Evaluating the Impact of Charter Schools on Student Achievement: A Longitudinal Look at the Great Lakes States

Appendix A

June 2007

EPRU | EDUCATION POLICY RESEARCH UNIT

Education Policy Research Unit
Division of Educational Leadership and Policy Studies
College of Education, Arizona State University
P.O. Box 872411, Tempe, AZ 85287-2411
Telephone: (480) 965-1886
Fax: (480) 965-0303
E-mail: epsl@asu.edu
http://edpolicylab.org

Education and the Public Interest Center
School of Education,
University of Colorado
Boulder, CO 80309-0249
Telephone: (303) 492-8370
Fax: (303) 492-7090
Email: epic@colorado.edu
http://education.colorado.edu/epic

- The policy brief is available online at: http://epsl.asu.edu/epru/documents/EPSL-0706-236-EPRU.pdf

This research was made possible by funding from the Great Lakes Center for Education Research and Practice

Appendix A
Methodology

Overview

This evaluation used a nonexperimental, longitudinal, and cross-sectional design to compare student math and reading achievement in charter and public schools in six Great Lakes states (Illinois, Indiana, Michigan, Minnesota, Ohio, and Wisconsin) over a five-year period. The time period was the most recent five years for which data were available for each of the states (typically the 2001-02 to 2005-06 academic years). The unit of analysis was individual schools in each state. One grade at each of the three school levels (i.e., elementary, middle, & high school) was selected for analysis (e.g., grades 4, 7, and 10). Ex post facto data were obtained from the National Center for Education Statistics Common Core of Data Web site and from each state's department of education Web site. Independent predictor variables were percentage minority, percentage low income (free/reduced lunch), percentage special education, percentage limited English proficiency, and urbanicity; and the dependent variable was achievement results on state assessment tests. Linear regression models were used to conduct residual gains analyses on school-level data that produced three estimates: (1) actual scores based on observed student achievement data provided by each school; (2) predicted scores based on an aggregate of actual scores for demographically similar public schools; and (3) residual scores, which identified the difference between charter school actual and predicted student achievement. Patterns of growth/decline were then analyzed over time. Two evaluation questions were used in this study: (1) How does student achievement in charter schools compare to demographically similar public schools? (2) Are charter schools an effective strategy for improving student achievement over time?

Scope and Design

This evaluation focuses exclusively on student achievement. Brief descriptions of charter school reform for the schools included in this study are provided. These descriptions contain general details regarding the age and relative size of the reforms as well as comments regarding whether the reforms are restrictive or permissive with regard to autonomy. Comments on the rigor of oversight are also included. An in-depth analysis of the differences among schools or states is beyond the scope of this current evaluation. Table 1 presents decision criteria and rationale regarding the scope and focus of the evaluation. Trade-offs and compromises always need to be made when narrowing a study. Though data collection and analysis challenges varied considerably by state, the intent of the criteria was to ensure that the study was as structured and systematic as possible.

Table 1. Decision Criteria and Descriptions

Topic	*Decision Criteria, Description, Rationale*
States Included	This evaluation was sponsored by the Great Lakes Center, so the decision to focus on these states is based on their location in the Great Lakes region (i.e., Illinois, Indiana, Michigan, Minnesota, Ohio, and Wisconsin)
Test/ Assessment	Only state assessments were used since all public schools, including charter schools, must participate in these assessments. In some states, other standardized tests also are administered,[1] but these typically include only a sample of schools or students. The state assessments are commonly viewed as high stakes tests, and they are familiar to a broad range of stakeholders.
Outcome Measure	Preference was given to the most sensitive test/assessment measure available in the following order: normal curve equivalent, percentile rank, scaled score (mean achievement test score for a school), and cut score (mean percentage of students meeting or exceeding state standards). All states in this study had cut score data. Additionally, Wisconsin and Indiana had scale score data equated from year to year over the previous 5 years. These were used instead of the cut scores, given their increased sensitivity to change over time.
Test Content	Math and reading tests were selected because these subjects had the best longitudinal data and typically comprise the high stakes component of state assessment programs.
Grade Levels	One grade at each of the three school levels (i.e., elementary, middle, & high) was selected for this study. Preference was given to the highest grade with longitudinal data at the elementary level.[2]
Years	Trends were analyzed over a five-year period, with preference given to the five most recent years in which data were available. For most states, this meant tracking data from 2001-02 to 2005-06. In Michigan and Indiana, we were able to obtain 2006-07 data.

Data Sources

Data including district and school name, district and school number, school enrollment, ethnicity, free and reduced lunch (low income), urbanicity (locale), and a charter school identifier were obtained from the National Center for Education Statistics Common Core of Data (CCD).[3] Data including district and school name, district and school number, special education enrollment, limited English proficiency enrollment, number of students tested, and achievement test scores were obtained from each state's department of education Web site.

Data Set Construction

Longitudinal data sets containing demographic and student achievement data for all charter and traditional public schools needed to be constructed for each of the six states. Data collection and preparation were conducted by several graduate students and research staff over the course of several months. Complications encountered included (1) extremely large and difficult to manage data files (many exceeding Microsoft Excel's capacity); (2) inconsistent data formatting, including changes in data structures and variable names from year to year; and (3) inconsistent and/or missing building or school codes, which were necessary for matching and merging data. In some cases, we needed to create unique identification numbers combining district and school numbers in order to merge CCD data sets to state board of education data sets. Data cleaning involved complex data transformations, recoding, and creating new variables since the source data often did not contain the specific predictor variables required for analysis. The preparation of these data sets literally meant that several gigabytes of files were downloaded from state or federal Web sites. After extracting or creating the variables of interest, cross-sectional data sets for each year in the trend were then merged into six longitudinal data sets (one for each state) that could be managed by common desktop computers and software. Although the process started with several gigabytes of information for each state, the final flat files were typically less than 10 megabytes in size.

Variables

Independent variables included percentage minority (MINORITY), percentage low income (LOW INCOME), percentage special education (SPED), percentage limited English proficiency (LEP), and population density where school is located (URBANICITY) for each school. The dependent variable was achievement test results (TEST) for each school. A moderator variable identifying charter school status was used to distinguish charter schools in the state. See Table 2 for study variables, codes, and operational definitions.

Table 2. Independent and Dependent Study Variables

Variable	***Operational Definition***
Minority	Percentage of students in each school who are American Indian/Alaska Native, Hispanic, Black, or mixed ethnicity (White and Asian/Pacific Islander students were intentionally excluded. Although Asian-American students are considered part of a minority group, they typically are the ethnic group that performs highest on standardized tests, followed by White or European Americans).
Low Income	Percentage of low income students in each school who are eligible to receive free or reduced lunch.
Special Education (SPED)	Percentage of students in each school who are identified as "special education" students (e.g., have disabilities, receive special education services, have individualized education plans/programs-IEPs)
Limited English Proficiency (LEP)	Percentage of students in each school with limited English proficiency

Variable	Operational Definition
Urbanicity (Locale)	8-category urbanicity rating for each school based on the community's population density: 1. **Large city**: A principal city of a metropolitan core based statistical area (CBSA), with the city having a population greater than or equal to 250,000 2. **Midsize city**: A principal city of a metropolitan CBSA, with the city having a population less than 250,000 3. **Urban fringe of a large city**: Any incorporated place, census-designated place, or nonplace territory within a metropolitan CBSA of a large city and defined as urban by the census bureau 4. **Urban fringe of a midsize city**: Any incorporated place, census-designated place, or nonplace territory within a CBSA of a midsize city and defined as urban by the census bureau 5. **Large town**: An incorporated place or census-designated place with a population greater than or equal to 25,000 and located outside a metropolitan CBSA or inside a micropolitan CBSA 6. **Small town**: An incorporated place or census-designated place with a population less than 25,000 and greater than or equal to 2,500 and located outside a metropolitan CBSA or inside a micropolitan CBSA 7. **Rural, outside CBSA**: Any incorporated place, census-designated place, or nonplace territory not within a metropolitan CBSA or within a micropolitan CBSA and defined as rural by the Census Bureau 8. **Rural, inside CBSA**: Any incorporated place, census-designated place, or nonplace territory within a metropolitan CBSA and defined as rural by the Census Bureau
Achievement Test (TEST)	The dependent variables for the analyses are state achievement test results for each school. Order of preference in selecting test score data was based on the sensitivity of the measure: scaled scores (mean achievement test score for a school) were preferred and used over cut scores (mean percentage of students meeting or exceeding standards) when possible.

Residual Gains Analysis

Analyses based on changes within individual students are more rigorous and desirable than analyses based on school-level data; however, only school-level data were readily available for states in this study. This is common in the country as a whole. Some states that have student level data sets are restricted from sharing this data with researchers. The few states that have student level data that could be used for evaluating the impact of charter schools on student achievement include Arizona,[4] Delaware,[5] Florida,[6] North Carolina,[7] and Texas.[8] Several more years will be needed before new value-added assessment systems can provide student level data for longitudinal designs. Thus, residual gains analysis was selected because it provides one of the most rigorous designs and methodological approaches suitable for analyzing group or school-level student achievement data. This approach has been used successfully for a number of state evaluations when individual student data were not available. In a recent evaluation of the Delaware charter school reform, a quasi-experimental design based on student-level data was used as well as a longitudinal residual gains analysis based on school-level data. This provided a unique opportunity to compare results from the best possible analyses of student-level data with the best possible analyses of school-level data. Results from the two approaches revealed nearly identical findings.[9]

Linear regression models were used to conduct residual gains analyses on school-level data to estimate growth/decline patterns of student achievement. To facilitate estimating these patterns, school-level performance needed to be tracked across time. Regression models were fit to each test content area (math or reading) and each grade level included in the study. Approximately 30 separate regression models were tested for each state. Independent variables used as predictors in the models included percentage of minority students (MINORITY), percentage of low-income students receiving free or reduced lunch (LOW INCOME), percentage of students qualifying for special education (SPED), percentage of students with limited English proficiency (LEP), and each school's urbanicity rating based on its community's population density (URBANICITY) (see Table 2). While these variables represent the desired set of predictor variables, substantial difficulty was encountered when obtaining school-level SPED and LEP data in a few of the states, either because these data were not available for 1 or more years or because these data were available only at district but not school levels.

From these regression models, three estimates were produced: (1) actual, or observed scores; (2) predicted, or expected scores; and (3) residual, or difference in scores. The actual performance scores presented in the tables for each state represent the cut scores or scaled scores reported by schools for a given grade and test content area in a given year.

Predicted scores are those that were anticipated in comparison with public schools (charter and noncharter) for a given grade in a given year. In other words, the predicted scores represent how a charter school is expected to score based on how demographically similar public schools perform. The predicted values were determined using an ordinary least squares (OLS) multiple regression procedure, in the form of the linear equation given in Equation 1

$$\hat{Y}_i = a + b_1 MINORITY_i + b_2 LOWINCOME_i + b_3 SPED_i + b_4 LEP_i + b_5 URBANICITY_i + \varepsilon_i \qquad (1)$$

where $\hat{Y}$ is the predicted value for a given school i, expressed in terms of the constant a of the intercept term; $MINORITY_i$ is the proportion of minority students (does not include White or Asian-American students) for a given school i; $LOWINCOME_i$ is the proportion of students receiving free or reduced lunch for a given school i; $SPED_i$ is the proportion of special education students for a given school i; LEP_i is the proportion of students who qualify for limited English proficiency accommodations for a given school i; $URBANICITY_i$ is the degree of urbanicity or population density for a given school i; and the error term ε_i. In this equation, the regression coefficients (bs) represent the independent contributions of each independent variable to the prediction of the dependent variable $\hat{Y}$.

Residual values e_i are simply the difference between the observed value Y_i and the fitted value (predicted) $\hat{Y}_i$ for given school i as shown in Equation 2. These residuals, or differences, indicate whether a school (or group of schools) is performing at, above, or below other demographically similar schools. A residual of 0 indicates that the school performs at the average of all other similar schools. A negative residual means the charter school is performing lower than predicted, and a positive residual indicates it is performing higher than predicted.

$$e_i = Y_i - \hat{Y}_i \qquad (2)$$

To obtain the observed, predicted, and residual scores aggregated for each state, a weighted mean was calculated for the observed value Y_i and the fitted value (predicted) $\hat{Y}_i$ from the school-level Y_is and $\hat{Y}_i$ s, from which state aggregate e_i is calculated. The weighted mean is determined simply by multiplying each school-level Y_i and $\hat{Y}_i$ by the number of test takers within each school. To obtain the state-level Y_is and $\hat{Y}_i$s, the school-level weighted means are averaged and divided by the number of schools in the state. In other words, the average across all the charter schools takes the number of students within those schools into account (for any given grade and any given year).

Average annual change (AAC) scores were computed for patterns of observed, predicted, and residual scores across time by subtracting the first score from the most recent and dividing by the number of observations (e.g., years) minus 1 (i.e., N-1). An example of the procedure for the average annual change in residual scores is shown in Equation 3.

$$\text{AAC} = (e_{2005} - e_{2001})/N\text{-}1 \qquad (3)$$

Typically, the predictor variables in the regression equations accounted for 45 to 65 percent (adjusted R^2 = .45-.65) of the variability in school-level outcome measures (e.g., scaled scores, cut scores), which suggests that these models were rather strong in terms of predicting school performance with a limited number of background indicators.

The focus of this report is on the aggregate results across all charter schools. At a later point in time, we will make available additional appendices with school level results for each charter school in the participating states that have at least two valid points of data on any particular test.

The discussion of methods in this report has been kept brief and relatively nontechnical. Readers interested in a more detailed exposition of methods may contact the authors or refer to our state evaluations of charter schools in Pennsylvania and Delaware where we provide further details and insights regarding the application of this methodology.

Limitations

Below, we summarize the key limitations of the evaluation:

1. This study is based on school-level rather than student-level analyses; thus, fully controlling for student mobility or identifying differences within schools was not possible.

2. Analyses were conducted on consecutive cohorts of students in identical grades (4th graders in 2003, 4th graders in 2004, 4th graders in 2005); therefore, each cohort group had different students. Data were not available to track the same cohorts of students as they progressed through grades (for example, from grade 4 in 2003 to grade 5 in 2004) because most states did not have tests in consecutive grades until very recently when NCLB mandated testing in grades 3-8.

3. The quality (sensitivity) of measures of student achievement varied by state, with all states reporting cut scores but only a few reporting mean scaled scores. Thus, the criterion variable in the regression equations varies by state.
4. Charter schools with missing or incomplete data were dropped from analyses. The most common explanation for missing data was that specific charter schools had too few test takers. (One of the most common measures to ensure the confidentiality of findings is to report performance results only when there are 10 or more test takers; in some states this threshold was as low as 5). The results from Ohio were particularly affected by incomplete data. Although Ohio has the most charter schools in the region, this state had the highest proportion of schools dropped from the analysis due to incomplete data (see Appendix F for more details).
5. Data on special education and limited English proficiency were not available in some states at the school level. Data on special education and limited English proficiency was not available in some states at the school level. Even when we could control for the percentage of special education students, we could not control for differences in the nature and degree of severity of disabilities. Our state evaluations revealed that charter schools have—on average—a substantially lower proportion of students with disabilities and the students with disabilities that enroll in charter schools tend to have less severe and less-costly to remediate disabilities.[10]

While the longitudinal design, broad scope, and overall quality of this study make it one of the most rigorous and comprehensive evaluations of charter school student achievement, these limitations should be considered when interpreting results.

Notes and References

[1] For example, the NAEP, college entrance examinations, or tests developed and administered for largely diagnostic purposes. The perceived importance of these other tests is negligible and varies by schools since they are not high stakes tests.

[2] Each state's accountability system has relied on a high stakes test at 3 or 4 grade levels over the past 7-10 years. More recently and in response to the requirements of NCLB, states have been adding high stakes tests at more grade levels until now when they all are testing at grades 3-8 as well as 1 or 2 high school grade levels. For our analysis it was important to follow relative progress over time, so we sought to include only grade level tests that could be tracked over 5 consecutive years.

[3] Retrieved [February 27, 2007] from the Web site for the Common Core of Data: http://nces.ed.gov/ccd/.

[4] Garcia, D. R., McIlroy, L., & Barber, R. (2007). Starting behind: A comparative analysis of the academic standing of students entering charter schools. *Social Science Quarterly*.

[5] Miron, G., Cullen, A., Applegate, B., & Farrell, P. (2007). *Evaluation of the Delaware charter school reform: Final report*. Dover: Delaware State Board of Education.

[6] Sass, T. R. (2006, Winter). Charter schools and student achievement in Florida. *Education Finance and Policy*, 91-122.

[7] Bifulco, R., & Ladd, H. F. (2004). *The impact of charter schools on student achievement: Evidence from North Carolina*. Chapel Hill, NC: Terry Sanford Institute of Policy. Working Papers Series SAN04-01. http://www.pubpol.duke.edu/people/faculty/ladd/SAN04-01.pdf

[8] Gronberg, T., & Jansen, D. (2005). *Texas charter schools: An assessment in 2005*. Austin: Texas Public Policy Foundation.

[9] Miron, G., Cullen, A., Applegate, B., & Farrell, P. (2007). *Evaluation of the Delaware charter school reform: Final report*. Dover: Delaware State Board of Education

[10] Miron, G., & Nelson, C. (2002). *What's public about charter schools: Lessons learned about choice and accountability*. Thousand Oaks, CA: Corwin Press.