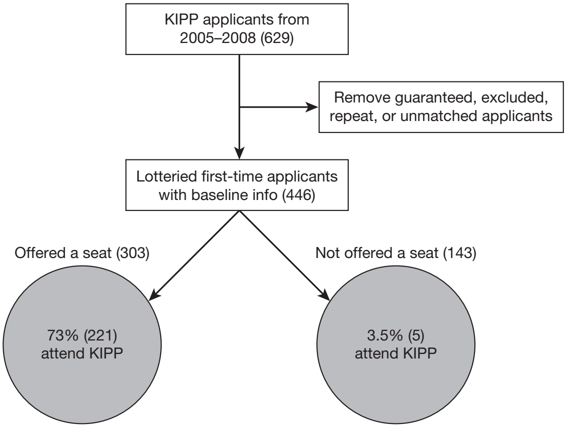
- A good (unbiased) VAM procedure compares *ex ante* similar kids

The SEII VAM Agenda

- Large urban districts like Boston use centralized assignment to offer parents a choice of schools
  - New Orleans, New York, Newark, Chicago, Washington, and Denver use something similar
- These schemes randomly assign seats when schools are over-subscribed
  - BPS combines centralized assignment for its traditional public and pilot schools with decentralized school-specific lotteries at charter schools
- A research/policy opportunity! SEII uses the lotteries embedded in the Boston assignment scheme to:
  - *Test the validity of VAM estimates*
    - we ask whether conventional regression-adjusted VAM predicts the effect of *random assignment* to schools
  - *Reduce bias*
    - we combine lottery info with regression-adjusted VAM to produce a better "hybrid VAM" mousetrap, less biased and more precise

How We Play the (School) Lottery: KIPP Lynn

FIGURE 3.1
Application and enrollment data from KIPP Lynn lotteries



Note: Numbers of Knowledge Is Power Program (KIPP) applicants are shown in parentheses.

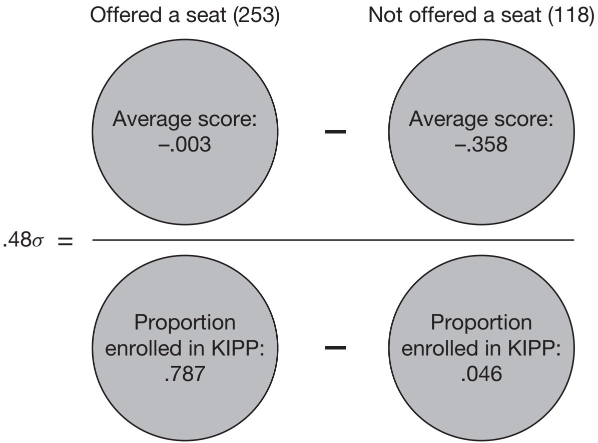
Comparing KIPP Applicants

TABLE 3.1
Analysis of KIPP lotteries

		KIPP applicants			
	Lynn public fifth graders (1)	KIPP Lynn lottery winners (2)	Winners vs. losers (3)	Attended KIPP (4)	Attended KIPP vs. others (5)
Panel A. Baseline characteristics					
Hispanic	.418	.510	−.058 (.058)	.539	.012 (.054)
Black	.173	.257	.026 (.047)	.240	−.001 (.043)
Female	.480	.494	−.008 (.059)	.495	−.009 (.055)
Free/Reduced price lunch	.770	.814	−.032 (.046)	.828	.011 (.042)
Baseline math score	−.307	−.290	.102 (.120)	−.289	.069 (.109)
Baseline verbal score	−.356	−.386	.063 (.125)	−.368	.088 (.114)

Waiting for Superman

FIGURE 3.2
IV in school: the effect of KIPP attendance on math scores



Note: The effect of Knowledge Is Power Program (KIPP) enrollment described by this figure is $.48\sigma = .355\sigma/.741$.

Boston Public

Schools: Traditional, Pilot, and Charter

Table 1: Boston students and schools

	Enrollment				Enrollment		
School	OLS sample	Lottery sample	Lottery school?	School	OLS sample	Lottery sample	Lottery school?
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Traditional publics				B. Pilots			
1	1,095	79	Y	1	538	310	Y
2	1,025	445	Y	2	1,260	433	Y
3	1,713	1,084	Y	3	585	296	Y
4	547	218	Y	4	78	5	
5	217	46		5	453	46	Y
6	1,354	581	Y	6	380	67	Y
7	263	44		7	242	179	Y
8	1,637	492	Y	8	558	73	Y
9	472	104		9	18	12	
10	1,238	591	Y	C. Charters			
11	537	11		1	738	406	Y
12	331	35	Y	2	361	23	
13	335	82		3	357	215	
14	952	232	Y	4	393	332	Y
15	294	71	Y	5	338	16	
16	333	90		6	511	115	Y
17	766	243	Y	7	71	8	
18	372	47	Y	8	300	23	
19	137	14		9	389	342	Y
20	1,091	225	Y	10	654	34	
21	1,086	127	Y	11	45	3	
22	577	104	Y	12	53	2	
23	622	61		13	415	305	Y
24	906	270	Y	14	70	6	
25 *(Ref.)*	267	19		15	104	23	
				16	701	92	
All schools:	27,864	8,718	28	17	85	37	

Notes: This table counts the students and schools included in the observational (OLS) and lottery samples. The sample covers cohorts attending 6th grade in Boston between the 2006-2007 and 2013-2014 school years. Traditional public school #25 is the designated omitted enrollment category for value-added estimation. Columns (4) and (8) indicate whether the school has enough students subject to

Sample, Data, and Models

- 28,000 6th graders attending 51 traditional, pilot, and charter schools in Fall 2006-13
  - Schools with fewer than 25 6th graders are excluded
  - 8,700 applicants to schools with at least 50 students subject to random assignment make up the lottery sample
  - Baseline: 5th grade math, ELA (standardized by subject/grade/year)
  - Outcome: similarly standardized scores from the end of 6th
- We compare four conventional VAM models
  - Uncontrolled (except for calendar year)
  - Demographic
    - Controls for sex, race, FRPL, SPED, ELL, and for baseline absences and suspensions
  - Lagged Score
    - Adds baseline math and ELA scores to the demographic model
  - Gains
    - Replaces score levels with grade-to-grade score changes in the demographic model

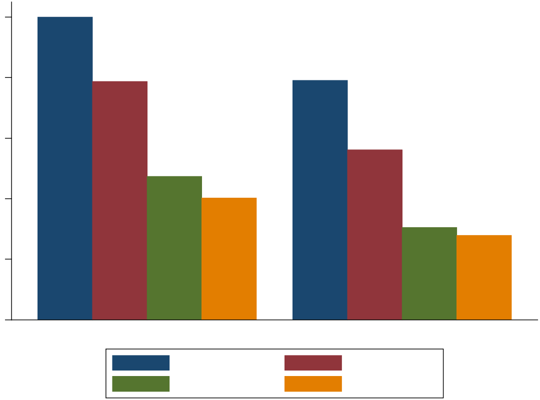
Descriptive Stats

Table 2: Descriptive statistics

	Means		Offer instrument balance			
	OLS sample	Lottery sample	All lotteries	Traditional	Pilot	Charter
Baseline covariate	(1)	(2)	(3)	(4)	(5)	(6)
Hispanic	0.345	0.354	-0.017	-0.007	0.003	-0.006
			(0.013)	(0.017)	(0.033)	(0.018)
Black	0.410	0.485	-0.011	-0.005	-0.052	-0.009
			(0.014)	(0.018)	(0.034)	(0.020)
White	0.122	0.072	0.010	0.006	0.005	0.009
			(0.007)	(0.008)	(0.015)	(0.010)
Female	0.490	0.504	0.017	0.034*	-0.013	-0.025
			(0.014)	(0.019)	(0.037)	(0.020)
Subsidized lunch	0.806	0.830	0.020*	0.020	0.006	-0.005
			(0.010)	(0.013)	(0.026)	(0.016)
Special education	0.208	0.195	0.006	-0.003	-0.022	0.015
			(0.011)	(0.013)	(0.030)	(0.016)
English-language learner	0.205	0.206	0.006	-0.001	0.018	0.004
			(0.011)	(0.014)	(0.027)	(0.016)
Suspensions	0.093	0.076	-0.025	-0.025	0.009	-0.016
			(0.016)	(0.023)	(0.025)	(0.017)
Absences	1.710	1.534	-0.087	-0.138*	-0.092	0.110
			(0.095)	(0.080)	(0.260)	(0.167)
Math score	0.058	0.004	0.022	-0.026	0.080	0.036
			(0.024)	(0.030)	(0.061)	(0.035)
ELA score	0.030	0.013	0.035	0.045	0.060	0.013
			(0.025)	(0.030)	(0.061)	(0.036)
N	27,864	8,718	8,718	4,849	1,303	3,655

Notes: This table reports sample mean characteristics and investigates balance of random lottery offers. Column (1) shows mean characteristics for all Boston 6th graders enrolled between the 2006-2007 and 2013-2014 school years, and column (2) shows mean characteristics for randomized

Conventional VAM Estimates Using Alternative Controls

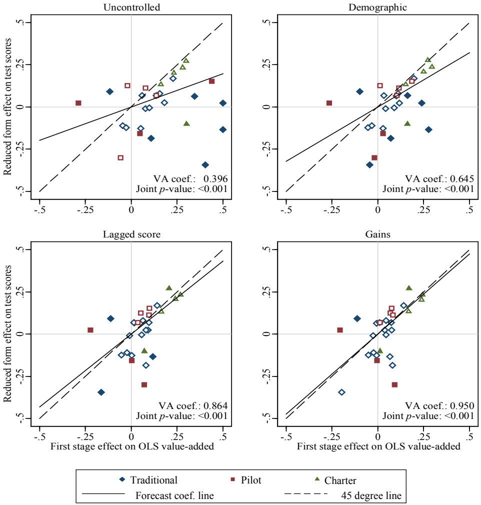


Testing VAM Validity

- Consider a district with three middle schools, A, B, and C; suppose that students pick one of these freely
  - Conventional VAM compares the difference in average 6th grade scores across these schools, after regression-adjusting for characteristics like subsidized lunch and 5th grade scores
  - VAM estimates in this case are adjusted contrasts between A and C and between B and C. Call these values VAM_A and VAM_B (to compare schools A and B, look at $VAM_A - VAM_B$)
- Suppose now that a lottery randomly matches students to schools
  - This generates apples-to-apples comparisons of students who differ only by virtue of school assigned
- Validation
  - Compare those randomly assigned to A and to C: is the achievement gap here equal to VAM_A, or only some fraction of this?
  - Compare those randomly assigned to B and to C: is the achievement gap here equal to VAM_B, or only some fraction of this?

Visual VAM Validation

Figure 2: Visual instrumental variables tests for bias



Notes: This figure plots lottery reduced form effects against value-added first stages from each of the 28 school lotteries. See the notes for Table 3 for a description of the value-added models and lottery specification. Filled markers indicate estimates that are significant at the 10% level. Slopes of solid lines correspond to the forecast coefficients from Table 3, while dashed lines indicate the 45-degree line.

Formal Tests

Table 3: Tests for bias in conventional value-added models

	Uncontrolled (1)	Demographic (2)	Lagged score (3)	Gains (4)	Lagged score, no charter lotteries (5)
Forecast coefficient	0.396	0.645	0.864	0.950	0.549
	(0.056)	(0.065)	(0.075)	(0.084)	(0.164)
First stage F-statistic	45.6	36.1	29.6	26.6	11.2
p-values:					
Forecast coef. equals 1	<0.001	<0.001	0.071	0.554	0.006
Overid. restrictions	<0.001	<0.001	0.003	0.006	0.043
All restrictions	<0.001	<0.001	<0.001	<0.001	0.002
All restrictions (bootstrap refinement)	<0.001	<0.001	<0.001	<0.001	0.002

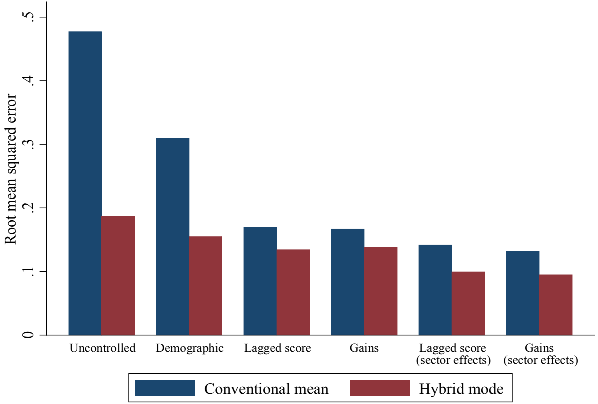
Notes: This table reports estimates of the VAM forecast coefficient and the results of tests for bias in conventional value-added models for 6th grade math scores. Estimated forecast coefficients are from regressions of 6th grade scores on fitted values from conventional value-added models, instrumented by the set of offer dummies for all school lotteries. Models are estimated via a two-step optimal GMM procedure that is efficient with arbitrary heteroskedasticity. Joint p-values come from OLS regressions of value-added residuals on offer dummies. The uncontrolled model includes only year-of-test indicators as controls. The demographic model adds indicators for student sex, race, subsidized lunch, special education, limited-English proficiency, and counts of baseline absences and suspensions. The lagged score model adds cubic polynomials in baseline math and ELA scores. The gains model includes the same controls as the demographic model and uses score gains from baseline as the outcome. Column (5) excludes charter school lotteries from the lottery sample in testing the lagged score model. All IV models control for lottery strata fixed effects, demographic variables, and lagged scores. Standard errors are reported in parentheses. Bootstrap p-values are based on 500 Bayesian bootstrap replications (see Appendix B for details).

Hybrid VAM

- We'd like to use lotteries to compare *all* schools in the district
  - In practice, some schools are under-subscribed; that is, everyone who applies gets a seat (good for them, of course!)
  - Lottery comparisons involve fewer students than the comparisons of all enrolled students that lie behind conventional VAM
- Can we have our unbiased lottery VAM cake and eat it with the statistical precision of conventional VAM too?
  - Our hybrid VAM strategy compares lottery and conventional VAM estimates to determine the extent of bias in the latter
  - With bias known and accounted for, we adjust relatively precise conventional estimates to build a better VAM mousetrap
- Estimating a random coefficients (empirical Bayes model), we use gaps between lottery and conventional VAM estimates to pin down the hyperparameters governing the distribution of bias
  - This allows us to bias-correct VAM for all schools

Help from the Hybrid

Figure 5: Root mean squared error for value-added posterior predictions



Notes: This figure plots root mean squared error for posterior predictions of school value-added. Conventional predictions are posterior means constructed from OLS value-added estimates. Hybrid predictions are posterior modes constructed from OLS and lottery estimates. Root mean squared error is calculated from 100 simulated samples drawn from the data generating processes implied by the estimates in Table 5. The random coefficients model is re-estimated in each simulated sample.

Sim City

VAM-Based Policy Predictions

Table 7: Consequences of closing the lowest-ranked district school for affected children

		Replacement school:			
Value-added model	Posterior method	Average school (1)	Average above-median school (2)	Average top-quintile school (3)	Average charter school (4)
-	True value-added	0.357	0.488	0.570	0.666
Uncontrolled	Conventional	0.036	0.057	0.064	0.345
	Hybrid	0.163	0.242	0.294	0.472
Demographic	Conventional	0.106	0.154	0.182	0.415
	Hybrid	0.188	0.266	0.321	0.496
Lagged score	Conventional	0.206	0.281	0.332	0.515
	Hybrid	0.266	0.363	0.434	0.575
Gains	Conventional	0.237	0.326	0.383	0.557
	Hybrid	0.290	0.394	0.467	0.610

Notes: This table reports simulated test score impacts of closing the lowest-ranked district school based on value-added predictions. The reported impacts are effects on test scores for students at the closed school. Column (1) replaces the lowest-ranked district school with an average district school. Columns (2), (3) and (4) replace the lowest-ranked school with an average above-median district school, an average top-quintile district school, or the highest-ranked district school. Column (5) replaces the lowest-ranked district school with an average charter school. See notes to Table 3 for a description of the controls included in each value-added model. Conventional empirical Bayes posteriors are means conditional on OLS estimates only, while hybrid posteriors are modes conditional on OLS and lottery estimates. All models include sector effects. Statistics are based on 100 simulated samples, and the random coefficients model is re-estimated in each sample.

Mistakes notwithstanding, VAM can support productive policy decisions

Lessons Learned

- The verdict on conventional VAM:
  - Biased, sometimes quite a bit
  - Can be improved markedly by lotteries
  - The worse the conventional VAM model, the larger the lotto gains
- Bias notwithstanding, conventional VAM is far from worthless
  - Except for the simplest unadjusted comparisons, school policy decisions using conventional VAM can yield large gains
  - Lotteries provide a low-cost way to make these gains even larger