On the prospective comparison of Charter and Public Schools using Propensity Score Analysis

Robert Pruzek*
SUNY Albany

The debate over effectiveness of charter schools in relation to public schools has been mounting for over a decade. Based in reform efforts, charter schools are usually autonomous public schools run by teachers, parents, and/or community organizations. A key argument often advanced to justify charter schools is that they can reduce bureaucracy and thereby improve efficiency. Charter schools are given autonomy and deregulated in exchange for a time-limited contract for student achievement. As of 2000, about one quarter of a million students in the US now attend charter schools, just under 1 percent of the total of K-12 public school population at the end of the last century (cf. Report cited in next paragraph).

Despite their growing popularity, and despite the advantages seen by many advocates, charter schools have not been systematically or comprehensively compared with public schools using scientific standards for evaluation. Numerous reports exist that testify to good charter school experiences, and beliefs, often aimed at supporting charter schools in relation to their better-established public school counterparts. But scientifically designed studies aimed at 'objective' comparisons of these two kinds of schools seem not to exist. Interestingly, the government document, 'The State of Charter Schools, 2000, Fourth Year Report,' ['http://www.ed.gov/pubs/charter4thyear/'— pdf version of report can be downloaded] offers *no evidential basis* for comparing achievements of charter and public students. Its authors use survey data to focus on descriptions of the numbers of charter schools, where they exist, how they have grown, and what purposes they are intended to serve. But the desired comparison data are typically incomplete, unfocused, and easily criticized as a basis for judgments about the relative virtues of these two kinds of schools.

Surely one would like to answer questions such as, 'What are the typical achievements of charter school students, as compared with their public school counterparts?' Or, better, 'How, and how much, do various achievements, and behavioral characteristics, of charter school students generally differ from those of their public school counterparts, having accounted for relevant differences between students in the respective schools?' Soundly researched answers to such questions seem essential if parents are to be able to make reasonable judgments about where their children should be educated, or for the general public to decide whether the charter schools 'experiment' should be continued, at least in some form. Moreover, if revisions are to be made in charters for such schools, evidence derived at least in part from student performance characteristics seems essential.

The chief premise of this note is that a newly developed methodology called propensity score analysis (PSA) can be especially helpful in answering such questions. Notably, use of PSA methods seems likely to improve upon traditional methodology most often used as a basis for educational treatment comparisons. Since charter school legislation varies state-by-state, charter schools generally vary across states and grade-levels. Most charter schools focus on grades K-3, but schools also exist in grade-level ranges of 4-8, and 9-12. Consequently, in many

states/communities there may be no charter school counterpart to the public schools, at least at some grade levels. Given these realities, as well as the fact that random assignments of students to charter and public schools would rarely be feasible, PSA methods seem worthy of examination for their potential to adjust for differences between students in the two kinds of schools. That is, a well-conceived PSA could account for selection bias, and thus provide a basis for sound comparison of these two kinds of schools, despite the unfeasibility of random assignments of students to schools. In what follows I shall sketch the logic of PSA in the context of comparing charter and public schools; it should be clear that the basic PSA method has potential applicability in many contexts.

The initial goal of PSA methods for comparing charter and public schools would be to observe as many covariates (predictors) as possible that might account for differences between students who enter these two kinds of schools. The best covariates in this context would be ones that best predicted the ***distinction*** between charter and public schools and had some relationship with the ultimate outcome or response variable(s). That is, the initial aim would be to learn how effectively to discriminate these kinds of schools. Student-level covariates of major interest might be (standardized) test scores for English-Language arts, reading comprehension, vocabulary, word attack; various components of mathematics achievement or skill, as well as science achievement scores. Additional student-level covariates to discriminate between charter and public school students could include measures of the numbers of books in the child's home, parents' education, family SES, attendance history, and measures of motivation to learn. Choosing effective covariates will surely involve many stakeholders in the educational enterprise; the more knowledge one has about the differences between the kinds of students that enter public and charter schools, the better equipped s/he would be to choose effective covariates.

Supposing that observations were made for a comprehensive selection of covariates, PSA methods entail using covariate information to model or predict the (binary) 'treatment assignment' variable, in this case, whether a student is in a charter or a public school. For example, logistic regression analysis could be used to generate a single covariate, itself a function of many covariates that would serve to predict the binary variate. Such a derived variable, based solely on the covariates, is called an estimated 'propensity score,' as each value (usually a score in the range zero to one) indicates the 'propensity' of a single student to have chosen charter over public schools. The complement, one minus the estimated propensity score, indicates the propensity to have chosen the other type of school. Modeling (the probability of) treatment assignment constitutes the *first phase* of a PSA, which generally has two main parts, or Phases.

Given a reasonable selection of covariates, students with *similar* propensity scores will be comparable to one another; that is, within each (narrow) band of propensity scores, covariate distributions will tend to be similar. The key point is that within any narrow band of propensity scores students will be comparable with respect to any covariate that distinguishes between these two kinds of schools because the bands have been derived by 'optimally' taking student differences into account. A particular PSA application might focus on discriminating between treatments within a particular state, city or community. Students with similar propensity scores will thus not only be similar to one another with respect to the covariates, they will be

comparable within a particular state, city or community. (This means that in practice the logistic regression model itself should be carefully selected, quite possibly after several alternatives have been examined.)

The basic strategy of PSA is to sort or rank students according to their estimated propensity scores and then stratify on these scores so that within strata, propensity scores will be similar (perhaps with as few as five strata). A standard approach is then to make comparisons within each PS stratum for *any* chosen outcome measure(s). This could be done at each of a number of grade levels. Assuming a continuous dependent or response variable, the basic comparison to be made for each response, in each domain of application, would usually be the ***mean response variable difference between charter and public schools***, computed initially within each stratum, and then averaged across strata.

Such a derived (average) difference score is called a 'direct estimate of treatment effect,' as it should (principally) reflect treatment differences, having accounted for the covariates that were used in deriving the propensity measure. Any such direct effect estimator might informatively be divided by the (pooled) standard deviation (across all treatment/strata combinations) to obtain an estimate of standardized effect-size; this would yield a mean difference expressed in standard deviation units. Significance testing may also be used, but this is not central to the method. Theory exists to support such treatment effect estimators as 'unbiased' for comparing treatment effects, conditional on the choice of covariates used in deriving (estimated) propensity scores (Rosenbaum and Rubin, 1983 [*Biometrika*]; and Rosenbaum, 2002, *Observational Studies, 2nd Ed.* [New York: Springer]). (NB: PSA is wholly different from analysis of covariance (which does NOT aim to control for selection bias), both in its logic and its methodology; PSA is also not properly described as an instrumental variable method, such methods having become popular in econometrics in recent years. To date, PSA has been used mostly in medical science.)

It is notable that initial strata obtained in Phase I of a PSA study need not be changed or modified for assessments across different outcome measures. So a single derived propensity-covariate could serve to adjust for relevant differences between charter and public school students for any outcome or response variable. Depending on the times when outcome measures or scores were obtained, 'treatments' might be conceived in terms of semesters, full years, or longer. Any set of outcome measures of interest to parents or educators could be used for comparisons, including, but not limited to, achievement measures, measures of behavior, interest, or even the likelihood of staying in school in the future.

In any application of PSA the collection of data, including all covariates and outcome measures, is likely to be somewhat difficult, and certainly would take concerted efforts. However, different covariates might reasonably be chosen in different states, communities or at different grade levels with different degrees of effort being used in different applications. Response variables can also be expected to differ across different applications of PSA. In many situations raw data for covariates and responses may reside in school databases, or other archives, since at least some response variable information will be part of any end-of-year summary data in a typical school. Moreover, evaluations at a single grade level, say grade three, would generally use records from end of second year as covariates; then the schools' grade three

summary data are likely to include at least some outcomes of interest for PSA. Still, it is likely that specific needs of a well-designed PSA study will require additional data, beyond what had been routinely collected, as well as additional resources to support data collection.

Since random assignment of students to the two kinds of schools will rarely be an option, methods such as PSA seem needed if the analyst is to control for selection bias. Interpretations of PSA results seem likely to be (much) more straightforward than usual comparisons of such observational data, since it may be learned that there is an evidential basis for comparing only certain kinds of students, as when some PSA strata have few or no students in either the charter or public school 'cell'. Therefore one can expect that the PSA stratification will itself provide important information about just what charter/public school comparisons are feasible, or not. In this sense PSA applications have the laudatory feature that they tend to entail self-criticism. In the end, evidence from a PSA could provide evidence that would favor the proponents of charter schools, or the advocates of public schools. A scientifically trained investigator will aim for an objective assessment of differences to the extent that they exist, having adjusted for as many relevant differences (selection bias) as possible between the two kinds of schools.

In many communities parents face decisions they did not face in the past concerning where their children should be educated, and the larger decisions as to whether, or how much, charter schools should be funded. Cost issues are also likely to figure in to any such comparison. Of particular relevance is that in the current popular and educational press there seems to be little understanding of the issue of selection bias, and how it generally confounds interpretations of educational data – which are nearly always observational, not experimental. While a great deal of attention is now being focused on making charter schools 'accountable' there seems to be no broad recognition that having students take tests, then recording, compiling and summarizing outcome information, no matter what its form, does not in itself provide a basis for making sound judgments about the relative merits or demerits of schools. So-called 'accounting' data may provide a reasonable *starting point* for knowing which educational modality is 'better,' or 'best,' but without taking selection bias into account, sound interpretations of such data cannot help but be difficult, and can easily become impossible, even for the experienced educator or evaluator. One wonders who is looking at the larger picture. Who is seriously trying to provide useful information to the parent who has no prior vested interest in one type of school or the other, but simply wants what is 'best' for her child?

Of course the general problems of educational treatment comparison extend far beyond comparing charter and public schools. But, when, as is usual, there is only an observational database available for making treatment comparisons, selection bias cannot be ignored. To compound this problem – and to further confound interpretations of data – there are currently few educational researchers with background or training that covers basic methodology of propensity score adjustment that corrects for selection bias. Dozens of PSA studies in the charter public school context are feasible, but just one good one is clearly needed. Who will plan it, get funding for it, or conduct it?

---

*With thanks for the assistance from Joseph M. Baltrus.