MEMORANDUM

To: Abigail Smith, DCPS Office of Transformation Management
Stefan Huh, Office of the State Superintendent of Education

From: Urban Institute, 21st Century School Fund, and Brookings Institution

Date: May 1, 2009

RE: School Enrollment Projections and Analysis

In the fall of 2008 DCPS contracted the Urban Institute, 21st Century School Fund, and Brookings Institution (Study Team) to provide enrollment analysis and grade-level projections for DCPS schools[1] for the 2009-10 school year. The Study Team has already provided enrollment analysis on the school closings and grade configuration changes to DCPS.[2] The Study Team provided grade-level school-specific enrollment estimates for the 2009-10 school year to DCPS on January 15, 2009, but at that time the statistical model we were developing for this purpose was not adequately tested or developed. Since then we modified the original statistical model and tested its ability to predict at the grade-level and school-specific levels, as well as at the school district level. The purpose of this final memo is to document:

1. The development and testing of a cohort statistical model for grade and school level enrollment projections;
2. Recommendations for its use as part of annual enrollment planning, analysis and projections; and
3. Suggestions for related research and studies that will inform the District's reform efforts, planning, budgeting and evaluation.

Enrollment Projections and Analysis

DCPS student enrollment projections and analysis are important for planning, budgeting and evaluation. Enrollment projections are a basis for both system-level and school level planning and budgeting. DCPS enrollment projections have taken on added weight due to changes in the District's funding process for public education. The DCPS operating budget used to be funded using the Uniform Per Student Funding Formula (UPSFF) applied to the *previous year's* audited enrollment. However, starting in 2007, the DCPS operating budget is now determined using its *next year* projected enrollment, the same method used to determine public charter school funding levels. As a result, DCPS's projection methods need to be explicit and clear to ensure they are of high quality.

---

[1] Excluding special education and alternative schools.

[2] Memos to DCPS from Study Team: Analysis on Impact of Grade Reorganizations, February 20, 2009; and Analysis of the Impact of DCPS School Closings for SY2008-2009, March 17, 2009.

School level enrollment projections are also important for DCPS schools. An accurate budget projection for a local school enables them to start the school year with an adequate plan and staff for the population to be served. When there are significant discrepancies at the local school level equalization of funding and staff occurs once the school year has already begun. Thus schools with projections that were lower than actual enrollments end up with a shortage of teachers until equalization is completed (sometimes up to 3 months into the school year) and schools with projections that were higher than actual enrollments end up with more teachers or staff than needed and staff are reassigned as part of the equalization process. This is an extremely disruptive process.

Enrollment analysis can be used to evaluate the effect and impact of system-wide and school-specific educational or organizational reforms or interventions. In the District's public education system, with so many school choices for families, changes in enrollment patterns are one indicator of how well DCPS is both delivering and communicating quality school options for families with school-age children. Finally, enrollment projections are a critical element of an educational facility master plan and important information for space planning.

Although important, enrollment forecasting for public schools in Washington, DC is unique and challenging. Few school districts have similarly liberal enrollment policies governing student options within their district ***and*** such a large and growing charter school sector. New Orleans may be the only comparable district. In a literature search, there were no useful studies on methods for enrollment projections in systems with such high levels of school choice. The work that DCPS is doing in this area is pioneering and needed, and it will also offer other school districts important guidance as new models of governance and choice develop in other communities.

Systemic Process for Enrollment Projections

The Study Team worked to develop enrollment planning processes and methods that can be used to prepare high-quality *next year* enrollment estimates and projections for DCPS. As part of a recommended process, the Study Team developed annual process protocols for making enrollment projections for DCPS. The steps involved in this process are:

1. Collecting, cleaning and merging administrative student-level and school-level panel data for analysis.
2. Collaborating with external research partners to analyze enrollment patterns, test the statistical projection models, and determine if any additional revisions to the model are necessary due to DCPS or public charter programmatic changes.
3. Applying updated statistical models to generate estimates for *next year* grade and school level enrollments using panel data.
4. Reviewing statistical model output in order to make and track adjustments at the grade, school and district levels, using a simple software data interface.
5. Documenting the statistical model and its assumptions and the technical and programmatic adjustments that are made to the model's estimates.

There is more work that needs to be done on the statistical model and process protocols, but tremendous progress has been made on this, which is documented here. The Study Team recommends that the revised statistical model and proposed process and protocols be fully piloted for the *next year* enrollment projections for the 2010-11 school year. After refining the development, testing, and use of process protocols, DCPS will have a high quality process and be able to manage its annual projections with in-house staff and at low cost.

Data Preparation

The first requirement of a high quality enrollment planning process is administrative data that is reliable, consistent, accessible, complete and timely. This is a challenge, but the District, including DCPS and the Public Charter School Board, has been making tremendous improvements in the student-level data that forms the foundation of the statistical modeling developed for grade level and school level projections.

In order to do statistical modeling for next year's enrollment projections, basic student and school level data needs to be collected, cleaned, and compiled. The data used by the Study Team for the 2009-10 student enrollment projections include:

- Six years of DCPS student enrollment data from DC STARS (October pre-audited data), including student IDs, home addresses, school attended, and grade.
- Six years of DCPS and public charter school inventory compiled by the Study Team including school names, locations, and grades offered.

In order to clean and compile the enrollment projection dataset, the Study Team geocodes every student's address. The Study Team runs the student-level data through their in-house geocoding macro, which corrects typical street misspellings and missing quadrants. After running through the geocoding macro that automatically fixes the majority of errors, Study Team staff spends significant time hand cleaning the problematic addresses that still could not be geocoded. The Study Team ensures that at least 95 percent of all student addresses are geocoded. For example, 98 percent of all students from 2008-09 were goeocoded. This process is time consuming; it took Study Team staff approximately 200 hours to clean and process the 2007-08 and 2008-09 student-level data.

The Study Team also ensures that the student-level data have the correct DCPS building numbers and grade assignments. Due to programmatic changes such as school consolidations and swing space, DCPS has inconsistently assigned students' school building numbers. The school building numbers are crucial in the student projection model and Study Team staff must decipher what the correct building code should be. Also, DCPS occasionally combines schools into one building number, such as Peabody and Emilia Reggio.

The Study Team compiles a longitudinal DCPS and public charter school-level dataset that includes every school (DCPS or public charter), its location (geocoded with longitude and latitude coordinates), and minimum and maximum grades starting with SY2003-04. This was a time consuming process to create school file for past years, as there is no consolidated history of the location and grade configurations of public charter schools in particular.

From the school-level, the Study Team then calculated the supply of grade-appropriate schools (DCPS and public charter) within a half mile, 1 mile, and 2 miles of every DCPS school for every year (taking into account the schools that changed the grades they offered). We then calculated the change in the number of schools offering particular grades between every two-year spread. (These variables were included in the basic model.)

To run the enrollment projection models, we compiled a panel dataset from the student-level and school-level file that included an observation for every DCPS school at every grade with the necessary dependent and independent variables (current enrollment, cohort enrollment, change in school supply at the three distances, and school and grade dummy variables). There was a total of 4,799 observations in the dataset that was used to run the cohort basic model.

Applying Models to Create Enrollment Projections

Although statistical modeling cannot perfectly predict the future, statistical modeling can be a powerful component of a high quality enrollment projection process and of enrollment choice analysis. The first step in modeling school and grade enrollments is understanding historic enrollment patterns and choices. Based on earlier studies done jointly by 21st Century School Fund, Urban Institute and Brookings Institution, we identified factors that we thought would predict future enrollments such as where a student was in school the year before. But in the District of Columbia, we knew this was not going be a close enough estimation. Although the District has experienced relatively stable total public school (DCPS and public charter) enrollment over the past decade, there has been considerable enrollment switch to public charters resulting in declining enrollment in DCPS overall and sizable enrollment movement between DCPS schools. For instance, in 2008-09, only 37% of all public school students (DCPS and public charter school students) attended their in-boundary DCPS school.

A major challenge for school level projections is that many parents do not commit to their school. We know from earlier analysis that 14 percent of public school students switched schools from one year to the next before completing the final grade offered at their original school.[3] Some of these school switches are due to students moving but another reason is that families are leapfrogging between schools.

Another challenge in forecasting school enrollments is that schools, communities and neighborhoods also change affecting future enrollments in both the near and long term. The reasons parents are moving their children may be school related but they may also be due to other changes in the life of the family or neighborhood. Other research has found that nearly 60% of student mobility can be attributed, at least in part, to residential change.[4] In analyzing address changes of students, we found that 20% of DCPS students had changed their home address or moved between the start of the 2007-08 and 2008-09 school years. More specifically,

[3] *Quality Schools and Healthy Neighborhoods: A Research Report*, 21st Century School Fund, Brookings Institution, Urban Institute, Washington DC, September 2008.
[4] "Patterns of Urban Student Mobility and Local School Reform," David Kerbow, University of Chicago, Report No. 5, 1996

for DCPS students who attended DCPS schools that were closed in 2008-09, 25% of these students' families had changed addresses between the two school years, 6 percentage points higher than students who attended schools that did not close.

There are also major changes in school supply that affect enrollment patterns and complicate enrollment projections. In recent history, DCPS closed and consolidated 23 schools and new public charter schools have opened, including the conversion of the 7 Catholic schools into public charter schools. In addition to new public charter schools, many public charter schools are still progressively adding grades and expanding campus locations. DCPS has also made changes in grade configuration both extending grades in some schools and reducing them at others.

The challenge of generating high quality school level projections can be seen in an analysis of previous DCPS projections. In previous years, DCPS generated its own "next year" student enrollment projections. We were not provided written documentation describing their earlier methods, but we understand that the approach was generally as follows:

- DCPS facilities planning staff requested from principals their grade by grade estimates for the next year.
- DCPS staff then adjusted these estimates based on the enrollment history of the school and other factors that they thought might affect enrollment.
- These projections were returned to the principals for review and principals with concerns worked their concerns on an ad hoc basis with the assistant superintendents.
- School level projections were used for the local school budget allocation and grade level projections were used for the out of boundary lottery.

This process resulted in projections that have been relatively accurate at the district level (within 7% overall for the 06-07 school year and within 1% for the 08-09 school year), but they have ranged broadly at the school level. In 2008-09 56% of all basic grade level schools fell within 10% of their projected levels. Table 1 shows the comparison of two years of projections versus actual enrollment done by DCPS.

Table 1: DCPS Projections vs Actual Enrollment for SY2006-07 and SY2008-09

	DCPS Projections for SY06-07				
	All Students Projected	**All Students Actual**	**Student Difference**	**Percent Difference**	**Share of Schools within 10% of Predicted Enrollment**
ES Totals	32,349	30,779	1,570	5%	73%
MS Totals	8,689	7,344	1,345	18%	40%
HS Totals	12,438	11,818	620	5%	65%
Total	**53,476**	**49,941**	**3,535**	**7%**	**67%**

	DCPS Projections for SY08-09				
ES Totals	25,749	25,683	66	0%	54%
MS Totals	5,853	5,339	514	10%	47%
HS Totals	12,918	12,913	5	0%	72%
Total	**44,520**	**43,935**	**585**	**1%**	**56%**

Note: The Study Team did not have access to the DCPS projections for SY07-08.

Development of a Statistical *Next Year* Enrollment Model

To try to address the challenges of the public school enrollment environment, establish and follow a clear, defined method of enrollment projections, and document the process for use in future years, the Study Team created a grade-specific school-specific econometric model that takes into account the change in school supply. The model is primarily based on the number of students enrolled in the previous grade the previous year (the grade cohort); however, the Study Team included an additional set of variables to capture change in school options nearby, specifically whether more or fewer grade-appropriate DCPS or public charter schools would be open near the school being estimated. The Student Projection Data Methodology appendix (Appendix A) describes the process of data cleaning, modeling, and testing the results of the student projection model. The following sections briefly summarize the Study Team's efforts.

The model that we recommend using for forecasting student enrollment for each grade at every school was finalized in early April 2007. (Models were developed as early as January 2009 but the Study Team continued to refine the model until April 2009. Appendix A describes the development of the models in more detail.) The latest revised model's unit of observation is a particular DCPS school (S) at a particular grade-level (G) in a particular year (T). The dependent variable is total enrollment (number of students) in a grade at a school for a given year -- $Enrollment_{S,G,T}$

The recommended model[5] is:

$$Enrollment_{S,G,T} = function\{Enrollment_{S,G-1,T-1} + + Enrollment_{S,G,T-1} + Wards_S \times Enrollment_{S,G-1,T-1+} Grade \times Enrollment_{S,G-1,T-1} + New\ Supply\ Dummy\ .5\ miles_{S,G,T} + New\ Supply\ Dummy\ 1\ mile_{S,G,T} + New\ Supply\ Dummy\ 2\ miles_{S,G,T} + Grade\ Dummies_T + DCPS\ School\ Dummies_T + Constant\}$$

[5] There were two variations to this model for grades in schools that did not have an earlier cohort, typically kindergarten and grades 5, 6, and 9. See the Student Projection Methodology appendix for more detail.

The recommended model is a function of last year's enrollment in the previous grade, the number of students enrolled in the same grade the previous year (or grade history), an interaction variable between the school's ward and cohort enrollment, an interaction variable between grade and cohort enrollment, the change in the number of schools (DCPS or public charter) offering the estimated grade with a half mile, 1 mile, and 2 miles of the school, and dummy variables for the grade being predicted and the school being predicted. (The Student Projection Methodology appendix describes the variables in more detail and the rationale for including them.)

We did not apply an econometric model to predict enrollment for non-mandatory grades (pre-school and pre-kindergarten) or special DCPS programs/schools, such as alternative education, adult education or special education schools. The size of the early education grades and special programs is typically determined by DCPS administrators.

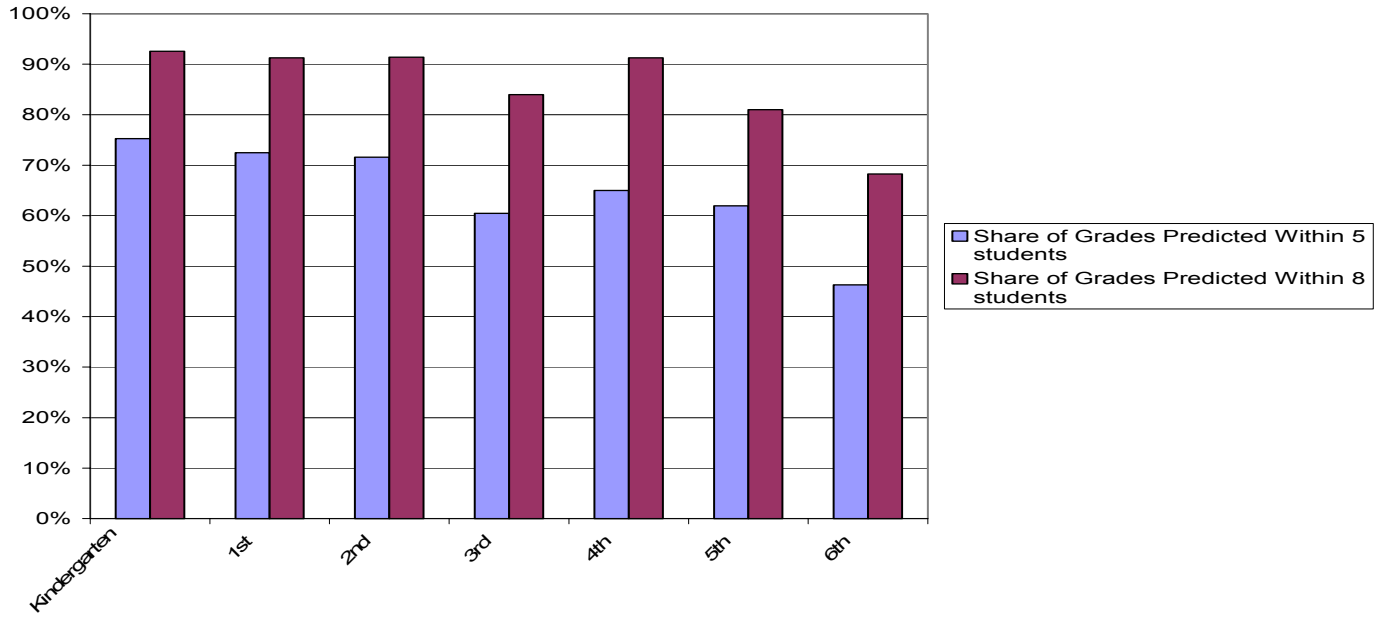
Testing the Statistical Model

In order to understand how well the model performs in estimating DCPS enrollments, we compared the model's grade-specific school enrollment estimates for the 2007-08 school year against the actual 2007-08 enrollment. (The 2007-08 school-grade enrollment estimates were derived from the model using all six years of DCPS data.) This comparison showed how accurate the model was in estimating at the grade-specific school level. Then we summed the grade-level estimates in every school to determine how accurate the aggregated estimates were at determining total enrollment at every school (all grades combined) in 2007-08. Finally, we summed all the aggregated DCPS school estimates together to determine how accurately the model estimates the total DCPS district enrollment.

Figures 1-3 summarize our model's ability to estimate enrollment for the 2007-08 school year within the tight standard of within 5 students and 8 students for the elementary school grades (K to 5th grade), within 10 to 15 students at the middle school grades (6th to 8th grades), and within 20 students and 30 students for senior high schools (9th to 12th grades). These thresholds were selected using the average citywide grade size at each of these levels and correspond to approximately 10 to 15% of the average grade. We set a high standard for the model since its practical use for budgeting staffing allocations at the individual school level requires a close fit between the projected enrollment and the actual enrollment. For purposes of our initial assessment of the model, we applied the standard without regard to the total enrollment across grades at each school.

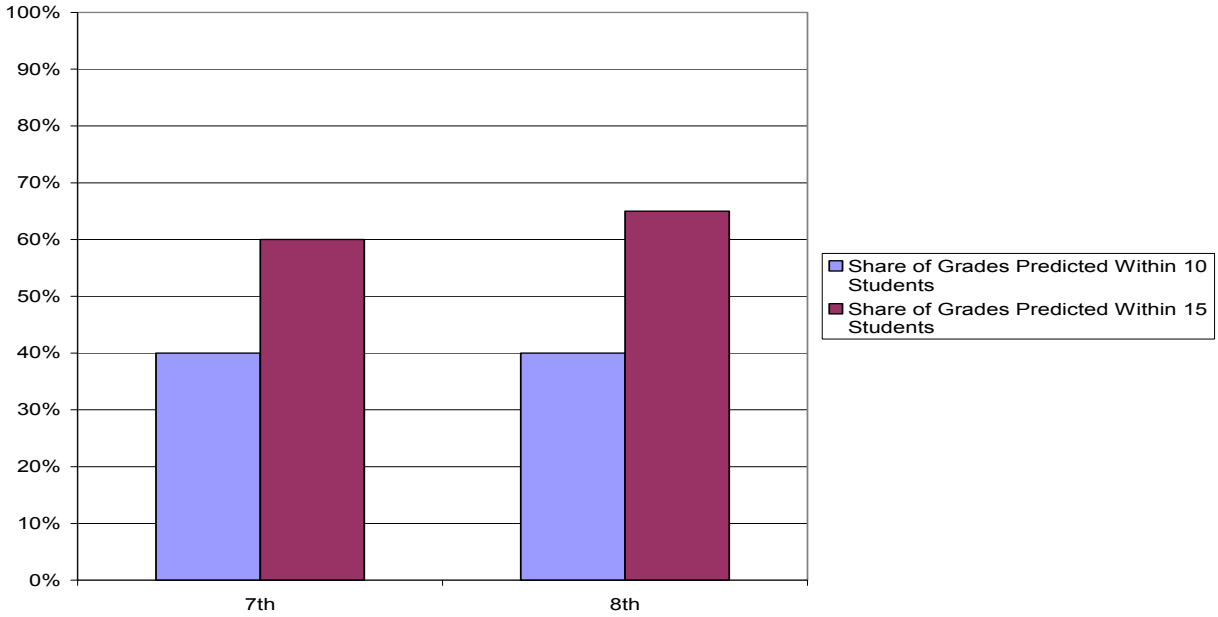
We found that, on average, the model estimated within 5 students per grade level for the elementary school grades 66 percent of the time and within 8 students 87 percent of the time. The model was particularly accurate within 5 students for grades K, 1, and 2 (see Figure 1).

Figure 1: Accuracy of SY2007-08 Grade-School Estimates, Elementary Grades



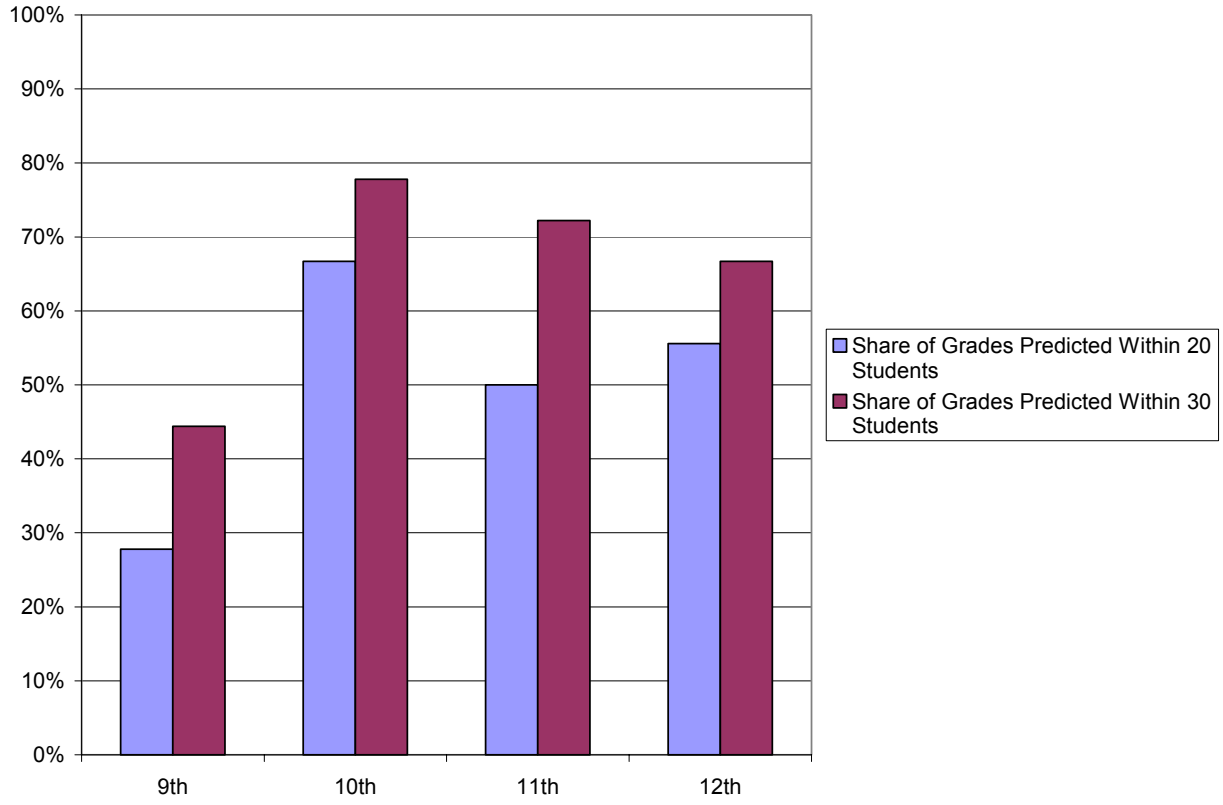
The model was less accurate meeting the standard of within 10 and 15 students when estimating for the middle school grades (7th and 8th grades): only 40 percent of the middle school estimates were within 10 students and 63 percent were within 15 students. The model accurately estimated within 10 students for 40 percent of both 7th and 8th grades and the model estimated slightly better for the 8th grade.

Figure 2: Accuracy of SY2007-08 Grade-School Estimates, Middle Grades



The model was less accurate at the high school grades (9th to 12th grades) compared to the elementary grades but more accurate than the middle school grades. The estimated school-grade enrollment was within 20 students for 50 percent of high school grades and within 30 students for 65 percent of the high school grades. As Figure 3 shows, the model was the most accurate estimating for grade 10 and had a very low accuracy rate for grade 9. Surprisingly, the estimates for the high school grades worsened with the introduction of the revised model developed in April compared to the original model submitted in January. However, if we extend the threshold to include 31 students (as opposed to 30 students), the expanded model estimated 72 percent of the share of school-grades for the high school grades.

Figure 3: Accuracy of SY2007-08 Grade-School Estimates, Senior High Grades



Total School Estimations

The model provides school-grade specific enrollment for every DCPS school, other than alternative, adult and special education schools. However, for planning, budgeting and evaluation, accurate school-level projections are most important (or forecasts for the enrollment of the entire school irrespective of grades). Therefore, the Study Team aggregated the individual grade level estimates at each school, which smoothed the variations between the individual grade estimates.

Although grade-specific prediction accuracy is very difficult to achieve at the high standards we set, the aggregated school-level estimations were extremely accurate. The model predicted 78%

of all schools within 10 percent of their actual enrollment, with the highest school-level accuracy at the high school level (94%) (see Table 2).

Table 2: Aggregated School-Level SY2007-08 DCPS Estimations vs Actual Enrollment

	Estimated SY07-08 Enrollment	Oct Unaudited SY07-08	# Difference	Percent Difference	# of Schools within 10%	# of Schools	Share Schools within 10%
ES/PS-8 Totals	19,949	19,772	177	1%	64	81	79%
MS Totals	4,362	4,294	68	2%	8	15	53%
HS Totals	12,880	13,516	-636	-5%	17	18	94%
Total	**37,191**	**37,582**	**-391**	**-1%**	**89**	**114**	**78%**

Similar to the experience of earlier DCPS projection efforts, the middle grades are the most difficult to predict (as found in Table 1). There are several likely reasons for this. First, there is high student mobility at the middle grades, both between sectors and among DCPS schools. Our earlier analysis showed that feeder patterns from elementary to middle schools are weak (the average DCPS middle school in 2006-07 enrolled students who attended 41 different elementary schools) and charter participation is highest at the middle grades. Our recent analysis of the 2008-09 grade configuration changes showed the middle grades trend to charters accelerated, i.e., the charter participation of 6th graders increased from 35.9% of all public school 6th grade students in 2007-08 to 45.9% in 2008-09. Second, there have been multiple changes in grade configurations at the middle grades, causing uneven enrollment trend lines making it difficult for the model to accurately estimate. For instance, in 2007-08, DCPS converted its junior high schools (serving grades 7-9) into middle schools (serving grades 6-8), and over the past several years it has staggered the transition of 6th graders out of elementary schools, meaning that each year the middle schools have a different number of elementary schools either holding or sending 6th graders.

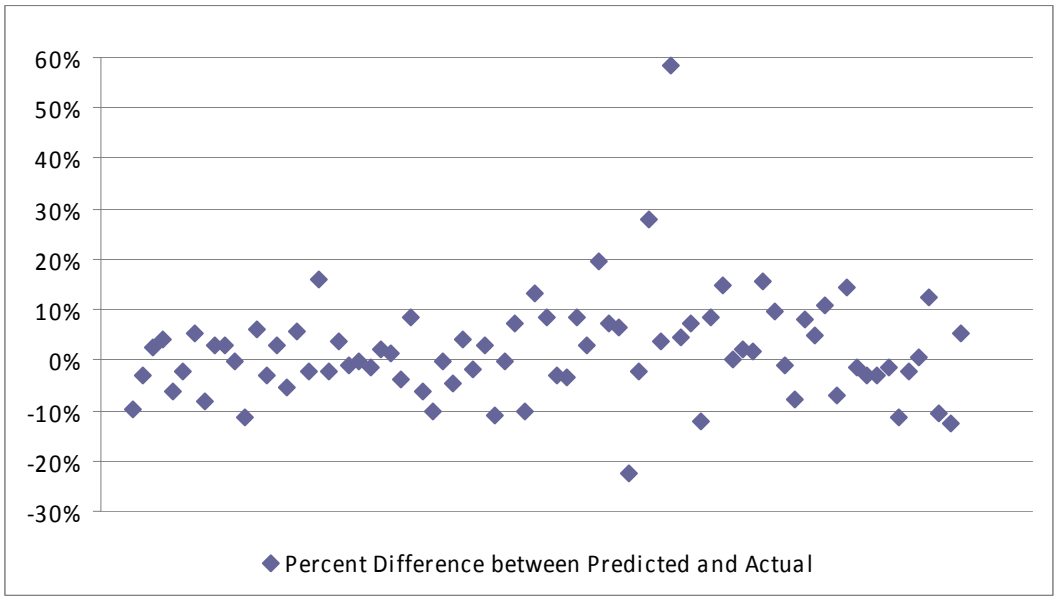
Going forward, the 2008-09 conversion of 14 elementary schools to PK-8 education centers will also impact middle school enrollment patterns. Finally, the supply of public schools has continued to increase at the middle grades, with new charter schools opening or established charters expanding grades, in addition to the new DCPS PK-8 schools. While the model accounts for new grade-appropriate school supply, often from charter schools opening or expanding, the school supply variable may not be sufficiently capturing the enrollment effect of these changes. These programmatic changes make accurate statistical estimation very difficult, and DCPS will need to adjust the model's predictions based on the process that we describe at the end of the memo.

Process Protocols: Reviewing Statistical Enrollment Estimates

Since no statistical model will exactly predict future enrollment, the process of reviewing how the model estimates enrollments for a year where the actual enrollments are known is critical and the first step in the review process protocol. Analysis of the actual versus predicted enrollment results may show that the model needs further modification due to unexpected or undetected data

errors. This section describes our review of the results of the application of the recommended model. For example, Figures 4, 5, and 6 show scatter plots of the percent difference between the estimated enrollments versus actual enrollment for SY07-08 – the smaller the difference, the more accurate the estimated enrollment.

Figure 4: Difference Between Estimated and Actual Enrollment, SY2007-08 Elementary Schools

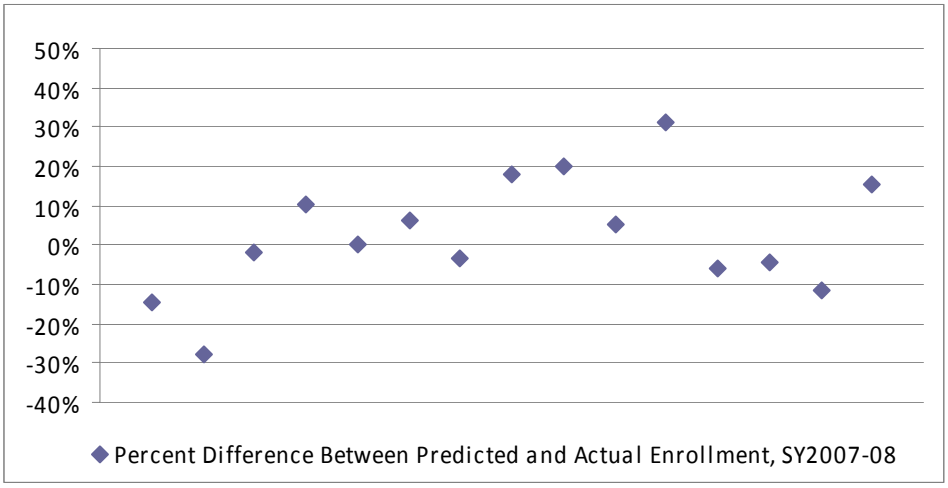


For the vast majority of elementary schools, the difference between estimated and actual enrollments clustered within +/-10 percent. There are understandable reasons why the estimated enrollment differed from actual enrollment by significantly more than 10 percent for some elementary schools. The most extreme outliers were Savoy Elementary (the model over estimated by 58%), River Terrace Elementary (the model over estimated by 28%) and Peabody Elementary (the model under estimated by 22%). In 2007, Savoy moved from its home location into swing space at Birney Elementary while Savoy was scheduled to begin a modernization and expansion. Parents who had been sending their children to Savoy at its Shannon Place location were not willing to have their children cross Howard Road, a dangerous and major intersection with Martin Luther King, Jr. Boulevard, to attend Birney, and as a result of the school's move, Savoy lost nearly half of its students. This one year change could not be adequately estimated by the model.

The explanation for River Terrace is less clear cut, but since the school has generally carried a large special education population we surmised that physically handicapped students who were often assigned to River Terrace may have been reassigned to the more modern and fully compliant new elementary schools.

The under prediction at Peabody appears to be due to the fact that the DCPS school numbers are the same for Peabody and Reggio Emilia School-within-a-School for the test year. This illustrates how reviewing results can help identify data errors as part of the quality check process, and the importance of such review by users familiar with the school and grade-level data.

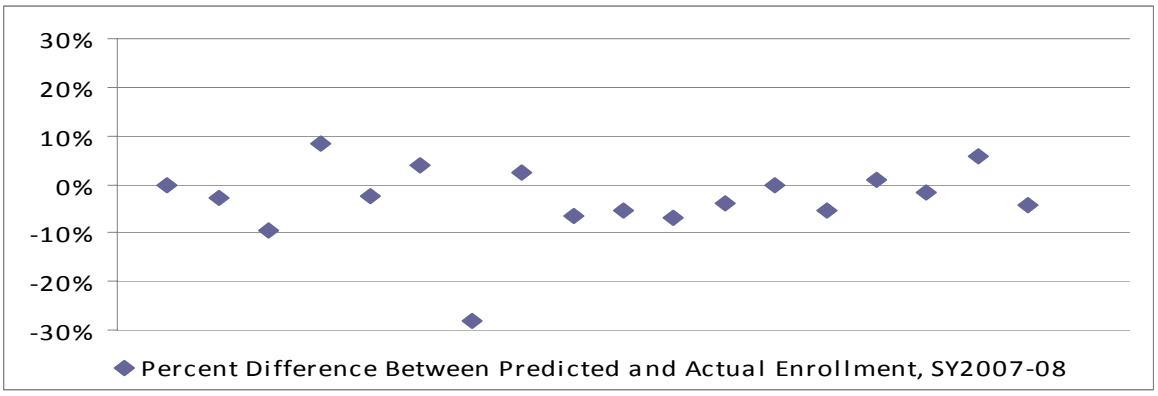
Figure 5: Difference Between Estimated and Actual Enrollment, SY2007-08 Middle Schools



As discussed, estimations for the middle schools were less successful. The difference between estimated and actual enrollment was particularly striking for four schools: Browne JHS (model under estimated by 28%), Johnson JHS (model over estimated by 18%), Kramer MS (model over estimated by 20%), and Kelly Miller MS (model over estimated by 31%). The underestimation at Browne is likely a data problem, as DCSTARS mistakenly switched the Browne Center and Browne JHS school IDs assigned to students. At Kelly Miller, there was a significant expansion of other school offerings at the middle grades, which made estimations less accurate: Cesar Chavez Parkside continued to expand its 6th-8th grades; Maya Angelou opened a middle school at Fletcher Johnson with 100 6-8th graders in SY07-08; and KIPP opened a middle school across the street from Fletcher-Johnson.

School-level predictions were strongest at the high school level, with only one outlier, Eastern HS, where the model under estimated by 28%. The test year, 2007-08, was the year that DCPS transferred 9th grade from the middle schools to the high schools, a change that would not have been reflected in the historic high school enrollment trends used by the model. Not only was this programmatic change not accurately reflected in Eastern's estimation, but many of the surrounding middle grade schools – Eliot, Hine, Jefferson, and Browne – were junior high schools that served 9th graders prior to 2007-08.

Figure 6: Difference Between Estimated and Actual Enrollment, SY2007-08 High Schools



In many of these outlier cases, recent programmatic changes could not be accurately incorporated by the model. Therefore, DCPS should expect to have to make school-level enrollment prediction adjustments to outliers. Appendix B provides the projections and actual enrollment for 2007-08.

City-wide grade level test of model

Another way to understand the success of the model is to look at how it estimates at each grade citywide. As Table 3 shows, the model performed best at a citywide aggregate level for grades K-4, 7, 8, and 11th, and struggled most with grades 6, 9, and 10. The strong performance at grades 7 and 8 appears contrary to our school level tests at the middle grades, suggesting that relatively equal amounts of under and over-predicting may have evened out to a better net performance. The under-predicting at 9th grade may be a result of the transition from junior highs to high schools, or the model may be insufficiently capturing the extent of 9th grade retention.

Table 3: Citywide Grade Level Comparison of Predicted and Actual Enrollment, SY2007-08

Grade Predicting	prediction_0708	Actual 2007 Enrollment	Difference between Predicted and Actual	Percent Difference Between Predicted and Actual
0 Kinder	3,268	3,251	17	1%
1st	3,249	3,222	27	1%
2nd	3,374	3,325	49	1%
3rd	3,158	3,094	64	2%
4th	3,148	3,131	17	1%
5th	2,967	2,857	110	4%
6th	1,604	1,518	86	6%
7th	2,059	2,089	(30)	-1%
8th	2,460	2,439	21	1%
9th	3,846	4,230	(384)	-9%
10th	3,203	3,425	(222)	-6%
11th	2,841	2,809	32	1%
12th	2,555	2,650	(95)	-4%
ALL GRADES	**37,733**	**38,040**	**(308)**	**-1%**

Application of Expanded Model for Enrollment Projections for SY2009-10

The models submitted on January 15, 2009 were not developed enough to use for the FY2010 local budgeting process, although we did provide preliminary grade level estimates, by school, derived from the basic statistical model. We have rerun these estimates by school-grade and by total school using the expanded model (see Tables 4 and 5). The projected change from the current year's enrollment appears to follow similar patterns as the tested 2007-08 estimates. While the accuracy of the projections cannot be determined until they are compared against the actual fall 2009 enrollment, examining the projected change now can help DCPS to anticipate budget and staffing needs, as well as where adjustments to the model may be necessary.

Overall, the model projects an increase of 753 students, or 2%, over the 2008-09 enrollment. The projected growth at the elementary school level is largely driven by the expansion of the new PK-8 schools, which are still adding grades. While one might expect this enrollment expansion to bring concurrent decline at the surrounding DCPS middle schools, our analysis of the grade reconfiguration changes suggests that the PK-8s do better retaining middle grade students in DCPS, and thus may contribute to an overall growth in DCPS population at these grades.

In addition, the stasis in the middle schools documented in Table 4 includes the addition of 6th grade at Deal. If this new grade, which is projected to enroll over 200 students, is removed from the middle school analysis, the model would project middle school enrollment to decline by 5 percent, which is more consistent with recent years' experience. It is also important to note that Table 5 shows the model projecting no change for middle schools in 2009-10, but Table 5 shows 7th and 8th grade increasing by 15 and 13 percent, respectively. This difference is likely driven by the differential enrollment patterns in 7th and 8th grade at the new PK-8s (growing) compared to existing middle schools (declining). The discrepancy suggests that in future years the citywide grade estimates should be broken out by grade type for better accuracy – elementary, education centers, middle, and high schools. The model is predicting steady enrollment at the senior high schools.

Table 4: Comparison of 2009-10 Model Predictions with Current Year's Enrollment

Study Team Model Projections SY09-10 compared to SY08-09 Actual Enrollment							
	All Students Projected, SY2009-10	All Students Actual, SY2008-09	# of Students Difference between Projected & Actual	Percent Difference between Projected & Actual	Share of Schools within 10%	# of Schools within 10%	# of Schools
ES Totals	22,517	21,820	697	3%	71%	60	84
MS Totals	4,850	4,834	16	0%	71%	10	14
HS Totals	12,985	12,945	40	0%	86%	18	21
TOTAL	**40,352**	**39,599**	**753**	**2%**	**74%**	88	119

Table 5 Predicted 2009-10 Enrollment Compared to Actual 2008-09 Enrollment by Grade Level

Grade Predicting	Predicted Enrollment 2009-10	Actual Enrollment 2008-09	# Difference	Percent Difference
Kindergarten	3,412	3,294	118	3.6%
1	3,289	3,487	-198	-5.7%
2	3,395	3,551	-156	-4.4%
3	3,538	3,632	-94	-2.6%
4	3,519	3,269	250	7.6%
5	3,024	3,102	-78	-2.5%
6	2,169	1,998	171	8.6%
7	2,639	2,295	344	15.0%
8	2,757	2,436	321	13.2%
9	4,417	4,366	51	1.2%
10	2,864	2,901	-37	-1.3%
11	2,700	2,676	24	0.9%
12	2,629	2,592	37	1.4%
Total	**40,352**	**39,599**	**753**	**1.9%**

Note: The 2009-10 predictions do not include special education or alternative education schools. The actual enrollments in 2008-09 only include those schools/grades that were predicted in 2009-10. In some cases, the number of grades differs between the two school years. In schools that dropped specific grades in 2009-10, for instance the elementary schools that will no longer serve 6th graders, those grades were not predicted for and their 2008-09 enrollment is not included in this table.

Appendix C includes the school-grade level SY2009-10 projections, as well as comparison of the school by school actual enrollment for 2008-2009 and the total enrollments estimated with the expanded model. Based on our model testing for the 2007-08 school year, we anticipate that there will also be outliers in these estimates. Specifically, we think we need to pay attention to enrollment estimates for the following types of schools:

- Schools affected by school closing and grade configuration changes. For example, some of the assigned receiving schools had enrollment spikes in 2008-09, such as Tubman that gained 80 students. This enrollment growth is not expected to continue in 2009-2010 but may result in an over-estimation for Tubman.
- Expansion of public charter schools, especially in wards 7 and 8, where rapid expansion may impact DCPS enrollment more than the model captures.
- New/modernized DCPS schools opening in 2009-10 may attract more student than they have in the past, resulting in an underestimation of the model. These schools are Savoy, Deal, School Without Walls, Addison/Hyde, and Walker Jones.
- For schools that have recently undergone programmatic changes (such as consolidations), DCPS STARS data may have resulting data glitches. For example, there are school building number problems with consolidated schools (i.e., Turner and Green). Also,

schools that have added new grades will not have historic data to draw from resulting in inaccurate predictions (i.e., Phelps and the new PK-8 schools).

Estimating and Projecting for Special Populations

The Study Team did not apply the school-student enrollment model to predict enrollment for non-mandatory grades or special DCPS programs/schools, such as pre-school and pre-kindergarten, adult/evening schools, alternative education schools, or special education schools. The size of the early education grades and special programs is typically determined by DCPS administrators.

In 2008-2009 there were 6,108 of these students enrolled in DCPS.

Specialized Student Groups	Enrollment, SY2008-09
Pre School Students	1,549
Pre Kindergarten Students	2,522
Alternative Education Students	98
Adult/Evening Education Students (STAY and Twilight Programs)	1,343
Special Education School Students (in special ed schools)	596
TOTAL	**6,108**

Since the city needs to plan and budget for these students and their programs, as well as regular education schools, it is helpful to understand recent trends among these students and where these students live. Citywide public early childhood enrollment (pre-school and pre-kindergarten) has increased by nearly 20 percent over the past 3 years, likely reflecting rising birth rates in the city. However, much of this increase is occurring in expanding pre-school and pre-kindergarten programs in the charter schools, as DCPS early childhood enrollment has remained relatively stable during this time period. There are DCPS schools in some neighborhoods that have experienced significant increases in their early childhood enrollments due to increased birth rates, changes in the neighborhood housing markets, and families committed to enrolling their children in the public school system (Capitol Hill is a strong example of this phenomenon). Also, the funding being provided for universal pre-school will likely increase the enrollment of the early childhood population in DCPS.

For the January 15 enrollment submissions, the Study Team provided DCPS with a matrix of pre-school and pre-kindergarten enrollment over time for each school matched to the birth and housing market trends in the corresponding neighborhood cluster. DCPS, along with special education specialists, then determined for each school whether pre-school and pre-kindergarten classes would remain the same size as the year before, decline, or increase for every DCPS school. The Study Team recommends that DCPS implement this process for next year.

The enrollment capacity in alternative education programs is also largely determined by DCPS central administration. The two CHOICE programs were consolidated for the 2008-09 school

year and the Youth Engagement Academy opened at Eliot serving 50 9th graders. Youth Engagement Academy will expand in 2009-10 to serve 100 9th and 10th grade students. The STAY program at Roosevelt has declined just over 10% in the past three years, while the Spingarn and Ballou STAY programs have both experienced significant enrollment increases (60% at Spingarn and 20% at Ballou). In addition, the Twilight program opened at Ballou with 66 students in 2008-09 to help students needing credit recovery to complete high school. Since the District of Columbia has had such a long history of dropouts, the demand for high quality alternative and adult education programs is an area that if it were invested in on the program side could see an increase in student demand and therefore increased enrollment.

Finally, as part of the 2008-09 reorganization, DCPS closed five of its special education centers (Moten, Browne, Taft, Washington, and Spingarn); those students distributed across the public (and some non-public) schools, with large concentrations at Shadd Transition Academy (25%) and Hamilton Center (13%). The enrollment did not change measurably at the other special education schools which remained open – Mamie D Lee, Sharpe Health, Prospect Learning Center and Jackie Robinson. There needs to be a process that incorporates projections for special education population that examines transition from private placement, efforts by DCPS and OSSE to improve programs and services and also looks at indicators for new identification—for example, retention.

Recommended Process for Enrollment Projections

The Study Team has made tremendous progress on developing an accurate, standardized, and transparent process for developing grade, school, and citywide enrollment projections. While the model and process needs to be further refined, the Study Team recommends that the statistical model and proposed process and protocols be fully piloted for the 2010-11 enrollment projections.

Table 5 describes the process steps and schedule for estimating the 2010-11 enrollment projections.

Table 5: Basic Activities and Schedule for Enrollment Projection Pilot

Schedule	Task
June/July 2009	Present current model and analysis of the model's accuracy to other researchers outside the Study Team for discussion and input. Consider and test modifications to current model, for example including school capacity, not just whether a school is in proximity. Explore modifications to the statistical methodology for the next year's citywide DCPS projection, for example using a time series method.
July/August 2009	Run additional tests on statistical model used for preliminary 2009-2010 projections. For example, test model for 2008-09 (using 2003-04 to 2007-08 data).

	Examine results of any modifications proposed to current model or statistical approach.
August/September 2009	Review and revise report format used for DCPS grade and school level enrollment projection review. Devise report format needed by DCPS for city-wide enrollment projection review.
September 2009	Use DCPS preliminary September 15, 2009 enrollment counts to compare to the Study Team's 2009-10 projections. Conduct analysis to determine if models were more or less accurate for specific schools, grades, wards, etc.
August/September 2009	Prepare ***school-level*** data for 2010-11 estimates, including grade ranges and new grades offered.
October/November 2009	Clean and geo-code 2009-10 October ***student level*** data. Data must be received by mid October to be completed by the end of November.
Early December	DCPS reviews planned programmatic adjustments – grade configuration changes, new programs, school closings, changes in special ed/adult/alternative ed, changes in PS/PK classrooms-- that DCPS anticipates may affect future enrollments and should be incorporated into the enrollment adjustment process.
Complete by mid-December	DCPS, with Study Team, creates estimates for special ed/adult/alternative ed schools and PS/PK classrooms using planned programmatic adjustments and enrollment, birth, and housing market trends.
Complete by mid-December	Apply statistical models to create school-level and grade-level enrollment estimates.
Complete by end December	Add PS/PK and special schools to school and district-level projections.
1st week of January	Provide DCPS with grade level estimates in user-friendly electronic format at school, grade, and district level (see sample form, Appendix D) DCPS Reviews grade level estimates and uses review format to capture and document adjustments for planned programmatic/leadership changes, neighborhood change
2nd week of January	Local schools review estimates, adjustments, and projections and provide comments to DCPS on adjustments. DCPS and local schools determine enrollment targets, where appropriate.
January 15, 2010	Enrollment projection process is complete.

Appendix A: Detailed Student Enrollment Projection Methodology

The purpose of this appendix memo is to provide DCPS with a comprehensive document that describes in detail the methodology and results for the enrollment model we used to predict student enrollment for the 2009-10 school year. Some information described in this appendix is also included in the School Enrollment and Projections memo but this appendix provides more detail overall.

Data Modeling for School Level Grade-By-Grade Enrollment Estimates

The Study Team scanned the literature searching for other school districts' methods of student enrollment projection. Most of the literature was about district-level predictions and the majority used a cohort-survival method. We found little literature on individual school enrollment projections and no examples of school districts predicting at the grade level. For those that used school-level predictions, most tweaked their district-level model, primarily the cohort-survival method, to include the smaller geographic areas. Atlanta, Georgia was one school district that projected student enrollment at the school level. Some of the academic literature discussed how the sample error is much larger when making school level versus district level predictions complicating the enrollment predictions.

We pursued two statistical models to determine which would better estimate DCPS enrollment for 2009-10: a statistically-advanced, computationally-complex conditional multinomial logit model (individual student choice model) and a straightforward but nuanced linear regression model (cohort survival model).

Individual student-choice model. Since we had data for every public school student (DCPS and public charter) – their demographic characteristics and home address, as well as where they attended school and the school's characteristics and location – we used a modeling method a discrete choice model known commonly as a "conditional multinomial logit" (CML). CMLs are commonly used by economists to estimate consumer choices in the market place, such as which hospital people will choose to use when a new hospital opens. This sophisticated modeling takes into account individual characteristics of the student (such as special education status, race, free/reduced price lunch) as well as school characteristics (distance between each school and student, average test scores) and estimates the probabilities of students choosing each available school. For the CML, the probability that student *i* chooses school *j* at time *t* is written as:

$$\Pr(Y_{it} = j) = \exp(\beta' z_{ijt} + \alpha_j' y_{it}) / \sum_{k=1}^{J_t} \exp(\beta' z_{ikt} + \alpha_k' y_{it})$$

where y_{it} denotes characteristics specific to student *i* at time *t*, and z_{ijt} denotes characteristics for school *j* at time *t* for student *i*.

While the CML model is commonly used in economic literature to predict consumer choices and can be a robust method for predicting student choices, we did not find other school districts that used CML modeling. One reason that other school districts have not used this model is that most

school districts do not have the breadth of choice that the District does. For most school districts, students are assigned to a public school and they either attend that school, choose to enroll in private school, or move. Other school districts such as Boston and Chicago have public charter schools but not as extensive a public charter system as Washington, D.C.'s.

There were challenges in compiling the data for the CLM model and running the model due to the extremely large matrix that the statistical package must run. For instance, we prepared a data file for every grade level that included one observation for each student for every possible school (DCPS or public charter) could have chosen in that grade. A "choice" flag indicated which school the student actually chose to attend. Each grade-level dataset included the student-level choice school and every other possible school for school years 2005-06, 2006-07 and 2008-09.

We split the data into individual grade level files for two reasons. First, practically speaking, the individual grade-level files were quite large: 100-500 MB for each grade. Second, we hoped to be able to identify the key school characteristics that influence choice at each grade level, assuming that different factors will come into play in elementary school than in high school, for example.

We first began using the statistical package Stata to run the model but ran into problems because 1) Stata imposes file size limits and had difficulty processing the large data files (we were forced to run individual grade level files over night) and 2) Stata automatically included school dummies into the model that created a multicollinearity problem stalling the processing and limited the model's ability to determine the effect of school characteristics.

Therefore, we switched to SAS to run the CML models. Unlike Stata, SAS does not add school dummies to the conditional logit model. This gave us hope that we could identify school characteristics that factor into students making their decision about what school to attend. An additional benefit was that SAS processed the models much more quickly, and without the file size limits that Stata imposes. On the downside, the SAS procedure does not permit us to include student characteristics in the model, such as race or whether a particular school is within a student's school boundary. Therefore, the CML model that includes just school characteristics can be used for policy purposes illuminating what school characteristics factor in a student's choice to attend public school.

The final problem with the CML model was that we were missing school characteristic indicators which we hypothesized would be crucial to choosing a school, such as teacher quality or student-teacher ratio. We also had school characteristic data for only some DCPS schools and not public charter schools, such as density factor. If DCPS elects to continue trying to use the CML model in the future, it would be worthwhile investing in improving the quality and access to school characteristics for all public schools.

Results for CML model for the 11th grade. Next we describe the statistical results from the CML model for the 11th grade. We focused primarily on the 11th grade because it had the smallest number of choices as compared to elementary school grades, so the data set was smaller and more manageable to work with.

Three independent variables were consistently statistically significant in virtually every model that included them: distance between student's home and school, percentage of free and reduced-price lunch students at the school, and enrollment at the school in the previous year. This means that students entering 11th grade are less likely to choose schools that are farther away, all other things equal. They are also less likely to choose schools with a larger share of students eligible for free or reduced-price lunch. However, they are slightly more likely to choose schools that had a larger student body the previous year.

While the CML model did not prove to be as fruitful as we had hoped, we think that there is promise in running the CML model with a more robust set of indicators about the characteristics of schools in the future that can be used for programmatic purposes.

Cohort-adjustment model. While we were working on the CML model, we were also simultaneously developing an ordinary least squared (OLS) linear regression model based on the number of students enrolled in the previous grade the previous year (the cohort). This is a method commonly used by other school districts at the district level and is similar to how DCPS previously projected their enrollment, although we included an additional set of variables to capture whether more or less grade-appropriate DCPS or public charter schools were located near the school being estimated. We tested several variations on this model in an effort to find the model that best predicted future enrollment.

The first basic model that we created for the enrollment predictions submitted on January 15, 2009 for the development of school-level budgets by DCPS estimated enrollment in a given grade for a particular school as a function of last year's enrollment in the previous grade plus the change in the number of schools (DCPS or public charter) offering the estimated grade with a half mile, 1 mile, and 2 miles of the school, as well as dummy variables for the grade being predicted and the school being predicted. The unit of observation is a grade-level (G) at a particular DCPS school (S) in a particular year (T). The dependent variable is total enrollment (number of students) in a grade at a school for a given year -- $\text{Enrollment}_{S,G,T.}$

The *basic* model is:

$$\text{Enrollment}_{S,G,T} = \text{function}\{\text{Enrollment}_{S,G-1,T-1} + \text{New Supply Dummy .5 miles}_{S,G,T} + \text{New Supply Dummy 1 mile}_{S,G,T} + \text{New Supply Dummy 2 miles}_{S,G,T} + \text{Grade Dummies}_{T} + \text{DCPS School Dummies}_{T} + \text{Constant})$$

We ran the basic model separately for three different grade configurations, elementary (1st through 5th grade), middle (6th through 8th grade), and high school (9th through 12th grade), as we believed that each of these grade levels are inherently different.

There were two variations to this basic model for grades in schools that did not have an earlier cohort, typically kindergarten and grades 5, 6, 9. For the kindergarten model, we did not think the cohort variable alone would sufficiently predict future kindergarten enrollments, as pre-kindergarten is not mandatory in the District. Therefore, we also included the number of children born five years earlier in each of the 39 D.C. Department of Planning's neighborhood clusters.

We also included the number of schools offering kindergarten within half mile. The variation on the basic model for kindergarten was:

$$\text{Enrollment}_{S,G,T} = \text{function}\{\text{Enrollment}_{S,G-1,T-1} + \text{Births}_{T-5} + \text{Supply Dummy .5 miles}_{S,G,T} + \text{Grade Dummies}_{T} + \text{DCPS School Dummies}_{T} + \text{Constant}\}$$

To estimate the grades at schools for which there was no previous grade or cohort enrollment (typically grades 5, 6, and 9), we substituted the number of students who were enrolled citywide in DCPS schools ($\text{Enrollment Citywide}_{G-1,T-1}$) from the previous grade the year before. We referred to the model for estimating starting grades with no available earlier cohort as the *citywide* model:

$$\text{Enrollment}_{S,G,T} = \text{function}\{\text{Enrollment Citywide}_{G-1,T-1} + \text{New Supply Dummy .5 miles}_{S,G,T} + \text{New Supply Dummy 1 mile}_{S,G,T} + \text{New Supply Dummy 2 miles}_{S,G,T} + \text{Grade Dummies}_{T} + \text{DCPS School Dummies}_{T} + \text{Constant}\}$$

Special programs. We did not run the basic model or create a new model to predict enrollment for non-mandatory grades or special DCPS programs/schools, such as pre-school and pre-kindergarten, alternative education schools, or special education schools. The size of the early education grades and special programs is typically determined by DCPS administrators.

During our review session with DCPS on January 14, we provided DCPS with a matrix of pre-school and pre-kindergarten enrollment over time for each school matched to the birth and housing market trends in the corresponding neighborhood cluster. DCPS, along with special education specialists, then determined for each school whether pre-school and pre-kindergarten classes would remain the same size as the year before, decline, or increase for every DCPS school.

Data. In order to run the basic cohort-adjustment model and its two variations (kindergarten and citywide), we used two sources of data, student-level data and school-level data. We used six years of un-audited October count student-level data from DCPS' STARS system (SY2003-04 through SY2008-09) to determine the number of DCPS students enrolled by grade and school over time. We had previously cleaned 2003-04 through 2006-07 student-level data for the Quality Schools and Healthy Neighborhoods research project funded by OSSE. On November 12, 2008, we received the SY2007-08 and SY2008-09 un-audited October count student-level data from DCPS. We completed the cleaning and processing of the data well enough that we could begin building our models by early December, although we continued to tweak the data up until early January. It took approximately 200 hours for two research assistants to clean and process the student-level data.

We also compiled a DCPS and public charter school-level dataset that included every school (DCPS or public charter), its location (geocoded with longitude and latitude coordinates), and minimum and maximum grades for SY2003-04 through 2008-09. Other school-level characteristics were provided by 21st Century School Fund, such as budget per school and density factor, but these are for use in an expanded CML model. We then calculated the supply

of grade-appropriate schools (DCPS and public charter) within a half mile, 1 mile, and 2 miles of every DCPS school for every year (taking into account the schools that changed the grades they offered). We then calculated the change in the number of schools offering particular grades between every two-year spread. (These variables were included in the basic model.)

To run the cohort-adjustment model, we then compiled a panel dataset from the student-level and school-level file that included an observation for every DCPS school at every grade with the necessary dependent and independent variables (current enrollment, cohort enrollment, change in school supply at the three distances, and school and grade dummy variables). There was a total of 4,799 observations in the dataset that was used to run the cohort basic model.

Testing the cohort model results. The cohort-adjustment model did a good job in predicting DCPS enrollment. We had few data problems c (i.e., missing observations) and R^2 was high for the basic model and two variations showing that the model did a good job fitting the historic data to actual estimated outcomes.

In order to determine whether the basic cohort and citywide models would sufficiently predict the 2009-10 school year enrollment, we tested the models to estimate enrollment for the 2007-08 school year and compared it to the actual 2007-08 enrollment. To do this estimation, we included all years of student-level data in the model, SY2003-04 through SY2008-09. We included SY2007-08 data to test the "goodness of fit" of the model. Tables 1 and 2 summarize our model's success in predicting enrollment for the 2007-08 school year. (Note: The columns, share within 5 students, 8 students, 20 students and 30 students, is an average of every grade in every school that falls into each grade level.)

Table A-1. Success Rate of Basic Cohort and Citywide Estimations for SY2007-08 By Grade Categories (January 15, 2009 model)

Grade Level	Share within 5 students, 2007	Share within 8 students, 2007	Share where 2007 estimation greater than actual 2007
Elementary Schools	65.0%	86.8%	55.4%
Middle Schools	23.9%	40.3%	49.3%

Grade Level	Share within 20 students, 2007	Share within 30 students, 2007	Share where 2007 Prediction greater than Actual 2007
High Schools	59.7%	74.6%	35.8%

Table A-2. Success Rate of Basic Cohort and Citywide Estimations for SY2007-08 By Individual Grades (January 15, 2009 model)

Grade Predicting	Share within 5 students, 2007	Share within 8 students, 2007	Share where 2007 estimation greater than actual 2007
Kindergarten	65.0%	85.0%	51.3%
1st	65.4%	92.3%	59.0%
2nd	74.7%	89.9%	54.4%
3rd	60.8%	83.5%	58.2%
4th	62.8%	89.7%	48.7%
5th	61.3%	80.0%	61.3%
6th	34.5%	44.8%	44.8%
7th	10.5%	21.1%	63.2%
8th	21.1%	52.6%	42.1%

Grade Predicting	Share within 20 students, 2007	Share within 30 students, 2007	Share where 2007 estimation greater than actual 2007
9th	56.3%	56.3%	50.0%
10th	58.8%	82.4%	17.6%
11th	82.4%	88.2%	47.1%
12th	41.2%	70.6%	29.4%

We found that, on average, the model adequately estimated the 2007-08 enrollment for the elementary school grades – 65 percent of the predictions for the elementary school grades were within 5 students and 87 percent were within 8 students. We were particularly successful in grades K, 1, and 2. We were less successful in predicting the high school grades. The basic

models predicted within 20 students 60 percent of the time for high school grades and 75 percent were within 30 students.

The model did very poorly when predicting for the middle school grades: only 24 percent of the predictions were within 5 students and 40 percent were within 8 students. Clearly the middle school grades are the most problematic to predict. This is understandable for several reasons: 1) a greater share of public school students enroll in public charters at the middle school grades than at other grade levels; 2) there have been several reorganizations of the DCPS middle and junior highs in recent years, including the movement of 6th graders from elementary to middle schools; and 3) some of the middle schools have very small 6th grades but much larger 7th grades, which makes the predictions difficult.

Improvements to the model. It was clear that the basic cohort and citywide models submitted for the January 15, 2009 deadline were not predicting with sufficient accuracy. Part of the problem was that we did not have enough time from the date we received the 2007-08 and 2008-09 student data to fully vet the model and how it predicted the 2007-08 enrollment before providing the 2009-10 enrollment projections. (In addition, there were anomalies in the data that had not yet been resolved, such as new schools that needed to be calculated, such as Phelps Academy, and the combination of Bell and Lincoln into Columbia Heights Education Campus.)

Between the end of January and the beginning of April we worked to improve upon the basic cohort and citywide models. For instance, for the citywide high school model, we tested whether using the enrollment from the feeder middle schools was more successful in predicting 2007-08 than using the citywide enrollment variable. However, the feeder pattern enrollment did not improve our predictions. We also tried predicting total citywide DCPS enrollment for 2009-10 using our basic model with the grade history variable (described below); however, it was clear that this model also over-predicted total enrollment.

We did successfully add three additional variables to the basic cohort and citywide models.

- First, we included the number of students enrolled in the same grade the previous year (or grade history). We included this variable to attempt to control for the model's tendency to over predict, as well address the bumps and spikes in enrollment between certain grades, such as the 9th grade retention bump and the enrollment spike from 6th to 7th grade. This additional variable marginally improved our 2007-08 enrollment predictions.

- Upon further analysis for the basic model, we discovered that the basic model predicted better for some wards rather than others. For instance, the basic model successfully predicted the 2007-08 enrollment within 5 students for over 60 percent of the schools in Wards 2, 3, and 8 as compared to Ward 6 where the model only successfully predicted for 46 percent of the schools.

Table A-3. Success Rate of Basic Cohort and Citywide Predictions for SY2007-08 By Ward (January 15, 2009 model)

Ward (2002)	Share within 5 students, 2007	Share within 8 students, 2007	Share within 20 students, 2007	Share within 30 students, 2007	Share where 2007 Prediction greater than Actual 2007
Ward 1	50.0%	76.1%	89.1%	91.3%	52.2%
Ward 2	63.6%	72.7%	90.9%	97.7%	45.5%
Ward 3	63.0%	85.2%	96.3%	96.3%	35.2%
Ward 4	56.6%	81.9%	96.4%	98.8%	51.8%
Ward 5	52.5%	71.3%	91.3%	96.3%	56.3%
Ward 6	45.5%	71.4%	90.9%	93.5%	49.4%
Ward 7	53.9%	73.0%	93.9%	96.5%	60.0%
Ward 8	60.6%	71.2%	92.3%	93.3%	56.7%

Therefore, we added a new interaction variable between ward and cohort to the model.

- Finally, we included an interaction variable between grade and cohort enrollment. The basic model's cohort variable was an average across all grades. Since we know that the cohort in the earlier grades are more likely to return the following year compared to the middle school grades, we included this new interaction term to tease apart the differences in cohort by grade.

The revised cohort model was the following (the new independent variables are bolded):

$$\text{Enrollment}_{S,G,T} = \text{function}\{\text{Enrollment}_{S,G-1,T-1} + \text{New Supply Dummy .5 miles}_{S,G,T} + \text{New Supply Dummy 1 mile}_{S,G,T} + \text{New Supply Dummy 2 miles}_{S,G,T} + \text{Grade Dummies}_{T} + \text{DCPS School Dummies}_{T} + \mathbf{Enrollment_{S,G,T-1} + Ward_S \times Enrollment_{S,G-1,T-1} + Grade \times Enrollment_{S,G-1,T-1}} + \text{Constant}\}$$

However, it should be noted that we could not apply these three new variables to any school that added a new grade in 2008-09, as there was no historic data to base these new variables on. For instance, the estimate for Langdon's 7th grade, as well as the 7th grade estimates for all the new K-8 schools, is based on the original basic model

Testing the results from the improved models. The inclusion of the three new variables marginally improved the R^2 and improved some of the grade level estimates. For instance, the improved model predicted within 5 students for the 2007-08 school year for 68 percent of the

elementary grades as opposed to 65 percent previously (see Tables 4 and 5). The new model did significantly better for grades K and 1.

The improved model also did better predicting for the middle school grades: the new model successfully predicted within 5 students for the 2007-08 school year for 35 percent of the middle school grades as compared to 24 percent previously. The new model did significantly better for grades 6 and 7, although the overall success rate is still much lower than the elementary school grades.

Upon further discussions, we realized that our enrollment threshold size for the middle school grades was too low. Therefore, we increased the middle school threshold to be within 10 and 15 students. Again the revised model was less accurate at the middle school grades at the higher threshold, only 40 percent of the middle school estimates were within 10 students and 63 percent were within 15 students. The model accurately estimated within 10 students for 40 percent of both 7th and 8th grades and the model estimated slightly better for the 8th grade.

Surprisingly, the predictions for the high school grades worsened with the introduction of the new model. However, the reason that the share of students within 30 students dropped for the new model compared to the January model was that the revised model predicted within 31 students for four of the grades. If we extend the threshold to include 31 students, the model predicted 72 percent of the share of grades/schools for the high school grades. So the latest model predicted slightly worse for the high school grades but not as severely as Table 4 suggests.

Table A-4. Success Rate of Improved Cohort and Citywide Estimations for SY2007-08 By Grade Categories (April 7, 2009 model)

Grade Range	N	Share within 5 students, 2007	Share within 8 students, 2007	Share where 2007 estimation greater than actual 2007
Elementary Schools	530	67.8%	88.6%	54.4%
Middle Schools	134	34.6%	50.6%	48.1%

Grade Range	N	Share within 20 students, 2007	Share within 30 students, 2007	Share where 2007 Prediction greater than Actual 2007
High Schools	92	50.0%	65.3%	44.4%

Table A-5. Success Rate of Improved Cohort and Citywide Estimations for SY2007-08 By Individual Grades (April 7, 2009 model)

Grade Predicting	N	Share within 5 students, 2007	Share within 8 students, 2007	Share where 2007 estimation greater than actual 2007
Kindergarten	87	75.3%	92.6%	54.3%
1st	88	72.5%	91.3%	57.5%
2nd	89	71.6%	91.4%	54.3%
3rd	89	60.5%	84.0%	55.6%
4th	89	65.0%	91.3%	46.3%
5th	88	62.0%	81.0%	58.2%
6th	57	46.3%	68.3%	46.3%
7th	39	20.0%	35.0%	45.0%
8th	38	25.0%	30.0%	55.0%

Grade Predicting	N	Share within 20 students, 2007	Share within 30 students, 2007	Share where 2007 estimation greater than actual 2007
9th	29	27.8%	44.4%	55.6%
10th	25	66.7%	77.8%	50.0%
11th	22	50.0%	72.2%	33.3%
12th	21	55.6%	66.7%	38.9%

Adding the ward and cohort interaction improved the model's prediction for some of the wards. For instance, the 2007-08 estimations for Ward 2 increased from 64 percent within 5 students to 71 percent and increased Ward 8's predictions from 61 percent to 67 percent. However, the new model also reduced the success of some of the wards, such as Ward 1 and Ward 3.

Table A-6. Success Rate of Improved Cohort and Citywide Estimations for SY2007-08 By Ward (April 7, 2009 model)

Ward (2002)	Share within 5 students, 2007	Share within 8 students, 2007	Share within 20 students, 2007	Share within 30 students, 2007	Share where 2007 estimation greater than actual 2007
Ward 1	41.5%	66.0%	88.7%	94.3%	54.7%
Ward 2	70.5%	84.1%	93.2%	95.5%	43.2%
Ward 3	61.4%	80.7%	93.0%	94.7%	38.6%
Ward 4	62.5%	81.8%	92.0%	94.3%	48.9%
Ward 5	52.5%	67.5%	90.0%	95.0%	48.8%
Ward 6	50.0%	75.0%	92.1%	96.1%	52.6%
Ward 7	57.8%	76.7%	91.4%	95.7%	62.1%
Ward 8	67.0%	79.2%	91.5%	93.4%	51.9%

Predicting 2009-10 enrollment. Once the final model was developed, we then calculated the 2009-10 enrollment projections. To calculate the enrollment projections for each grade at each school, we took the revised model (either cohort or citywide depending on the grade) and plugged in the corresponding enrollment data and school capacity data for each school and grade for 2009-10. We had calculated the change in school supply for 2009-10 (the independent school supply variable) and since the model uses the previous year's cohort and grade history enrollment data, we plugged in the 2008-09 enrollment data. We then used the model's parameters to predict enrollment for 2009-10.

Because the model is a linear regression ($y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_i + \epsilon_i$ where y is the dependent variable [or the grade/school being predicted], β is the model parameter, x is the independent variable, and ϵ is the error term), we then multiplied the 2009-10 independent variable data (that differs for each grade at each school) by the model's corresponding parameter and added the products together resulting in the grade/school specific 2009-10 projection.

We could not use the most recent revised model to predict enrollment for new grades or new schools added in 2009-10 because there was no 2008-09 enrollment data for the same grade in the previous year. Therefore, we used the basic January model to forecast for these new grades and schools.

Appendix C shows the predicted 2009-10 enrollment for every school and grade, excluding special education and alternative education schools.

Further improvements for 2010-11 enrollment predictions

The Study Team has made significant progress in developing the statistical student enrollment projections. We recommend making the following revisions to the modeling process for the 2010-11 school year:

- To get a better sense on how the model estimates (or get a "goodness of fit"), we need to run the model estimating for additional years in addition to 2007-08. We should run the model to estimate SY2006-07 and SY2008-09. These estimations would include all years of data in the model, including the year that we are estimating.

- In addition, we should calculate predictions for multiple years beyond just 2009-10. (This would include all data in the model except for the year being predicted.) It would be helpful to see how accurate our predictions for SY2008-09 and SY2007-08 would be next to the actual historic enrollment.

- It may be that having a school supply variable based on location and whether a new grade is offered is not sufficient. The Study Team should explore including more robust measures of new school supply. To start with, we could explore measuring the capacity of a school -- how many seats or how many classrooms offered.

- Our outside econometric consultant recommended that we use a time series approach to adjust for our grade level over prediction. We could estimate a single time series of overall DCPS enrollment using traditional, straightforward statistical time series methods. In addition to estimating a total DCPS enrollment, we would use the difference between the most recent enrollment and the forecasted future enrollment to determine the percent change in total enrollment over time. This percent change could then be applied to the individual grade or school projections to weight them down, if the basic grade-level school-level model continues to over predict. It's not clear if this would improve our accuracy but is worth trying.

- In order to improve the student choice model, we need to further explore the CML procedures being used. There is most likely a way to overcome the modeling package's resistance to including individual student characteristics but it will take discussions with other econometricians.