EXECUTIVE SUMMARY (SPRING 2013)

Dissertation: "The Effects of School Autonomy on Students' Reading Achievement in Early Grades: A Dose-Response Treatment Approach."

Esther F. S. Carvalhaes

Problem Statement/Policy Issues

School autonomy is at the core of substantial educational policies aimed at improving school effectiveness and students' academic performance both in the United States and abroad. Initiatives such as School-Based Management or Site-Based Management (SBM), and the charter school movement have placed increased school autonomy at the heart of their reform efforts in response to the challenge of turning low-performing, often high-poverty schools into centers of true educational opportunity (Wohlstetter, 1995; Barrera-Osorio et al., 2009).

The appeal of increased school autonomy as a strategy for school reform is understandable: firstly, a transfer in power and responsibility is a low-cost type of reform. Secondly, demands for local control of schools find wide support from distinct constituencies, from advocates of community-governed schools to proponents of privatization of the school system (Levin, 2001). Thirdly, analyses of countries participating in international standardized achievement tests reveal a high positive correlation between school autonomy and students' performance (Barrera-Osorio et al., 2009; OECD, 2004). Lastly, increased autonomy is a conditional offer, granted in exchange for increased accountability and better educational outcomes (CREDO, 2009; Zajda & Gamage, 2009; Barrera-Osorio et al., 2009).

Despite the prominence of these reform proposals in the current policy environment, the amount of credible evidence demonstrating the effectiveness of autonomous schools on students' learning is surprisingly limited and presents inconsistent results. Part of the problem lies in the difficulty of conducting random experiments to learn about school effects: schools are not typically assigned to different levels of autonomy at random, nor are students assigned to schools randomly. Observational studies, on the other hand, face the challenge of minimizing selection bias in the estimation of causal effects.

Research Question

This dissertation will investigate the following question: <u>What are the effects of increased school autonomy on schools' average reading gains in Kindergarten and first grade?</u> I focus on early literacy which sets the foundation for later reading proficiency, the quintessential basic skill required for learning in advanced grades. By examining the above research question, my purpose in this dissertation is three-fold: (i) to assess the plausibility of claims that increased school autonomy leads to increases in student learning; (ii) to demonstrate how propensity scores for dose-response models can be applied to critical educational policy questions; (iii) to capitalize on the availability of a nationally representative database of schools offering Kindergarten to examine school-level effects when they first occur.

Theoretical Framework

To address the problem of estimation bias in observational studies, I utilize Rosenbaum and Rubin's (1983) counterfactual framework of causality in which they proposed the use of propensity scores and demonstrated their superiority in reducing estimation bias when compared to traditional methods of adjusting for covariates. A propensity score is defined as the conditional probability of exposure to a treatment, given a set of covariates. It is a way to model treatment assignment in an effort to account for pretreatment differences between groups in nonrandom studies (i.e. self-selection). Units with the same propensity score are thought to be similar in the distribution of covariates but different only in the treatment they received thus mimicking an experimental design.

Propensity scores are typically used to model treatment assignment to binary treatments. In this dissertation, I use an extension of propensity scores proposed by Joffe and Rosenbaum (1999) to estimate causal effects in the case where the treatment is not binary, but ordinal instead. The independent variables – school autonomy and its dimensions – are treated as multi-valued treatments where schools receive different levels or doses of the treatment (from low to high levels of autonomy) and the interest is in learning the effects of incremental doses of the treatment on the outcome variables. This can be accomplished through the use of balancing scores for dose-response models as recommended by Joffe and Rosenbaum (1999), which provide the theoretical foundation and estimation methods employed in this study.

Data

The Early Childhood Longitudinal Study-Kindergarten (ECLS-K) will be used for this study. Kindergarten is an auspicious grade to start examining school-level effects since it captures children's learning when they are first exposed to school life. Moreover, utilizing a nationally representative sample of schools allows the estimation of effects that are unlikely to be obtained in traditional experimental fashion. Our school-level analyses utilize a total of 633 schools. The interest is in an analytical sample of schools offering conventional Kindergarten programs.

Dependent Variables. (1) school mean reading item response theory (IRT) gain scores, Kindergarten (school mean score based on student-level reading gains, computed by taking the difference in student's reading IRT scores between the Spring 1999 and the Fall 1998); (2) school mean reading IRT gain scores, first grade (mean score is based on individual gains, obtained by subtracting the student's reading IRT score in the Spring 1999 from the score obtained in the Spring 2000). In both cases, IRT scores are selected because they are appropriate to track achievement gains over time.

Independent Variables. In this study, an autonomous school is defined as one in which the school's internal decision-making power (exercised by one or various agents) is higher than the power of external agents (exercised by one or various individuals or groups) over given areas of control (e.g. instruction, personnel/administrative issues). Accordingly, school autonomy is measured by a set of variables indicating the extent to which different decision-makers control key areas of the school management, namely, hiring and firing teachers, selecting textbooks, setting curriculum and standards, establishing evaluation policies and practices, deciding on budget, and planning professional development.

The measures of school autonomy come from a series of questions about decision-making answered by the principal in each school. Based on the principals' answers to these questions, three independent variables are constructed: (1) school autonomy on instructional matters (SAIN), (2) school autonomy on personnel/administrative issues (SAPA), and (3) a global measure of school autonomy (SAGL).

Control Variables. Extensive sets of school characteristics and contextual variables (thought to predict treatment assignment) are used for the estimation of balancing scores. Child- and teacher-level variables are aggregated to the school-level. The control variables encompass characteristics of the schools, their organization and policy, classroom resources and environment, teachers' practices, education and background, student body composition and family background, as well as the parents' involvement in the students' progress.

Methods

Propensity scores are usually employed to model treatment assignment to binary treatments. A propensity score, let us call it X, is a summary of pretreatment variables that predicts whether a unit is assigned to the treatment or control group. The set of variables in X function as a conditioning variable: within each level of X, individuals in the treatment and the control group are thought to be similar to each other except for their observed assignment, thus mimicking a randomized experiment; comparing the outcomes of treated and control units within each level of X yields unbiased treatment effects for that stratum under standard assumptions in the counterfactual framework of causality, namely, the stable unit treatment value assumption (SUTVA) and the strong ignorability of treatment assignment (Rubin, 1980). When modeling treatment assignment the goal is to have within each level of X sufficient overlap - though not perfect - in the distribution of X between the treatment and the control groups to allow for comparisons (Morgan & Winship, 2007; Rosenbaum & Rubin, 1983). When that happens, the data are said to be 'balanced'. Once balance is achieved, obtaining average causal effects is straightforward: weighted averages of these stratum-level estimates are computed based on their marginal probability in the sample.

In dose-response models, the impact of variations in the dosage of the treatment is the focus rather than the presence/absence of treatment (Joffe and Rosenbaum, 1999; Hirano & Imbens, 2004; Imbens, 2000). Modeling treatment assignment requires the use of a balancing score for each unit (this is equivalent to a propensity score in the binary case). A propensity score is defined as "*the conditional probability of the actual, perhaps multivariate, treatment given the observed covariates*" (Imai & Van Dyck, 2004, p. 856). To arrive at our desired effects, the steps to be adopted are similar to those described for the binary case: the computation of propensity scores for each school, a demonstration of the covariates balance, and the estimation of effects.

Computing the propensity scores. Joffe and Rosenbaum (1999) set up the conditions under which a single variable, say a function of all the covariates X represented by the scalar b(X), can serve as a balancing score for an ordinal treatment variable. These conditions are true in McCullagh's (1980) ordinal logit model, which can be used for estimating the propensity score. For each school, the logit of being observed in a particular level/dose of autonomy is computed, given the set of pretreatment variables:

$$\log \frac{p\ (T_j \geq d)}{p\ (T_j < d)} = \tau_d + \beta^T X_j\ , \quad \text{for } d = 2, 3 \ \text{ (T with 3 doses).}$$

The linear predictor $\beta^T X_j$ in the logit model equation estimates b(X). Three balancing scores per school are computed, one for each independent variable: global measure of school (SAGL), school autonomy over personnel/administrative issues (SAPA), and school autonomy over instructional matters (SAIN).

Covariates balance. To demonstrate that $\hat{\beta}^T X_j$ in fact balances the covariates, each covariate in X is regressed on the corresponding treatment variable T conditional on the balancing score, $\hat{\beta}^T X_j$. The treatment variable coefficient in each regression should be unrelated to the covariate after controlling for $\hat{\beta}^T X_j$ (Imai & Van Dyk, 2004) in order to satisfy the conditional independence of treatment assignment and covariates. Once the covariates are balanced, the effects can be estimated.

Estimation of Treatment Effects. The effects will be estimated using two different procedures. First, using a general outcome model (to be computed separately for each of the three independent variables): schools' reading gains are regressed on the treatment levels (doses) conditional on the corresponding propensity score (parametric solution). To verify the plausibility of the parametric results, the same effects will also be estimated in a nonparametric way: subclassification, which is shown to further reduce bias in the estimation of causal effects (Imai & Van Dyk, 2004). I will use a five-class, seven-class and a ten-class stratification scheme. Lastly, sensitivity analyses will be conducted to examine the robustness of the results (Rosenbaum & Rubin, 1983; Rosenbaum, 1984).

Preliminary Findings

Prior to the propensity score modeling, exploratory analyses were conducted to evaluate the association between the School Autonomy (SA) measures and: (i) school demographics, *Table 1*; (ii) school organizational characteristics, *Table 2*; (iii) Kindergarten reading gains, controlling for key socio-economic characteristics of students and schools, *Table 3*.

Table 1 reveals significantly higher levels of autonomy among private schools and schools serving economically advantaged students. On average, private schools score higher than public schools in autonomy by almost an entire standard deviation (.83) on the instruction measure, and about half of a standard deviation (.52) on the personnel/administrative measure. The disparity is confirmed when looking at poverty levels only, irrespective of school sector: schools with wealthier students are higher in SAIN (equivalent to r=.23) and in SAPA (r=.14). Schools in rural areas score higher on SAIN when compared to those in urban areas (.21 of a standard deviation), but lower on SAPA compared to central city schools (.34 of a standard deviation).

Table 1. Association between School Autonomy (SA) Measures and School Demographic Characteristics: Bivariate Regression Models. All Schools (N=633)

	SA Global (SAGL) (Scale: -1 to +1)			**SA Instruction (SAIN)** (Scale: -1 to +1)			**SA Personnel/ Administrative (SAPA)** (Scale: -1 to +1)		
Regression 1	*Estimate*	*St. Error*		*Estimate*	*St. Error*		*Estimate*	*St. Error*	
Intercept	-0.045	0.015	**	-0.079	0.022	**	-0.011	0.014	
Private School	0.224	0.024	***	0.290	0.029	***	0.158	0.031	***
Regression 2									
Intercept	0.005	0.023		0.030	0.033		-0.019	0.023	
Urban/Large Town	-0.025	0.027		-0.076	0.035	*	0.025	0.029	
Central City	0.043	0.028		-0.015	0.034		0.100	0.032	**
Regression 3									
Intercept	-0.259	0.049	***	-0.367	0.061	***	-0.149	0.058	*
% Students Above Poverty Level	0.322	0.056	***	0.430	0.069	***	0.211	0.067	***

† p < .1 * p < .05 ** p < .01 *** p < .001

Do schools of choice and those adopting SBM score higher on autonomy than other public schools? Schools of choice are no different than other schools on any measure of autonomy (Table 2). Surprisingly, schools that adopt SBM are on average about .39 of a standard deviation below other schools on SAIN. The direction of the findings show that both types of schools tend to be lower on SAIN but score higher on SAPA, although significant results are found only for SBM schools. This might be due to the fact that adoption of these interventions is more common among low-performing schools.

Table 2. Association between School Autonomy (SA) Measures and School Organizational Characteristics: Bivariate Regression Models. Public Schools Only (N=466)

	SA Global (SAGL) (Scale: -1 to +1)			**SA Instruction (SAIN)** (Scale: -1 to +1)			**SA Personnel/ Administrative (SAPA)** (Scale: -1 to +1)		
Regression 1	*Estimate*	*St. Error*		*Estimate*	*St. Error*		*Estimate*	*St. Error*	
Intercept	-0.047	0.013	**	-0.077	0.020	**	-0.017	0.014	
School of Choice	0.018	0.037		-0.021	0.051		0.059	0.041	
Regression 2									
Intercept	-0.017	0.022		0.020	0.028		-0.053	0.024	*
School Has SBM	-0.039	0.023		-0.136	0.030	***	0.057	0.029	†
Regression 3									
Intercept	-0.050	0.019	**	-0.122	0.025	***	0.022	0.021	
School Receives Title 1 Funds	0.007	0.022		0.063	0.030	*	-0.049	0.027	

† p < .1 * p < .05 ** p < .01 *** p < .001

Is school autonomy associated with reading gains? Table 3 shows that of all school autonomy measures, only SAPA is positively associated with increased reading gains. Initial associations for SAGL and SAIN are fully explained by school sector. The coefficient for SAPA, however, is only partially explained by school sector (regression 2) and remains significant even after controlling for students' poverty level. Introducing the poverty variable on regression 3 clearly reduces the estimated effect of school sector on reading gains while the estimated effect of SAPA remains nearly unchanged. This finding suggests that schools with increased personnel/ administrative autonomy might have a positive impact on reading outcomes that goes beyond school differences in socio-economic background. The robustness of this association requires further investigation. Follow-up analyses utilizing propensity scores are currently underway.

Main Conclusions:

- Schools in more resourceful communities tend to display more local control over decisions involving instruction and personnel/administrative matters. Disparities between rural and urban/more central areas exist, but are not as pronounced.
- Contrary to expectations, schools of choice and those adopting SBM do not exhibit higher autonomy than other public schools. In fact, schools with SBM committees show significantly lower levels of autonomy over instructional matters; they are marginally higher in personnel/administrative autonomy.
- Schools with higher personnel/administrative autonomy exhibit higher reading gains after controlling for key social-economic indicators. Further investigation is required to examine the relationship between these variables.

Significance of the Preliminary Study:

- It allows the empirical verification of claims that schools of choice and those adopting SBM are more autonomous than others. This is a requisite piece of information in studies evaluating the impact of these types of programs/policies on learning. If these schools are no higher than others in levels of autonomy, any positive effects on learning (if observed) must be attributed to factors other than local school control.
- It identifies one type of school autonomy - personnel/administrative - that is positively associated with reading gains (Kindergarten) above and beyond existing economic conditions of schools and students. This finding sets up the stage for further research on the potential policy mechanisms that can plausibly improve learning. Follow-up analyses using propensity scores methodology are currently underway.

Table 3. School Average Reading Gain Scores (Kindergarten) and School Autonomy Measures: Regression Models. All Schools (N=633)

Model 1				**Model 2**				**Model 3**			
Regression 1	*Estimate*	*St. Error*		**Regression 1**	*Estimate*	*St. Error*		**Regression 1**	*Estimate*	*St. Error*	
Intercept	11.711	0.152	***	Intercept	11.740	0.153	***	Intercept	11.681	0.152	***
SA Global	1.831	0.581	**	SA Instruction	0.916	0.463	†	SA Personnel/Admin.	1.853	0.555	**
Regression 2				**Regression 2**				**Regression 2**			
Intercept	11.375	0.178	***	Intercept	11.345	0.180	***	Intercept	11.343	0.175	***
SA Global	1.043	0.619		SA Instruction	0.209	0.496		SA Personnel/Admin.	1.373	0.569	*
Private	1.318	0.369	***	Private	1.492	0.371	***	Private	1.337	0.353	***
								Regression 3			
								Intercept	9.304	0.712	***
								SA Personnel/Admin.	1.287	0.571	*
								Private	0.963	0.373	*
								% At or Above Poverty	2.522	0.854	**
								Regression 4			
								Intercept	9.464	0.787	***
								SA Personnel/Admin.	1.323	0.576	*
								Private	0.890	0.404	*
								% At or Above Poverty	2.471	0.861	**
								School Has SBM	-0.164	0.341	
								Regression 5			
								Intercept	9.295	0.806	***
								SA Personnel/Admin.	1.315	0.579	*
								Private	0.816	0.413	*
								% At or Above Poverty	2.429	0.864	**
								School Has SBM	-0.229	0.348	
								Urban/Large Town	0.382	0.401	
								Central City	0.326	0.396	

† $p < .1$ * $p < .05$ ** $p < .01$ *** $p < .001$

References

Barrera-Osorio, F., Fasih, T., Patrinos, H. A., & Santibáñez, W. (2009). Directions in Development: decentralized decision-making in schools. The theory and evidence on school-based management: The World Bank.

Center for Research on Education Outcomes, CREDO (2009). Multiple Choice: Charter School Performance in 16 States.

Hirano, K., & Imbens, G. W. (2004). The Propensity Score with Continuous Treatments. In A. Gelman & X.-L. Meng (Eds.), Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives: An Essential Journey with Donald Rubin's Statistical Family. New York: Wiley.

Imai, K., & Dyk, D. A. V. (2004). Causal Inference with General Treatment Regimes: Generalizing the Propensity Score. Journal of the American Statistical Association, 99(467), 854-866.

Imbens, G. W. (2000). The Role of the Propensity Score in Estimating Dose-Response Functions. Biometrika, 87(3), 706-710.

Joffe, M. M., & Rosenbaum, P. R. (1999). Invited Commentary: Propensity Scores. American Journal of Epidemiology, 150(4), 327-333.

Levin, H. M. (Ed.). (2001). Privatizing Education. Can the Marketplace Deliver Choice, Efficiency, Equity, and Social Cohesion? Boulder, CO: Westview Press.

McCullagh, P. (1980). Regression Models for Ordinal Data. Journal of the Royal Statistical Society. Series B (Methodological), 42(2), 109-142.

Morgan, S. L., & Winship, C. (2007). Counterfactuals and Causal Inference. Cambridge, MA: Cambridge University Press.

OECD (2004). Policy Brief: Raising the Quality of Educational Performance at School. (2004). Organisation for Economic Co-operation and Development.

Rosenbaum, P. R. (1984). From Association to Causation in Observational Studies: The Role of Tests of Strongly Ignorable Treatment Assignment. Journal of the American Statistical Association, 79(385), 41-48.

Rosenbaum, P. R., & Rubin, D. B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. Biometrika, 70(1), 41-55.

Rubin, D. B. (1980). Randomization Analysis of Experimental Data: The Fisher Randomization Test Comment. Journal of the American Statistical Association, 75(371), 591-593.

Wohlstetter, P. (1995). Getting school-based management right. Phi Delta Kappan, 77(1), 22.

Zajda, J., & Gamage, D. T. (Eds.). (2009). Decentralisation, School-Based Management, and Quality. Dordrecht, Heidelberg, London, New York: Springer.