

Un-spinning the Numbers on Charter Schools, Demographics and Performance in New Jersey

Bruce D. Baker

Department of Educational Theory, Policy and Administration
Graduate School of Education
Rutgers, The State University of New Jersey

Bruce.baker@gse.rutgers.edu

Recent New Jersey headlines once again touted the supposed successes of New Jersey Charter Schools.¹

The Star Ledger reporters, among others, were essentially reiterating the information provided them by the New Jersey Department of Education. Here's their story.

And here's a choice quote from the press release:

"These charter schools are living proof that a firm dedication to students and a commitment to best education practices will result in high student achievement in some of New Jersey's lowest-income areas," said Carlos Perez, chief executive officer of the New Jersey Charter School Association. He pointed to NJASK data for third grade Language Arts, where more than half the charters outperformed the schools in their home districts, and of those, more than 75 percent were located in former Abbott districts."

No spin there. Right? Just a balanced summary of achievement data, with thoughtful interpretation of what they might actually mean. Not really.

There are many, many reasons why the comparisons released yesterday are deeply problematic, and well, quite honestly, pretty darn meaningless. I could not have said it better than Matt DiCarlo of Shanker Blog did here:

"Unfortunately, however, the analysis could barely pass muster if submitted by a student in one of the state's high school math classes (charter or regular public)."²

¹ http://www.nj.com/news/index.ssf/2011/01/gov_christie_releases_study_sh.html

Here are some guidelines I have posted in the past, regarding appropriate ways to compare New Jersey Charter Schools to their host districts on various measures including outcome measures:

1. *When comparing across schools within poor urban setting, compare on basis of free lunch, not free or reduced, so as to pick up variation across schools. Reduced lunch income threshold too high to pick up variation.*
2. *When comparing free lunch rates across schools either a) compare against individual schools and nearest schools, OR compare against district averages by GRADE LEVEL. Subsidized lunch rates decline in higher grade levels (for many reasons, to be discussed later). Most charter schools serve elementary and/or middle grades. As such they should be compared to traditional public schools of the same grade level. High school students bring district averages down.*
3. *When comparing test score outcomes using NJ report card data, be sure to compare General Test Takers, not Total Test Takes. Total Test Takers include scores/pass rates for children with disabilities. But, as we have seen time and time again, in charts above, Charters tend not to serve these students. Therefore, it is best to exclude scores of these students from both the Charter Schools and Traditional Public Schools.*

The NJDOE and Star Ledger's primary violation involves #3 above, but also relates to the first two basic rules. Let's do a quick walk through, using the 2009 data, because the 2010 school level school reports data are not yet posted on the NJDOE web site. The bottom line is that it is relatively meaningless to simply compare raw scores or proficiency rates of charter schools to host district schools – as done by NJDOE and the Star Ledger. That is, it is meaningless unless they actually serve similar student populations, which they do not.

Below, I walk through a few quick examples of student population differences in Newark, home to the state's high-flying charter schools (North Star Academy³ and Robert Treat Academy⁴). Next, I construct a statistical model of school performance including New Jersey Charter schools and traditional public schools in their host district, controlling for student demographics and location. I use that model to show adjusted performance comparisons on a few of the tests, and then I use a variation of that model to test the proficiency rate difference – on average statewide – between charter schools and schools in the host district. Finally, I address one additional factor which I am unable to fully control for in the model – the fact that some New Jersey Charter Schools – high performing ones – seem to have unusually high rates of cohort attrition between grade 6 and 8, concurrent with rising test scores. I raise this point because *pushing out of*

² <http://shankerblog.org/?p=1646>

³ <http://schoolfinance101.wordpress.com/2010/12/02/truly-uncommon-in-newark/>

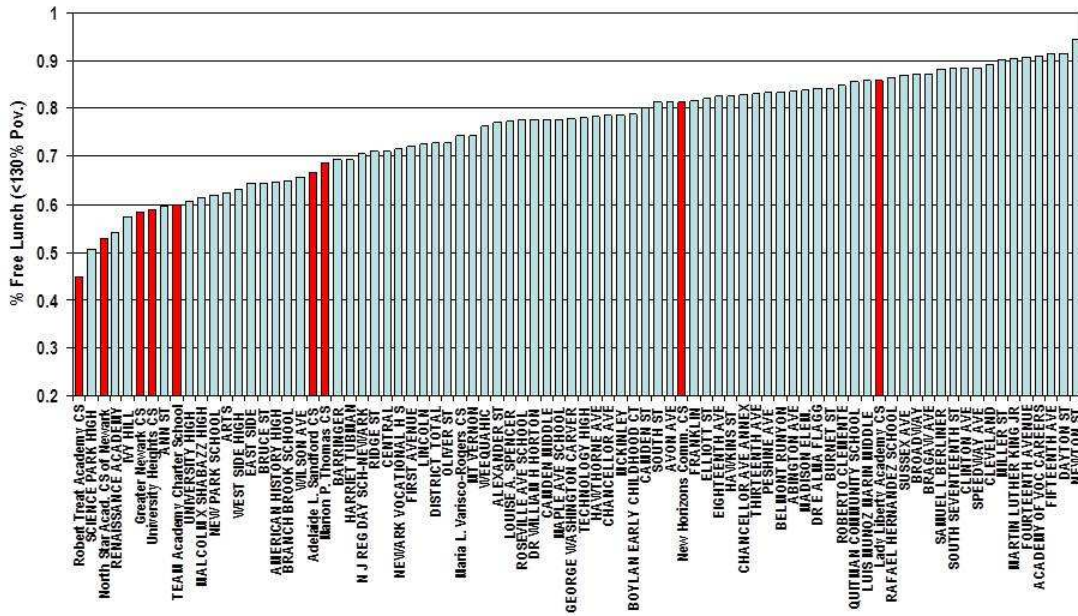
⁴ <http://schoolfinance101.wordpress.com/2009/11/05/replicating-robert-treat-academy/>

students is not an option available to traditional public schools. In fact, it is the traditional public schools that must take back those students pushed out.

Demographic Examples from Newark

Here are a few slides from previous posts on the demography of Newark Charter Schools in particular, compared to other Newark Public Schools. Here are the shares of kids who qualify for free lunch by school in Newark (city boundaries). Clearly, most of the charters fall toward the left hand side of the graph with far fewer of the lowest low-income children.

Newark Charters & NPS Schools % Free Lunch 2009-10 (<130% Pov. Threshold)

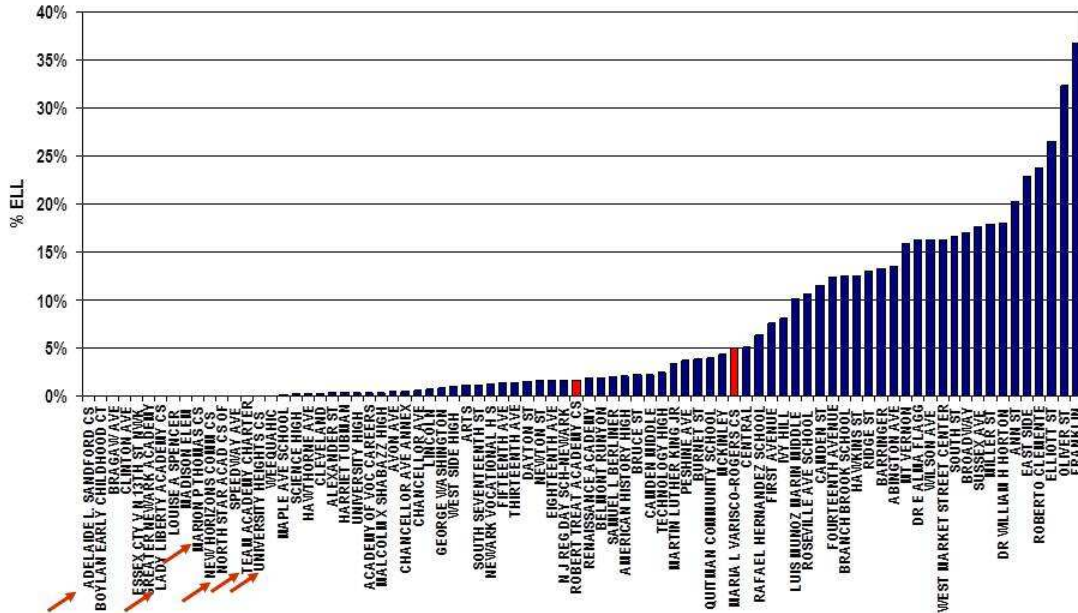


http://www.state.nj.us/education/data/enr/enr10/stat_doc.htm

Bruce D. Baker © 2010

The shares of English Language Learners look similar if not more dramatic. Many NPS schools have very high rates of English Language Learners while few charters have even a modest share.

Newark Charters & NPS Schools % ELL 2009-10



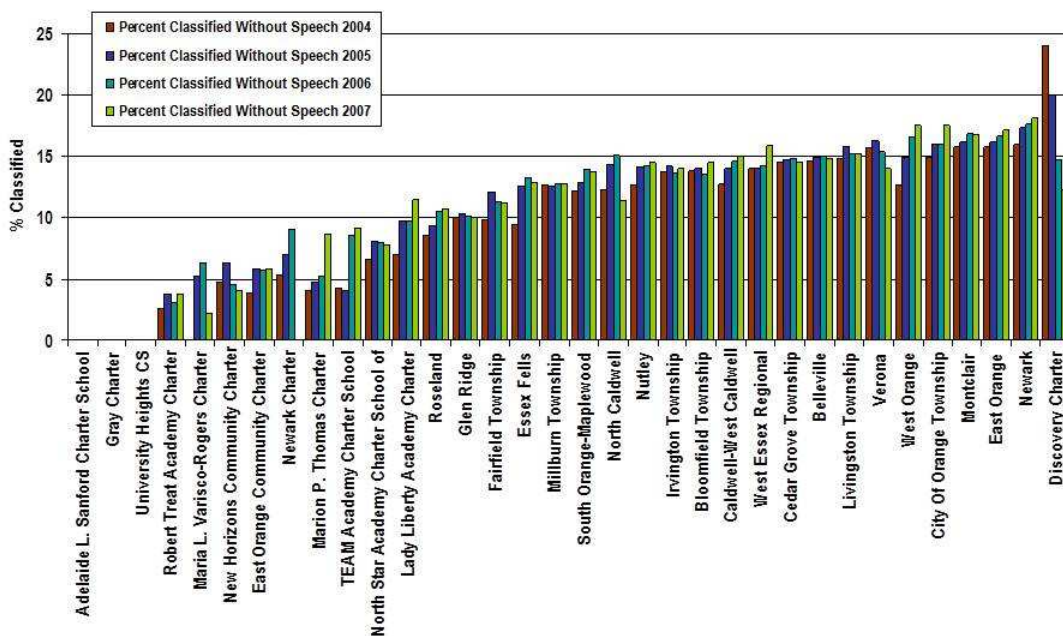
http://www.state.nj.us/education/data/enr/enr10/stat_doc.htm

Ertuce D. Baker © 2010

Finally, here’s a 4 year run of the most recent available special education classification rate data (More recent years of data have a dead link on the classification rates⁵). This graph compares Essex County charter schools with Essex County public school districts. Charter Schools have invariably low special education rates, but for those focused on children with disabilities.

⁵ <http://www.state.nj.us/education/specialed/data/2009.htm>

Newark Charters & Essex Co. Districts % Classified for Special Education 2004-2007 (excluding speech)



<http://www.state.nj.us/education/specialed/data/>

Bruce D. Baker © 2010

One cannot reasonably ignore these differences when comparing performance outcomes of kids across schools.

The Outcomes Corrected for the Demographics

So then, what happens if we actually use some statistical adjustments to evaluate whether the charter schools outperform (on average proficiency rate) other schools in the same city on the same test? Well, I've done this for charter data from 2009 and previous years and will do it again for the 2010 data when available. I use variables available in the Fall Enrollment Files⁶ and from the School Report Card⁷ and information on school location from the NCES Common Core of Data⁸ in order to create a model of the expected scores for each charter school and each other school in the same city. In the model, I use only the performance of GENERAL TEST TAKERS, so as to exclude those scores of special education students (who, for the most part don't attend the charter schools). The model:

⁶ <http://www.state.nj.us/education/data/enr/>

⁷ <http://education.state.nj.us/rc/>

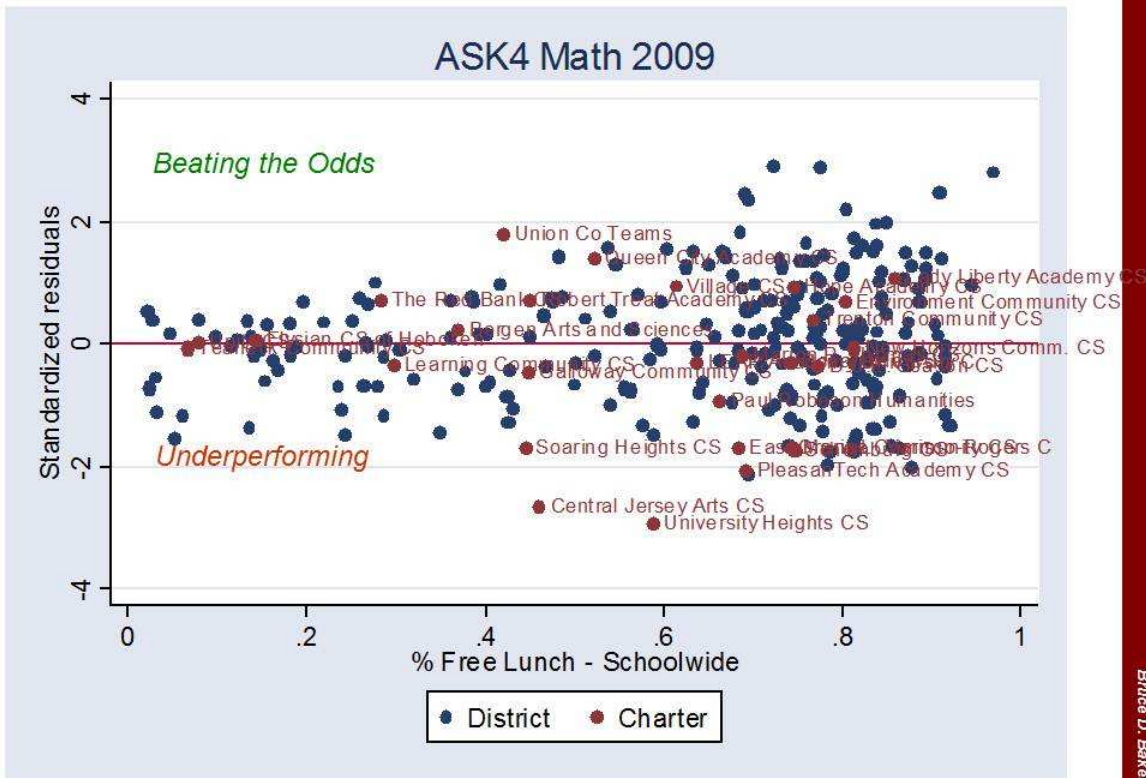
⁸ <http://nces.ed.gov/ccd/>

Outcome = $f(\text{Poverty, Race, Homelessness, City, Tested Grade, Subject})$

Is use the model to create a predicted performance level (proficiency rate) for each school, considering which grade level test we are looking at, in which subject, the race/ethnicity of the students (where Hispanic concentration is highly correlated with available ELL data, and Hispanic concentration data are more consistently reported), the share of students qualifying for free lunch, the percent identified as homeless and the city of location for the school. That is, each charter school is effectively compared against only other schools in the same geographic context (city).

This is a CRUDE model, which can't really account for other factors, such as the possibility that some charter schools actually shed, or push out, lower performing students over time. More on that below. So, for each school, I get a predicted performance level – what that school is expected to achieve given the children it serves and the location. I can then compare the actual performance to the predicted performance to determine whether the school beats expectations or falls below expectations.

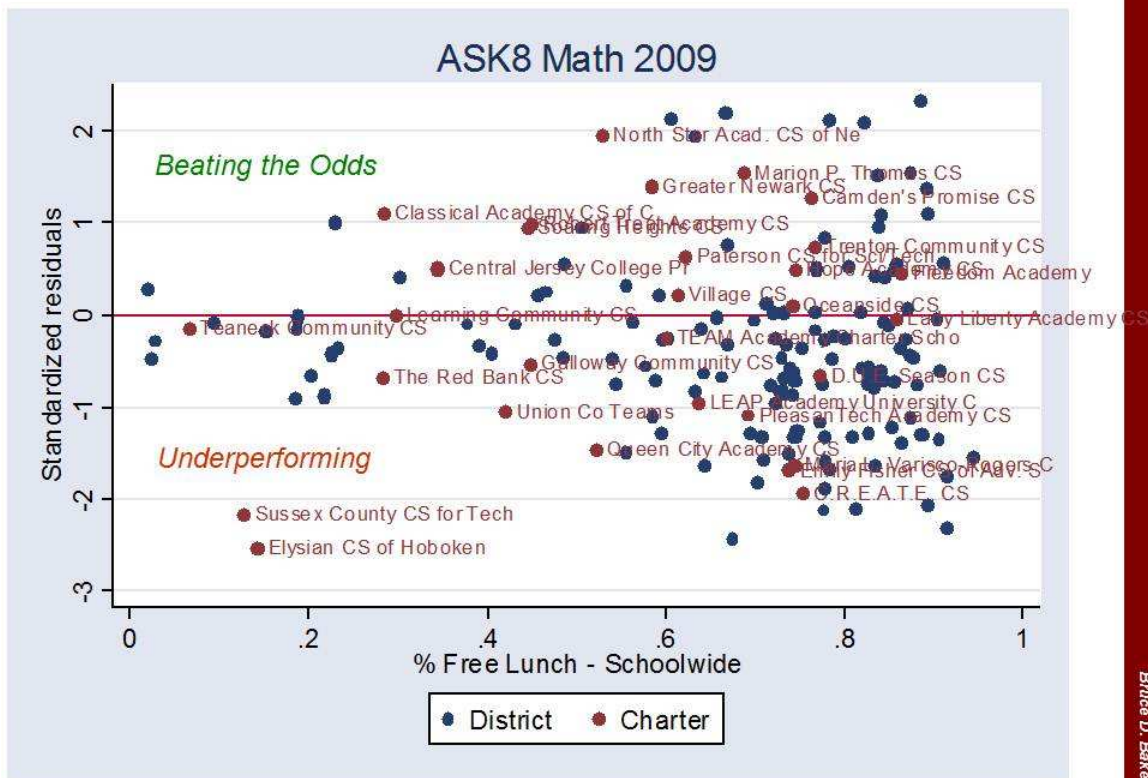
The next two graphs provide a visual representation of schools beating the odds and schools under-performing with respect to expectations. Charters are identified in red and named. Blue circles are traditional public schools in the same district. Note that there are about the same number of charters beating expectations as there are falling short. The same is true for non-charters. On average, both groups appear to be about average.



MODEL: % Proficient or Higher = f(Test Program, Subject, % Free Lunch, %Black, % Hispanic, % Homeless, City)

Bruce D. Baker © 2010

8th Grade Math performance looks much like 4th grade. Charters are evenly split between “good” and “bad,” as are the traditional public schools in their host districts.



MODEL: % Proficient or Higher = f(Test Program, Subject, % Free Lunch, %Black, % Hispanic, % Homeless, City)

Bruce D. Baker © 2010

The Overall Charter Difference (Or Not?)

Now, the above graphs don't directly test whether the average charter performance is better or worse than the average non-charter performance on the same test, same grade and in the same location. But, conducting that test (for these purposes) is as simple as adding into the statistical model an indicator of whether a school is a charter school. Doing so creates a simple (oversimplified, in fact) comparison of the average performance of charters to the average performance of non-charters in the same city (on the same test, in the same grade level), while "correcting" statistically for differences in the student population. I SHOULD POINT OUT THAT ONE CAN NEVER REALLY FULLY CORRECT FOR THOSE DIFFERENCES!

Using this oversimplified method, the analysis (statistical output) below shows that the charter average proficiency rate is about 3% higher than the non-charter average – BUT THAT DIFFERENCE IS NOT STATISTICALLY SIGNIFICANT. That is, there really isn't any difference. THAT IS, THERE REALLY ISN'T ANY DIFFERENCE.


```
. xi: reg perpass pctfree perblack perhisp perhome charter i.city i.testprg i.subj, robust cluster(codistsch)
i.city          _Icity_1-27          (_Icity_1 for city==ASBURY PARK omitted)
i.testprg       _Itestprg_1-7        (_Itestprg_1 for testprg==ASK3 omitted)
i.subject       _Isubject_1-3        (_Isubject_1 for subject==L omitted)
```

Linear regression Number of obs = 3327
F(37, 402) = .
Prob > F = .
R-squared = 0.5567
Root MSE = 15.461

(Std. Err. adjusted for 403 clusters in codistsch)

	perpass	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]
School Demog.	pctfree	-38.46536	6.270445	-6.13	0.000	-50.79232 -26.1384
	perblack	-32.72807	4.722989	-6.93	0.000	-42.01291 -23.44323
	perhisp	-9.014046	5.445967	-1.66	0.099	-19.72018 1.692086
Non-sig. Charter Effect	perhomeless	-75.97004	78.64483	-0.97	0.335	-230.5765 78.63647
	charter	3.319592	1.980754	1.68	0.095	-.5743387 7.213523
	_Icity_2	14.87147	3.385499	4.39	0.000	8.215979 21.52697
City Dummy Variable	_Icity_3	-6.582229	3.857923	-1.71	0.089	-14.16645 1.001996
	_Icity_4	-4.133042	3.556005	-1.16	0.246	-11.12373 2.857648
	_Icity_5	-6.132628	3.737832	-1.64	0.102	-13.48077 1.215512
	_Icity_6	17.23631	4.04591	4.26	0.000	9.282521 25.19009
	_Icity_7	7.055071	3.325915	2.12	0.035	.5167124 13.59343
	_Icity_8	-1.304334	3.302291	-0.39	0.693	-7.796251 5.187584
	_Icity_9	7.777475	3.61024	2.15	0.032	.6801657 14.87478
	_Icity_10	-1.968968	3.802094	-0.52	0.605	-9.443439 5.505503
	_Icity_11	-3746705	3.904522	-0.10	0.924	-8.050503 7.301162
	_Icity_12	8.44878	4.283924	1.97	0.049	.027088 16.87047
	_Icity_13	4.775536	3.157402	1.51	0.131	-1.431547 10.98262
	_Icity_14	4.005859	3.519986	1.14	0.256	-2.914019 10.92574
	_Icity_15	3.296921	4.02698	0.82	0.413	-4.619649 11.21349
	_Icity_16	4.667076	4.000766	1.17	0.244	-3.197962 12.53211
	_Icity_17	12.83232	3.290988	3.90	0.000	6.362623 19.30202
	_Icity_18	4.68808	3.629585	1.29	0.197	-2.447259 11.82342
	_Icity_19	4.962666	3.529469	1.41	0.160	-1.975855 11.90119
	_Icity_20	4.72629	3.650696	1.29	0.196	-2.45055 11.90313
	_Icity_21	5.710325	5.519077	1.03	0.301	-5.139534 16.56018
_Icity_22	6.264448	3.613737	1.73	0.084	-.8397346 13.36863	
_Icity_23	2.535433	4.430867	0.57	0.567	-6.175131 11.246	
_Icity_24	-5.602715	3.76845	-1.49	0.138	-13.01105 1.805616	
_Icity_25	-8.723081	5.500633	-1.59	0.114	-19.53668 2.090517	
_Icity_26	4.770722	4.253334	1.12	0.263	-3.590834 13.13228	
_Icity_27	-5.244844	3.111728	-1.69	0.093	-11.36214 .8724475	
Test Grade & Subj.	_Itestprg_2	1.69956	.78272	2.17	0.030	.1608245 3.238296
	_Itestprg_3	1.751707	.9178945	1.91	0.057	-.0527658 3.55618
	_Itestprg_4	-.429556	1.075458	-0.40	0.690	-2.543781 1.684669
	_Itestprg_5	-1.986832	1.141023	-1.74	0.082	-4.22995 .2562851
	_Itestprg_6	4.857523	1.093846	4.44	0.000	2.707149 7.007896
	_Itestprg_7	7.145066	1.957783	3.65	0.000	3.296295 10.99384
	_Isubject_2	4.119017	.5186227	7.94	0.000	3.099465 5.138568
_Isubject_3	24.43041	.7425033	32.90	0.000	22.97074 25.89009	
_cons	90.34768	3.921469	23.04	0.000	82.63853 98.05683	

Bruce D. Baker © 2010

Some Other Intervening Factors: Cohort Attrition, or Pushing Out

As I mentioned above, even the “tricky statistics” I used cannot sort out such things as a school that systematically pushes out lower performing students, where those lower performing students end up back in the host district. Such an effect would simultaneously boost the charter performance and depress the host district performance (if enough kids were pushed back).⁹

In this figure, we can see that for the 2009 8th graders, North Star began with 122 5th graders and ended with 101 in 8th. The subsequent cohort also began with 122, and ended with 104. These are sizable attrition rates. Robert Treat, on the other hand, maintains cohorts of about 50 students – non-representative cohorts indeed – but without the same degree of attrition as North Star. Now, a school could maintain cohort size even with attrition if that school were to fill vacant slots with newly lotteried-in students. This, however, is risky to the performance status of the school, if performance status is the main selling point.

⁹ <http://schoolfinance101.wordpress.com/2010/12/02/truly-uncommon-in-newark/>

Cohort Attrition Patterns

Charter School	Grade 5 2006	Grade 6 2007	Grade 7 2008	Grade 8 2009
6665-GRAY CS	34	39	30	24
7100-LADY LIBERTY ACADEMY CS	40	54	52	54
7210-MARION P THOMAS CS	43	40	48	36
7280-GREATER NEWARK CS	43		55	55
7320-NORTH STAR ACAD CS OF NEW	122	126	111	101
7325-TEAM ACADEMY CHARTER SCHO	88	95	90	79
7730-ROBERT TREAT ACADEMY CS	52	53	51	55
7735-MARIA L VARISCO-ROGERS CS		34	32	35

Charter School	Grade 5 2007	Grade 6 2008	Grade 7 2009	Grade 8 2010
6665-GRAY CS	27	28	17	
7100-LADY LIBERTY ACADEMY CS	36	31	33	31
7210-MARION P THOMAS CS	40	44	41	34
7280-GREATER NEWARK CS		51	46	39
7320-NORTH STAR ACAD CS OF NEW	122	129	119	104
7325-TEAM ACADEMY CHARTER SCHO	196	179	177	183
7730-ROBERT TREAT ACADEMY CS	50	49	54	49
7735-MARIA L VARISCO-ROGERS CS	14	32	34	

Charter School	Grade 5 2008	Grade 6 2009	Grade 7 2010
6665-GRAY CS	33	31	
7100-LADY LIBERTY ACADEMY CS	38	39	34
7210-MARION P THOMAS CS	45	40	42
7280-GREATER NEWARK CS	51	56	42
7320-NORTH STAR ACAD CS OF NEW	128	122	112
7325-TEAM ACADEMY CHARTER SCHO	196	175	162
7730-ROBERT TREAT ACADEMY CS	50	53	49
7735-MARIA L VARISCO-ROGERS CS	25	34	34

<http://www.state.nj.us/education/data/enr/>

Here, I take two 8th grade cohorts and trace them backwards. I focus on General Test Takers only, and use the ASK Math assessment data in this case. Quick note about those data – Scores across all schools tend to drop in 7th grade due to cut-score placement (not because kids get dumber in 7th grade and wise up again in 8th). The top section of the table looks at the failure rates and number of test takers for the 6th grade in 2005-06, 7th in 2006-07 and 8th in 2007-08. Over this time period, North Star drops 38% of its general test takers. And, cuts the already low failure rate from nearly 12% to 0%. Greater Newark also drops over 30% of test takers in the cohort, and reaps significant reductions in failures (partially proficient) in the process.

The bottom half of the table shows the next cohort in sequence. For this cohort, North Star sheds 21% of test takers between grade 6 and 8, and cuts failure rates nearly in half – starting low to begin with (starting low in the previous grade level, 5th grade, the entry year for the school). Gray and Greater Newark also shed significant numbers of students and Greater Newark in particular sees significant reductions in share of non(uh... partially)proficient students.

Test Taker Attrition

(General Test Takers Only)

2007-08 8th Graders

School	6th Grade % Failing	6th Grade Test Takers	7th Grade % Failing	7th Grade Test Takers	8th Grade % Failing	8th Grade Test Takers	% Drop in Test Takers	% Drop in "Failures"
Gray CS	23.5	34	48.1	27	62.5	24	29%	(1.66)
Greater Newark CS	48.8	41	45.9	37	25.0	28	32%	0.49
Lady Liberty Academy	83.7	43	50.0	42	62.5	32	26%	0.25
Maria L. Varisco-Roge	35.3	34	54.8	31	69.0	29	15%	(0.95)
Marion P. Thomas CS	55.3	38	63.9	36	48.6	35	8%	0.12
North Star Acad. CS o	11.7	60	4.0	50	-	37	38%	1.00
Robert Treat Academy	2.1	48	4.3	47	2.0	49	-2%	0.05
TEAM Academy Charter	46.1	89	32.2	87	43.8	80	10%	0.05

2008-09 8th Graders

School	6th Grade % Failing	6th Grade Test Takers	7th Grade % Failing	7th Grade Test Takers	8th Grade % Failing	8th Grade Test Takers	% Drop in Test Takers	% Drop in "Failures"
Gray CS	11.4	35	50.0	24	8.3	24	31%	0.27
Greater Newark CS	60.5	38	81.8	33	17.9	28	26%	0.70
Lady Liberty Academy	53.1	49	65.1	43	50.0	42	14%	0.06
Maria L. Varisco-Roge	47.1	34	87.0	23	57.1	28	18%	(0.21)
Marion P. Thomas CS	42.9	35	57.6	33	20.0	30	14%	0.53
North Star Acad. CS o	8.0	113	6.8	88	4.5	89	21%	0.44
Pleasant Tech Academy C	39.1	64	67.9	56	63.6	55	14%	(0.63)
TEAM Academy Charter	10.0	21	35.1	77	44.4	81		

My point here is not that these are bad schools, or that they are necessarily engaging in any particular immoral or unethical activity. But rather, that a significant portion of the apparent success of schools like North Star is a) attributable to the demographically different population they serve to begin with and b) attributable to the patterns of student attrition that occur within cohorts over time.

Understanding Parent versus Public Policy Perspectives

Some will say, why should I care if charters are producing higher outcomes with similar kids? What matters to me is that they are producing higher outcomes! Anyone who produces higher outcomes in Newark or Trenton should be applauded, no matter how they do it. It's one more high performing school where there wasn't one previously.

It is important to understand that comparisons of student outcomes that ignore differences in student populations reward – in the public eye – those schools that manage to find a way to serve more advantaged populations, either by achieving non-representative initial lottery pool or by selective attrition. As a result, there is a disincentive for charter operators to actually make greater effort to serve higher need populations – the ones who really need it! And there are many out there who see this as their real mission. Those charter operators who do try to serve more ELL children, more children in severe multi-generational poverty, and children with disabilities often find

themselves answering tough questions from their boards of directors and the media regarding why they can't produce the same test scores as the high-flying charter on the other side of town. These are not good incentives from a public policy perspective. They are good for the few, not the whole.

Further, one's perspective on this point varies whether one is a parent looking for options for his/her own child, or a policymaker looking for "scalable" policy options for improving educational opportunities for children statewide. From a parent (or child) perspective, one is relatively unconcerned whether the positive school effect is function of selectivity of peer group and attrition, so long as there is a positive effect. But, from a public policy perspective, the "charter model" is only useful if the majority of positive effects are **not** due to peer group selectivity and attrition, but rather to the efficacy and transferability of the educational models, programs and strategies. Given the uncommon student populations served by many Newark charters and even more uncommon attrition patterns among some -- not to mention the grossly insufficient data -- we simply have no way of knowing whether these schools can provide insights for scalable reforms.

As they presently operate, however, many of the standout schools do not represent scalable reforms. And on average, New Jersey charters are still just average.