KINGSBURY RESEARCH CENTER

An Application of Propensity Score Stratification Using Multilevel Models

Do Charter Schools Make a Difference in Student Achievement and Growth?

Yun Xiang
Northwest Evaluation Association

Shudong Wang
Northwest Evaluation Association

Paper presented at the annual meeting of the National Council on Measurement in Education. April 28-30, 2013, San Francisco, CA.

Send correspondence to:
Yun Xiang
Northwest Evaluation Association (NWEA)
121 NW Everett St.
Portland, OR 97209
Yun.Xiang@NWEA.org
(503) 915-6118

Abstract

This study explores the question of how to reduce bias and improve the accuracy of results when a randomized experimental design is rarely feasible in educational settings. It demonstrates how the method of propensity score stratification in the multilevel model can be applied to reduce the bias from confounding variables. We investigate the long-debated question of whether going to charter school positively impacts student learning. In particular, we evaluate the impact of transferring from public schools to charter schools on student achievement and growth in multiple states. To balance key covariates between the treatment and control groups of students as well as taking into account undefined factors that vary across schools due to school membership, we first estimate propensity score by using multilevel logistic regression models, and then use the propensity score stratification to estimate the effect of charter school attendance on student achievement and growth. Our findings showed that by matching students based on their probability of transferring to charter schools ("assignment" before "treatment"), charter school students performed significantly better than public school students at a few strata. The pattern, however, was mixed since it did not persist across every strata or every year. The results also showed no significant difference in achievement growth between charter school and public school students. By using propensity score stratification and multilevel models the study challenges the simple conclusion that charter schools have no effect and provides a more complex picture of charter school effect within a quasi-experimental context.

Background

With concerns that U.S. rarely scored on top in international comparison tests such as TIMSS and PIRLES, as well as the fact that United States only ranks 16th worldwide in the proportion of its citizens ages 25 to 34 with degrees[1], many sense that our schools need to catch up to help kids face the challenge of increasingly intense global competitions. Many policies, programs and interventions aiming to improve teaching and student outcomes have been investigated and, with investment from both public and private sectors, brought into the field to decode what works in the field and what interventions can improve student learning.

[1]According to data from the Organization for Economic Co-operation and Development

Experimental Design and Quasi-Experimental Design

Evidence-based educational research, repeatedly shown and recommended in a wide variety of educational literature and documents, often refers to experimental designs when random assignments are feasible (i.e., assign students to control and treatment conditions). The underlying assumption for advocating randomized experimental designs is that by having two equivalent groups, we are able to find out what factors affect student learning, and hence specific interventions can be designed to improve outcomes (Holland, 1986; Rubin, 2007).

The assumption is straightforward and easy to understand, but it has limitations (Cook, Shadish & Wong, 2008).The first difficulty researchers often encounter when conducting experimental studies is that randomly assigning students is impractical and/or unethical in many situations. Logistical and equity issues constrain feasibility of experimental studies in our educational systems. Rarely can researchers control all the important variables which may, directly and indirectly, influence program effectiveness. Additionally, experimental designs are always conducted in specific settings, while natural environments are much more complex than defined laboratory settings. In general, educational studies are situated in dynamic and constantly changing environments, which challenge the feasibility of true experimental design. These limitations often lead researchers to a quasi-experimental design, which does not require random assignment. Without randomization, however, it's difficult to determine causal relationships in observational studies since we cannot rule out effects of confounding variables (Robson & etc., 2001).

Propensity Score Estimation Using Multilevel Models

Random assignment ensures that both treatment and control groups are equivalent in order to reduce the bias from confounding variables. Techniques can be used to model and factor in the effects of confounding variables, thereby improving the accuracy of results obtained from quasi-experiments. Propensity score is one such technique to reduce selection bias by matching groups based on key covariates (Austin, 2011; Oakes & Johnson, 2006; Pearl, 2009; Rosenbaum & Rubin, 1983a; Stuart, 2008).

The propensity score was introduced by Rosenbaum and Rubin (1983a), where it was defined as "the conditional probability of assignment to a particular treatment given a vector of observed covariates." To ensure the two groups are equivalent, the propensity score, the distribution of key covariates of the treatment group and the control group are the same. The propensity score is often estimated by the logistic regression model, where a set of observed covariates are chosen to predict treatment status (treated or untreated). Using the propensity scores to help draw causal inferences, we need to be aware that propensity scores are probabilistic, not causal indicators (Rosenbaum & Rubin, 1983). Another limitation is that the propensity scores, like many other statistical approaches, cannot control for unobserved confounders (Newson, 2008).

To remove the effects of confounding variables and make the control and treatment groups comparable, four different propensity score techniques are used: propensity score matching (match each treatment unit with one or multiple control units), propensity score stratification (subjects are ranked and stratified into groups based on propensity score), covariate adjustment using the propensity score (using a multivariable regression model) and inverse probability of treatment weighting (IPTW) using the propensity score (Austin & Mamdani, 2006; Austin, 2011; Rosenbaum & Rubin, 1983a, 1984).

An important feature of educational studies is that interventions are always implemented in the hierarchical data structure (students are nested within teachers/classrooms which are nested within schools and districts). Multilevel models have begun to gain popularity by handling the nesting data structure and taking into account random effects across upper-level units (i.e. class, school or district) (Raudenbush & Bryk, 2002).

Using propensity score within a multilevel model to estimate treatment effects has only recently been investigated and researched (Grisword, Localio & Mulrow, 2010; Hong & Yun, 2008; Kim & Seltzer, 2007; Wyse, Keesler & Schneider, 2008). By applying propensity score in multilevel models, the approach addresses the two major concerns of quasi-experimental studies: 1) reducing the selection bias by matching students between the treatment and the control groups based on a set of relevant covariates; 2) reducing the bias by accounting for the random effects across units in a hierarchical data structure. For example, Hong and Raudenbush (2006) estimated kindergarten retention policy by estimating propensity score in a multilevel model and using these estimates to evaluate kindergarten students' reading achievement. In another application, inverse-propensity-score-weighted scores were used to evaluate an instructional treatment effect (Hong & Raudenbush, 2008).

Purpose of the Study

This study explores the question of how to reduce bias and improve the accuracy of results when a randomized experimental design is rarely feasible in educational studies. It demonstrates how to apply multilevel models to estimate propensity score and use propensity score stratification within multilevel models to reduce the bias from confounding variables. We investigated the long-debated question of whether going to charter school has short-term and long-term impacts on student learning. In particular, we evaluated the impact of transferring from public school to charter school has on student achievement and growth in multiple states. Students in this study are nested within the public schools they attended before transferring to a charter school. We first estimate propensity score by using multilevel logistic regression models, then use propensity score stratification to estimate the effects of charter school on student achievement and growth.

The study pursues two research questions:

1) Did transferring from public school to charter school improve student achievement?
2) Did transferring from public school to charter school improve student growth?

Methodology

Data and Sample

We investigated these two research questions by looking into the academic performance and growth of students taking computer-adaptive assessments that use the Rasch score model. One important criterion for conducting a growth study is that the measurement scales remain stable over time. It is that constancy of the MAP (Measures of Academic Progress) scale in NWEA that allows us to measure growth and compare student performance across time. In this study, we focus on student mathematics achievement.

We first selected students who transferred from public school to charter school at grade six. We then track them one year backward to include their previous achievement results and three years forward to understand their growth after transferring. We also track whether they stayed in charter schools after they transferred.

Two cohorts of students were selected for this study. We tracked the first cohort of students through grades 5, 6, 7 and 8 in 2006-07, 2007-08, 2008-09 and 2009-10 and the second cohort of students through the same grades (5, 6, 7 and 8) in 2007-08, 2008-09, 2009-10 and 2010-11. There are 4260 students in Cohort One, among which 535 students transferred to charter schools at grade six and 3725 did not. Cohort One students came from 78 public schools in grade five and in grade six, attended 28 charter schools and 147 public schools. There are 4207 students in Cohort Two, among which 693 students transferred to charter schools at grade six and 3514 did not. Cohort Two students came from 66 public schools in grade five and in grade six, attended 40 charter schools and 136 public schools.

Cohort One and Cohort Two consisted of, respectively, 51% and 50% female students, 41% and 45% minority students, 44% and 44% urban students. Free and reduced lunch (FRL) students in the two cohorts are 40% and 46%, respectively.

Analyses

We will compare student achievement and growth of two groups of students: the treatment group (students who transferred from public schools to charter schools at grade six) and the control group (students who came from the same public schools as the treatment group and attended public schools in grade six). Because of lack of randomization, they are not really treatment and control groups in the randomized trial study context. In this study, the treatment group refers to charter school students, and the control group refers to public school students.

The multilevel logistic regression model is applied to estimate propensity score. Students from the same schools are very likely to share commonalities which affect whether they transfer from public school to charter school; therefore, it's problematic to assume student sample independence within schools. Propensity score estimation in a multilevel model addresses the selection bias from students coming from different schools. By accounting for the school random effect, which can include various observed concerning factors (i.e., students in different schools may have gone to charter schools for various reasons), this approach can more effectively match students by controlling for both school membership (random effect) and school key covariates (fixed effects).

The propensity scores are estimated by using the following two-level logistic model (i refers to the student level and j refers to the school level):

$$\mathrm{P}(T) = \gamma_{00} + \gamma_{10} X_{ij} + u_{0j} + u_{1j} X_{ij} + r_{ij}, \qquad (1)$$

$$u_j \sim N(0, \tau);\; r_{ij} \sim N(0, \sigma^2)$$

where T is treatment dummy variable (public school=0, charter school=1); X_{ij} indicates the key covariates used to estimate propensity score; γ_{10} indicates the estimates at the individual level; u_{1j} indicates the estimates at the school level. We ran the models by applying the GLIMMIX procedure in SAS 9.3.

There are several ways to use estimated propensity score to control for confounding, among which are propensity score matching, propensity score stratification and regression adjustment. In this study, subjects in both treatment and control groups are ranked based on logit of propensity score and then stratified into mutually exclusive groups. Cochran (1968) demonstrated using five groups can remove approximately 90% of the bias from a confounding variable. When the sample size is large, a larger number of groups is recommended in order to yield greater homogeneity among units within each group or stratum (Faries, Leon, Maria Haro & Obenchain, 2010).We divide each cohort of students into eight strata on the basis of logit of propensity score, as recommended by Hong & Randenbush (2006).

To estimate charter school effect on achievement, we compare student achievement at each stratum. The multilevel model is used to estimate charter school effect on achievement growth. The two-level repeated measure models (three-year scores nested within individual students) are applied to compare growth rates between charter school students and public school students. The estimation of the charter school effect on growth is modeled as follows:

$$Y_{ij} = \gamma_{00} + u_j + \hat{\delta} a_{ij} + r_{ij} \qquad (2)$$

$$u_j \sim N(0,\tau);\ r_{ij} \sim N(0,\sigma^2)$$

where $\hat{\delta}$ is the estimated treatment (transferring to charter or not); a_{ij} is the time effect (year=0, 1, 2).

Stratum, as a grouping factor, is then included in the model to compare achievement growth between two groups of students for each propensity score stratum. The estimation of the charter school effect on growth at each stratum is modeled as follows:

$$Y_{ij} = \gamma_0 + u_j + \hat{\delta} T_{ij} + \sum_{g=1}^{7} \gamma_g e_{0gij} T_{ij} + r_{ij} \qquad (3)$$

$$u_j \sim N(0,\tau);\ r_{ij} \sim N(0,\sigma^2)$$

where e_{0gij} , g=1,…7, are dummy indictors for seven of the eight propensity strata; $\hat{\delta}$ is the estimated treatment (transferring to charter school or not).

Results

We first apply propensity score estimation and stratification and demonstrate how we balance the two groups—the charter school students ("treatment") and the public school students ("control"). We then estimate the causal effect of transferring to charter school at grade six by comparing student achievement and student growth between the two groups. The results for the two different cohorts are presented.

To balance the two groups, we used a set of variables to estimate the probability of each student transferring to charter school ("treatment assignment"). The variables are adopted from the two major data sources: the Growth Research Database by Northwest Evaluation Association (NWEA) and National Center for Educational Statistics (NCES) (see Table 1 for details). At the individual student level, we used student demographic background (ethnicity and gender) and student previous achievement (test score at the end of grade five). Our hypothesis is that a student's previous performance impacts whether they transfer to charter school later. At the school level, we used school type (urban, suburban, town and rural), major school characteristics (percent of FRL students and percent of minority students), school resources at the district level (total local revenue in a district, total current expenditures for elementary/secondary education, and payments to charter schools). Meanwhile, we included the variable which shows whether or not a school is in a district with charter schools to control for the factor of whether charter school is a convenient choice for some students. Last, we considered peer effect by including the proportion of students who transferred to charter schools at every school. After controlling for these individual level and school-level variables, the two groups can be closely "equivalent" in probability of transferring to a charter school. Hence, when we compare student achievement and growth, we have more confidence that the causal effect of charter school is not confounded with the factors included in the model.

Propensity Score Stratification (Apply Two-Level Logistic Regression Models)

We used the two-level logistic regression model (students nested within schools) to estimate propensity score. By investigating school random and fixed effects, we hypothesize that which elementary schools students attend impacts whether or not they transfer. In our analysis, school random effect was significant for both cohorts.

Cohort One

For Cohort One, among the variables used to predict student "assignment" (transfer or not), we found these students were more likely to transfer to charter schools at grade six: 1) female students ($p<.05$), 2) students with higher achievement score at the end of grade five ($p<.05$), 3) students in a district with charter schools (compared to districts without charter schools) ($p<.001$), 4) students in a school with a lower percent of minority students ($p<.01$), 5) students in suburban, town or rural schools ($p<.05$), 6) students in a school to which more students tend to transfer ($p<.001$).

In order to ensure that the two groups are "equivalent" in their treatment assignment (transfer or not), we divided the cohort of students into eight strata on the basis of logit of propensity score (e_0), as recommended by Hong & Raudenbush (2006). Please refer to Table 2 to see how the two groups are balanced. The table shows that the two groups are very similar in terms of logit of propensity score at each stratum, besides stratum 8 (e_0 =7), where 168 students who transferred to charter schools do not have a matched group, since the probability to transfer for those students is close to 1.

Cohort Two

For Cohort Two, among all the variables used to predict student "assignment" (transfer or not), we found these students were more likely to transfer to charter schools at grade six: 1) female students ($p<.05$), 2) minority students ($p<.05$), 3) students with higher achievement score at the end of grade five ($p<.01$), 4) students in a district with charter schools (compared to districts without charter schools) ($p<.1$), 5) students in a school with a lower percent of minority students ($p<.01$), 6) students in suburban, town or rural schools ($p<.05$), 7) students in a district with a higher ratio of local revenue to expenditures on elementary/secondary education ($p<.01$), 8) students in a school to which more students tend to transfer ($p<.001$). For this cohort of students, we also found that school random effect was significant ($p<.05$).

Identical to Cohort One, we divided the cohort of students into eight strata on the basis of logit of propensity score. Please refer to Table 3 to see how the two groups are balanced. The table shows that the two groups are very similar in terms of logit of propensity score at each stratum, besides stratum 8 (e_0=7), where 326 students who transferred to charter schools do not have a matched group, since the probability to transfer for those students is close to 1.

Causal Effect on Student Achievement

Cohort One

When we simply compared student achievement for charter school and public school students after one year of transfer, we found no significant difference between the two groups on their achievement at grade six spring term (μ_0=224.08, μ_1=225.16) . However, when we matched the two groups of students by their probability of transferring to charter school, the picture was different. Charter school students performed significantly better than those who did not transfer to charter schools at a few strata (see Table 4), which means that for the few groups of students sharing probabilities of transferring based on the model we defined, the charter school students performed better. The pattern is not consistent across all the strata, and thus we urge caution in drawing the conclusion that transferring to charter school and staying at charter school for one year significantly improved student achievement.

When we track these students who stayed in charter schools for two and three years after their transfer, again, without matching the two groups, the two groups were not significantly different in terms of their achievement (μ_0=230.63, μ_1=231.54 for the second year; μ_0=236.89, μ_1=236.68 for the third year). When we look at the matched groups (strata), we found that a few groups of the charter school students performed significantly better than the public school students who share the same probabilities of transferring based on the model we defined. The achievement differences for these students range from 6 points to 9 points, which is more than one year's growth (see Tables 5 and 6).

Cohort Two

Similar to Cohort One, charter school students and public school students did not differ in their average achievement if we didn't match them by the probability of transferring before they transferred to charter school. Tables 7, 8 and 9 shows that without matching, we found no significant difference in student achievement of the two groups for year one (μ_0 =224.60, μ_1 =224.51), year two (μ_0 =231.44, μ_1 =232.12) and year three (μ_0 =236.83, μ_1 =236.70).

When we looked at each stratum, although the charter school students performed better than the public school students at the end of the first year at most strata, these differences were not statistically significant (see Table 7). Moreover, for stratum one, where students have an extremely low probability of transferring to charter schools, we actually found that students who did not transfer performed significantly better than those who transferred. At the end of the second year, the pattern persists, besides one stratum where charter schools students performed significantly better (about one and a half years of growth). At the end of the groups' third year, the differences widened, ranging from 8 to 10 points, indicating that after staying at charter school for three years, these groups of students performed much better than those who did not transfer.

Summary

Without matching students at each stratum, the two cohorts showed the same picture—there was no significant difference in student achievement during the three years after the transfer. By matching students based on their probability of transferring to charter schools, we found that for a few middle strata, charter school students performed significantly better than public students. The pattern, however, was mixed, since it does not persist across all strata.

Causal Effect on Student Growth (Apply Two-Level Repeated Measure Models)

One research question we want to address in this study is whether students grow at different rates after they transfer to charter school compared with those who do not. By applying the two-level repeated measure model (three-year scores nested within students), we compare the achievement growth for students who transferred to charter school and stayed there for three years to the achievement growth for those who never transferred to charter school (public school students). Not only did we compare student growth at the group level, we also applied a model including the stratum interaction effect and charter school effect (see Equation (3) for details of the model) to investigate the question "Do charter school students and public school students grow at different rates when they are matched?"

Cohort One

For Cohort One, we found that student achievement scores increased 6 points every year, but the two groups grew slightly different rates along the three years (see Table 10). It indicates that public school students grew half a point more every year than charter school students. When we compared the two groups of students at each stratum, there was no significant difference in growth between the two groups (see Table 11).

Cohort Two

Cohort Two presents a slightly different picture about charter school effect on student growth. Student achievement scores increased 5.3 points every year, but the two groups grew at similar rates along the three years (see Table 12). When we compared the two groups of students at each stratum, we found no significant difference in growth between the two groups (see Table 13).

Summary

When we matched students based on their probability of transferring to charter school, we found no significant difference in achievement growth between charter school and public school students.

Conclusions, Limitations and Future Directions

Overall the study—based on two cohorts of students across four years (one year before transfer and three years after)—presented a mixed picture of charter school effect. First, we found no particular pattern that minority students or students from poorer schools tended to attend charter schools. Second, without matching students, the results were consistent: there was no charter school effect. However, when we matched students based on their probability of transferring to charter schools, the picture became more complex. In general, for the few strata where both groups of students had the same level of probability of transfer, charter school students performed significantly better. The differences can be more than one year's worth of growth. However, these differences were not consistent across strata. In terms of achievement growth, the results were more straightforward. After matching charter school students and public school students, they grew at similar rates. No significant difference in growth was found, whether or not students go to charter schools. One limitation of the growth results is that the sample was limited to students who stayed in charter schools for three years, while some of the students transferred from charter schools during the three years. We need to examine how attrition impacts our conclusion in our next step.

Another limitation: based on available data, we understand that our variable list is not exhaustive, and we can miss some variables that impact whether or not students transfer to charter school. For example, we don't have access to students' free and reduced lunch status, which is usually an important indicator of student socioeconomic status. But we have access to students' previous scores and school level FRL percentage, which helps us indirectly control for student socioeconomic status. Conducting sensitivity analysis for the next step will help us understand whether our model, based on selected and available variables, can sufficiently support our conclusion.

With these limitations in mind, our study does show that using multilevel models and propensity score stratification challenges the simple conclusion that charter schools have no effect and provide a more complex picture in the quasi-experimental context.

Our results were mostly consistent with the mixed findings in existing literature on charter school effect (Zimmer, Gill, Booker, Lavertu, Sass & Witte, 2009). To obtain a comprehensive picture of charter school effect, we need to expand the analysis to multiple grades. This will help us provide systematic evidence of charter school effect (or no effect) and inform the debate on charter schools. Moreover, instead of drawing a simple conclusion on whether or not charter school improves student achievement, the next step for the study is understanding what types of charter schools work better than others.

Table 1: List of Variables Used in Propensity Score Estimation

Student Level	Ethnicity Gender Previous test score
School Level	School poverty rate (FRL%) School location (urban, suburban, town, rural) School percent minority
District Level*	A district with charter schools vs. a district without Total local revenue Total current expenditures for elementary/secondary education Payments to charter schools

**From NCES District Finance Survey (F-33)*

Table 2: Distribution of the Logit of Propensity Score for Students Who Transferred or Did Not Transfer to Charter Schools (Cohort One)

	Public			Charter		
Stratum	n	mean	SD	n	mean	SD
$e_0=0$	2165	-3.25	.298	86	-3.13	.258
$e_0=1$	836	-2.50	.157	92	-2.50	.149
$e_0=2$	338	-1.98	.134	52	-1.97	.125
$e_0=3$	257	-1.46	.176	55	-1.38	.201
$e_0=4$	46	-.93	.136	31	-.93	.110
$e_0=5$	62	-.33	.156	29	-.27	.178
$e_0=6$	21	.61	.386	22	.71	.324
$e_0=7$		n/a		168	5.90	.863

Table 3: Distribution of the Logit of Propensity Score for Students Who Transferred or Did Not Transfer to Charter Schools (Cohort Two)

	Public			Charter		
Stratum	n	mean	SD	n	mean	SD
e_0 =0	1681	-3.58	.425	36	-3.54	.539
e_0 =1	1040	-2.59	.217	79	-2.57	.210
e_0 =2	352	-1.98	.135	54	-1.95	.134
e_0 =3	225	-1.47	.197	75	-1.41	.216
e_0 =4	120	-.86	.136	48	-.86	.118
e_0 =5	83	-.28	.201	60	-.29	.192
e_0 =6	13	.57	.170	15	.66	.220
e_0 =7		n/a		326	6.01	1.606

Table 4: Achievement Comparison Between Public School and Charter School Students at the End of the First Year of Transfer (Cohort One)

Stratum	Public			Charter			Mean Difference	p
	n	mean	SD	n	mean	SD		
$e_0 = 0$	2112	224.33	16.170	101	223.35	16.592	-.979	>.05
$e_0 = 1$	775	223.65	15.556	87	224.17	15.027	.516	>.05
$e_0 = 2$	328	223.81	14.563	49	225.81	14.094	1.991	>.05
$e_0 = 3$	249	224.15	15.265	52	230.39	11.787	6.243	.006
$e_0 = 4$	45	226.41	14.349	31	231.14	17.157	4.739	>.05
$e_0 = 5$	60	221.07	15.142	29	227.45	7.172	6.385	.008
$e_0 = 6$	20	222.57	14.013	19	221.84	17.760	-.733	>.05
$e_0 = 7$		n/a		163	223.80	12.878	n/a	n/a
Total	3589	224.08	15.784	531	225.16	14.408	1.08	>.05

Table 5: Achievement Comparison Between Public School and Charter School Students at the End of the Second Year of Transfer (Cohort One)

Stratum	Public			Charter			Mean Difference	p
	n	mean	SD	n	mean	SD		
$e_0=0$	1632	230.59	16.136	88	228.93	18.527	-1.658	>.05
$e_0=1$	659	230.86	16.628	77	229.26	15.155	-1.601	>.05
$e_0=2$	274	232.12	16.678	31	230.29	20.027	-1.825	>.05
$e_0=3$	187	227.90	15.492	38	234.64	14.721	6.746	.014
$e_0=4$	61	231.99	18.066	37	242.07	13.025	10.079	.002
$e_0=5$	41	232.66	13.000	19	238.52	10.087	5.859	>.05
$e_0=6$	14	219.62	14.512	15	221.05	18.463	1.426	>.05
$e_0=7$		n/a		153	231.29	15.006	n/a	n/a
Total	2868	230.63	16.285	458	231.54	16.366	.91	>.05

Table 6: Achievement Comparison Between Public School and Charter School Students at the End of the Third Year of Transfer (Cohort One)

Stratum	Public			Charter			Mean Difference	P
	n	mean	SD	n	mean	SD		
e_0 =0	1451	236.99	16.675	65	234.34	17.479	-2.648	>.05
e_0 =1	509	236.68	17.502	62	235.65	16.286	-1.031	>.05
e_0 =2	238	236.85	17.726	33	235.18	20.915	-1.669	>.05
e_0 =3	187	235.57	15.854	24	242.33	14.567	6.759	.049
e_0 =4	53	238.08	15.568	35	247.42	10.378	9.333	.002
e_0 =5	33	242.58	16.257	26	235.49	18.960	-7.083	>.05
e_0 =6	7	230.98	15.812	4	231.70	13.583	.717	>.05
e_0 =7		n/a		148	235.35	15.803	n/a	n/a
Total	2478	236.89	16.862	397	236.68	16.683	-.179	>.05

Table 7: Achievement Comparison Between Public School and Charter School Students at the End of the First Year of Transfer (Cohort Two)

Stratum	Public			Charter			Mean Difference	p
	n	mean	SD	n	mean	SD		
$e_0 = 0$	1639	224.75	15.575	85	219.65	14.230	-5.097	.003
$e_0 = 1$	976	225.21	15.612	79	222.51	15.211	-2.695	>.05
$e_0 = 2$	334	223.22	15.526	53	227.38	13.554	4.165	>.05
$e_0 = 3$	222	224.76	15.035	70	228.17	15.562	3.413	>.05
$e_0 = 4$	117	224.79	15.376	48	225.86	14.028	1.074	>.05
$e_0 = 5$	79	218.90	13.760	60	223.30	13.872	4.399	>.05
$e_0 = 6$	13	224.90	13.875	14	228.89	8.750	3.989	>.05
$e_0 = 7$		n/a		318	224.86	15.106	n/a	>.05
Total	3380	224.60	15.515	727	224.51	14.819	-.762	>.05

Table 8: Achievement Comparison Between Public School and Charter School Students at the End of the Second Year of Transfer (Cohort Two)

Stratum	Public			Charter			Mean Difference	P
	n	mean	SD	n	mean	SD		
$e_0 = 0$	1450	231.23	16.463	33	224.90	11.390	-6.324	.004
$e_0 = 1$	762	231.86	15.746	60	230.92	14.425	-.937	>.05
$e_0 = 2$	202	232.30	15.590	35	230.75	14.749	-1.551	>.05
$e_0 = 3$	126	232.06	19.353	39	237.06	16.719	4.998	>.05
$e_0 = 4$	59	231.10	16.753	34	237.38	15.559	6.280	>.05
$e_0 = 5$	49	224.54	15.689	43	233.86	16.495	9.319	.007
$e_0 = 6$	13	238.36	22.185	13	239.85	14.299	1.487	>.05
$e_0 = 7$		n/a		263	231.40	17.143	n/a	n/a
Total	2661	231.44	16.389	520	232.12	16.314	.680	>.05

Table 9: Achievement Comparison Between Public School and Charter School Students at the End of the Third Year of Transfer (Cohort Two)

Stratum	Public			Charter			Mean Difference	P
	n	mean	SD	n	mean	SD		
e_0 =0	1455	237.23	15.661	27	233.79	13.868	-3.443	>.05
e_0 =1	623	236.09	17.077	54	236.99	11.975	.897	>.05
e_0 =2	175	238.63	17.011	19	237.47	17.016	-1.165	>.05
e_0 =3	93	237.24	19.681	34	245.82	14.785	8.588	.022
e_0 =4	51	232.16	19.901	34	240.40	17.858	8.241	.049
e_0 =5	43	229.18	14.139	39	239.18	16.784	10.004	.004
e_0 =6	13	243.02	22.411	11	246.75	12.392	3.730	>.05
e_0 =7		n/a		188	233.54	16.236	n/a	n/a
Total	2453	236.83	16.450	406	236.70	16.013	.130	>.05

Table 10: Growth Comparison Between Public School and Charter School Students During Three Years (Cohort One)

Fixed Effects		Coefficient	se	p value
Average score for non-transfer group		224.83	.27	.000
Transfer effect		.83	.77	.000
Average growth rate		6.0	.08	.000
Transfer effect*year		-.53	.22	.018
Random Effects	**Variance Component**	**Std.Error**	**Wald Z**	**p value**
Individual intercept	227.02	5.77	39.31	.000
Individual growth rate	13.74	1.17	11.68	.000

Table 11: Growth Comparison Between Public School and Charter School Students at Each Stratum (Cohort Two)

Fixed Effects	Coefficient	se	p value
Average score for non-transfer group	227.56	8.16	.000
Stratum 1 e_0 =0	-2.83	2.56	>.05
Stratum 2 e_0 =1	-1.68	2.57	>.05
Stratum 3 e_0 =2	-2.56	2.63	>.05
Stratum 4 e_0 =3	-3.32	2.66	>.05
Stratum 5 e_0 =4	-2.9	2.67	>.05
Stratum 6 e_0 =5	-4.11	2.73	>.05
Stratum 7 e_0 =6	n/a	n/a	n/a

Random Effects	Variance Component	Std. Error	Wald Z	p value
Individual intercept	225.94	5.76	39.22	.000
Individual growth rate	13.74	1.18	11.67	.000

Table 12: Growth Comparison Between Public School and Charter School Students During Three Years (Cohort Two)

Fixed Effects		Coefficient	se	p value
Average score for non-transfer group		224.70	.27	.000
Transfer effect		.94	.69	>.05
Average growth rate		5.34	.08	.000
Transfer effect*year		-.07	.22	>.05
Random Effects	**Variance Component**	**Std. Error**	**Wald Z**	**p value**
Individual intercept	226.18	5.62	40.27	.000
Individual growth rate	15.08	1.21	12.49	.000

Table 13: Growth Comparison Between Public School and Charter School Students at Each Stratum (Cohort Two)

Fixed Effects		Coefficient	se	p value
Average score for non-transfer group		225.12	5.94	.000
Stratum 1 $e_0 = 0$		-.69	1.86	>.05
Stratum 2 $e_0 = 1$		-.11	1.79	>.05
Stratum 3 $e_0 = 2$		-1.56	1.97	>.05
Stratum 4 $e_0 = 3$		-1.15	1.89	>.05
Stratum 5 $e_0 = 4$		.71	1.92	>.05
Stratum 6 $e_0 = 5$		.0001	1.93	>.05
Stratum 7 $e_0 = 6$		n/a	n/a	n/a
Random Effects	Variance Component		Wald Z	p value
Individual intercept	224.34	5.62	40.27	.000
Individual growth rate	14.89	1.20	12.36	.000

References

Austin, P. C. (2011). An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behav Res* 2011, 46: 399–424.

Austin, P. C. & Mamdani, M. M. (2006). A comparison of propensity score methods: A case-study estimating the effectiveness of post-AMI statin use. *Statistics in Medicine* 25: 2084–2106. doi:10.1002/sim.2328

Cochran, W. G. (1968). The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics* 24: 295–313.

Cook, T. D., Shadish, W. R. & Wong, V. C. (2008). Three Conditions under which Experiments and Observational Studies Produce Comparable Causal Estimates: New Findings from Within-Study Comparisons. *Journal of Policy Analysis and Management* 27 (4): 724–750.

Grisword, M. E., Localio, A. R. & Mulrow, C. (2010). Propensity score adjustment with multilevel data: setting your sites on decreasing selection bias. *Annals of Internal Medicine* 152 (6): 393.

Faries, D. E., Leon, A. C., Maria Haro, J. & Obenchain, R. L. (2010). Analysis of observational health care data using SAS®. Cary, NC: SAS Institute, Inc.

Flom, P. L., McMahon, J. M. & Pouget, E. R. (2007). Using PROC NLMIXED and PROC GLMMIX to analyze dyadic data with a dichotomous dependent variable. Paper 179-2007. SAS Global Forum.

Holland, P. W. (1986). Statistics and Causal Inference. *Journal of the American Statistical Association* 81: 945-70.

Hong, G. & Raudenbush, S. W. (2006). Estimating Kindergarten Retention Policy. *Journal of the American Statistical Association* 101 (475): 901-910.

Hong, G. & Raudenbush, S. W. (2008). Causal Inference for Time-Varying Instructional Treatments. *Journal of Educational and Behavioral Statistics* 33 (3): 333-362.

Hong, J. & Yu, B. (2008). Effects of kindergarten retention on children's social-emotional development: An application of propensity score method to multivariate, multilevel data. *Developmental Psychology* 44: 407–421.

Kim, J. & Seltzer, M. (2007). Causal inference in multilevel settings in which selection processes vary across schools. National Center for Research on Evaluation, Standards and Student Testing (CRESST), CSE Technical Report 708, University of California, Los Angeles.

Kleyman, Y. & Hansen, B. (2008). Test for Covariate Balance in Comparative Studies. Paper presented at the Joint Statistical Meetings 2008 annual conference.

Newson, R. B. (2008). Causality, confounders, and propensity scores. Presented at Asthma Club, December 11, 2008. Retrieved from http://www.bristol.ac.uk/alspac/

Oakes, J. M. & Johnson, P. J. (2006). Propensity score matching for social epidemiology. *Methods in Social Epidemiology,* J.M. Oakes and S. Kaufman (Eds.), pp. 364-386. Jossez-Bass.

Pearl, J. (2009). Causal inference in statistics: An overview. *Statistics Surveys* 3: 96-146.

Raudenbush, S. & Bryk, A. S. (2002). Hierarchical linear models: Applications and data analysis methods. Sage, Newbury Park, CA, second edition.

Robson, L. S., Shannon, H. S., Goldenhar, L. M., Hale, A. R. (2001). Quasi-experimental and experimental designs: more powerful evaluation designs, Chapter 4 of *Guide to Evaluating the Effectiveness of Strategies for Preventing Work Injuries: How to show whether a safety intervention really works.* Institute for Work & Health, Canada.

Rosenbaum, P. R. & Rubin, D. B. (1983a). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika* 70: 41–55.

Rosenbaum, P. R. & Rubin, D. B. (1983b). Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society,* Series B. 45: 212–218.

Rosenbaum, P. R. & Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association*. 79: 516–524.

Rubin, D. (2007). The design versus the analysis of observational studies for causal effects: Parallels with the design of randomized trials. *Statistics in Medicine* 26: 20–36.

Stuart, E. A. (2008). Developing practical recommendations for the use of propensity scores: Discussion of 'A critical appraisal of propensity score matching in the medical literature between 1996 and 2003' by Peter Austin. *Statistics in Medicine* 27: 2062-2065.

Wyse, A. E., Keesler, V., Schneider, B. Assessing the effects of small school size on mathematics achievement: A propensity score-matching approach. *Teachers College Record,* 2008. 110: 1879–1900.

Zimmer, R., Gill, B., Booker, K., Lavertu, S., Sass, T. & Witte, J. (2009). Charter Schools in Eight States: Effects on Achievement, Attainment, Integration and Competition. Santa Monica, CA: RAND Corp. Retrieved from http://www.rand.org/pubs/monographs/2009/RAND_MG869.pdf

Learn more about our research at NWEA.org/Research.

NWEA has nearly 40 years of experience helping educators accelerate student learning through computer-based assessment suites, professional development offerings, and research services.

July 2014

NWEA™