Lessons Learned from Nicaragua's School Autonomy Reform
A Review of Research by the Nicaragua Reform Evaluation Team of the World Bank

Robert Kaestner
Institute of Government and Public Affairs
University of Illinois at Chicago
815 West Van Buren Street, Suite 525
Chicago, IL 60607
Tel: 312/996-8227
kaestner@uic.edu

Alec Gershberg
Associate Professor
Robert J. Milano Graduate School
of Management & Urban Policy
New School University
72 Fifth Avenue, 5th floor
New York, NY 10011
Ph: (212) 229.5311 x 1412
Fax: (212) 229.5404
gersh@newschool.edu

Paper written for the project "Empowering Parents While Making Them Pay: Autonomous Schools in Nicaragua," funded by the Tinker Foundation. New York: New School University, Milano Graduate School of Management & Urban Policy. See:
http://www.newschool.edu/milano/cdrc/schoolreport/index.html
for the project web site.

Revised October 29, 2002

Executive Summary

In a series of papers, researchers from the World Bank's Nicaragua Reform Evaluation Team (NRET) assessed the effects of Nicaragua's School Autonomy Reform.[1] Begun in 1993, and greatly expanded in 1995, Nicaragua's School Autonomy Reform was intended to increase enrollments and improve student learning by decentralizing the governance and administration of primary and secondary schools. Accordingly, the NRET investigated the effects of this reform on the extent and nature of school participation in the program; on the governance and administration of primary and secondary schools; and on student achievement, as measured by test scores. In this report, we critically review the NRET research in order to identify what policy relevant information this research provides about the effects of the Nicaragua school reform program.

The NRET research is based on matched-comparison design where schools in the reform program were matched to schools not in the program on the basis of enrollment and urban-rural location. Data for the analysis was generated from two household-school surveys implemented in 1995 and 1997, which collected information about schools (e.g., school decision processes, enrollment, teacher characteristics, physical resources) and family characteristics (parental education, socioeconomic status), and two achievement tests implemented in 1996 and 1997, which tested a random sample of students from each school in math and Spanish. The NRET recognized the limitations of this research design for the purposes of obtaining causal estimates of the effect of the reform program, and accordingly attempted to overcome these limitations with established statistical methods.

The major findings of the NRET studies are:

1. that school participation in the reform program was associated with both school-level (age and education of principal and teachers, school infrastructure) and community-level variables (e.g., per capita expenditure, presence of post office, travel time to school); among primary schools, community-level variables tended to be more important, and among secondary schools, school-level characteristics tended to be more important.

2. that program participation was associated with a significant increase in local control, but that greater periods of program participation did not increase local control—i.e., there

[1] The four papers reviewed in this report are:

Nicaragua's School Autonomy Reform: A First Look. Elizabeth King, Laura Rawlings, and Berk Ozler. September 30, 1996.

What's Decentralization Got to Do with Learning? The Case of Nicaragua's School Autonomy Reform. Elizabeth King and Berk Ozler. April 27, 1998.

Nicaragua's School Autonomy Reform: Fact or Fiction? Elizabeth King, Berk Ozler, and Laura Rawlings. September 1999.

Decentralization and Student Achievement: The Case of Nicaragua's School Autonomy Reform. Berk Ozler. May 2001

was no dose-response effect; most of the increase in local control was focused on administrative decisions and not pedagogy.

3. that de facto autonomy, as measured by the percentage of decisions made locally, was associated with an increase in math tests for primary students and Spanish test scores for secondary students.

Our assessment of the NRET's research leads to a more cautious conclusion. From our reading of the evidence, the clearest result of the Nicaragua School Autonomy Reform was a broad based increase in local control of schools. Nicaragua has seen an increase in school-based management (especially of administration) in many if not the majority of its schools (both autonomous and non-autonomous). This indicates that there was a significant latent talent at the local and school level in the early 1990s and that school stakeholders were ready and willing to exercise greater decision-making capacity. It is likely, but not certain, that the Nicaragua School Autonomy Reform was responsible for this change. On the important question of whether reform resulted in an improvement in student achievement, we do not believe that the research of NRET is adequate to answer this question because of the non-random participation of schools in the reform program. Although the NRET paid significant attention to this problem, in the end, the statistical procedures employed by the NRET were demonstrably not effective. Thus, the credibility of their findings is in doubt and it remains an open question as to what effect school reform had on student performance.

What factors determined school participation in Nicaragua's school reform program?

Summary:

The Nicaragua Reform Evaluation Team (NRET) used data measuring school-level and municipal-level characteristics to investigate what factors determined school participation in Nicaragua's school reform program. The sample of schools selected included a random sample of schools participating in the school reform program (target schools) and a random sample of non-participating schools (comparison schools). Target and comparison schools were selected after schools were stratified on the basis of enrollment and urban-rural location. The school-level measures include characteristics of the principal and teachers (e.g., age, education, and experience), enrollment and infrastructure. Community-level variables come from the Nicaragua Living Standard Measurement Study of 1993. Multivariate regression models were estimated for two dependent variables—a dichotomous variable indicating participation and a variable measuring the number of months until participation—using a variety of statistical algorithms (e.g., maximum likelihood probit, ordinary least squares, and maximum likelihood Weibull). The results indicate that both school-level (age and education of principal and teachers, school infrastructure) and community-level (e.g., per capita expenditure, presence of post office, travel time to school) variables were associated with school participation decisions. Among primary schools, community-level variables tended to be more important, and among secondary schools, school-level characteristics tended to be more important.

Assessment:

This is a straightforward analysis and the results are credible. The most important conclusion of this analysis is that there is significant selection on observable characteristics: schools that participate in reform have characteristics unlike schools that did not participate. The NRET concludes that: "These findings strongly argue for using an econometric model when comparing the program and non-program schools in order to explicitly account for differences in observable characteristics" (p. 13, Ozler 2001). However, what the NRET fails to explicitly acknowledge is that significant selection on observable characteristics strongly suggests that there is selection on unobservable characteristics, and it is the latter type of selection that undermines attempts to identify causal relationships between school reform and outcomes of interest (e.g., test scores).

The substantial likelihood that there is a significant amount of selection on unobserved characteristics is a major obstacle for the NRET to overcome. Although there are several theoretical solutions to this problem, in practice, these solutions are often not feasible because of data limitations. Moreover, forging on with these solutions when it is unlikely that required conditions for them to be effective have been fulfilled can result in quite misleading inferences. To preview the rest of my comments, I do not believe the NRET has adequately addressed this concern, and consequently, the results of their analyses are ultimately uninformative for policy.

Did Nicaragua's school reform program change school governance and school administration in primary and secondary schools?

Summary:

The NRET used data from a questionnaire given to school principals and a random sample of council members and teachers about the locus of decision making for 25 school functions to investigate whether or not school participation in reform affected school administration and governance—specifically, whether or not control over school administration and pedagogy was local. The sample of schools in this analysis was the same as those described above. Simple differences indicated that on average, stakeholders (principals, teachers, and parents) in participant schools responded that decisions were more likely to be made at the school level than stakeholders in non-participant schools. However, there was significant variation among schools with some non-participant schools having more local control than participant schools. Multivariate analyses were also implemented to explain variation in the percentage of stakeholders who indicated local, relative to central, control of school functions. In these analyses, the likelihood that there was significant selection on unobservable variables was acknowledged and accordingly, the researchers used an instrumental variables (IV) procedure to obtain estimates of the effect program participation on local control. The results of the IV analyses indicated that program participation was associated with a significant increase in local control, but that greater periods of program participation did not increase local control—i.e., there was no dose-response effect. Finally, the researchers also showed that most of the increase in local control was focused on administrative decisions and not pedagogy.

Assessment:

It is not unreasonable to believe that providing schools with de jure autonomy will result in greater local control, and in this respect the findings of the analysis are credible. However, there is significant variation in the data and a significant amount of local control even for traditional schools, particularly among secondary schools. For example, in 1997, approximately 70% of respondents in traditional (non-program) secondary schools reported that the school controlled its budget and plan; 81% of respondents in traditional secondary schools reported local control over teacher evaluation; and 55% of respondents in traditional secondary schools reported local control over classroom and pedagogy. Corresponding figures for autonomous schools are 92%, 81% and 64%. Moreover, between 1995 and 1997, there were significant increases in the amount of local control reported in traditional, as well as autonomous secondary schools. Among primary schools, the differences between traditional and autonomous schools are larger, with respondents from autonomous schools reporting much greater local control.

The reported differences between program and non-program schools in local control of school functions are not compelling, particularly at the secondary level, and they suggest that local control was possible even without reform and that reform did not always lead to greater local control. Furthermore, the growth in local control between 1995 and 1997 for both traditional and autonomous schools suggests that the movement to local control was general and not unique to autonomous schools, which raises the possibility of reverse causality—increasing local control caused the government to grant de jure autonomy. This is especially true since the

schools allowed to enter autonomy were chosen by the Ministry of Education based in part on school staff and community members with a demonstrated propensity and capacity for local control The NRET acknowledges some of these points: "In addition to delays and problems implementing real change within schools, the fact that traditional or non-program schools can also petition the Ministry of Education to take certain actions legally granted only to program schools contributes to the overlap between program and non-program schools" (p. 9, King, Ozler and Rawlings 1999).

All of this leads to the following possibility:

- there was a greater exercise of local control among school personnel and parents (i.e., stakeholders) and a desire to formalize this control;
- the government responded to this by granting de jure autonomy;
- schools with the most capable and motivated stakeholders apply for de jure autonomy;
- de jure autonomy is associated with greater local control, but not because it caused greater control, but because it was the result of greater local control, or because it was associated with more capable and motivated stakeholders who exercised greater local control.

To address these concerns, the NRET used an instrumental variables (IV) procedure to estimate the effect of de jure autonomy on the degree of local control. The efficacy of the IV procedure depends on two criteria: the validity of the exclusion restrictions (variables used to predict de jure autonomy, but which are unrelated to the degree of local control) and the strength of the correlation between the excluded instruments and the endogenous variable—in this case de jure autonomy. As implemented by the NRET, the IV procedure does not meet the necessary requirements for it to be valid. The variables used as instruments, and excluded from the equation predicting the degree of local control, are four community variables: paved road, post office, literacy program and family planning program. It is difficult to believe that these community characteristics are randomly distributed among communities and that they are not correlated with other community characteristics that would determine the capability and motivation of school stakeholders to seek greater local control of the schools. Thus, it is unlikely that these variables are valid instruments. On this point, the NRET is in agreement: "The endogeneity of program participation presents a problem because many of the factors that determine this participation could also influence de facto autonomy" (p. 10, King, Ozler and Rawlings 1999).

The absence of a credible research strategy to address the endogeneity of program participation seriously undermines the results of this analysis. In addition, the descriptive data documents that greater school autonomy was occurring broadly among Nicaragua's schools regardless of de jure autonomy status. Finally, the absence of a dose-response effect is consistent with the hypothesis that greater local control is a function of school characteristics and not the program. It is reasonable to expect that local control would increase the longer schools were in the program, but this was not the case. In sum, while there is some evidence that local control was greater among autonomous schools than traditional schools it is not clear that this was the result of school reform.

Did Nicaragua's school reform affect student learning in primary and secondary schools?

Summary:

For each school in the sample, researchers from the NRET randomly chose 10 to 15 students (from the third grade of primary school and the second year of secondary school) in November-December 1995 and administered tests of math and Spanish in November 1996 (fourth grade of primary school and third year of secondary school). The researchers use these data to estimate the parameters of an expanded education production function (i.e., test scores) that includes the school management regime (i.e., autonomous or traditional). Other variables included in the model are student characteristics (e.g., age and sex), parent characteristics (e.g., mother's education and household wealth), school characteristics (e.g., teacher characteristics, access to text books, and in some case school fixed-effects), and community characteristics (e.g., municipal expenditures). Two statistical problems addressed by the NRET are attrition and the endogeneity of program participation. Accordingly, for some models estimates are obtained by an instrumental variables procedure—to control for the endogeneity of program participation, and in King and Ozler (1998) the IV procedure is placed in the context of a sample selection model in order to address attrition. OLS and IV estimates (in King and Ozler 1998, and Ozler 2001) indicated that de jure autonomy had no effect on math or Spanish test scores for primary or secondary students. OLS estimates indicated that de facto autonomy, as measured by the percentage of decisions made locally, was associated with an increase in math tests for primary students and Spanish test scores for secondary students. IV estimates indicated that de facto autonomy was associated with an increase in math test scores for primary students. Finally, the NRET also estimated the effect of school reform on test scores using a fixed-effect methodology, and found results that are consistent with the IV estimates.

Assessment:

The NRET's analysis of the relationship between school reform and student test scores provides virtually no reliable information. The primary reason for this assessment is that the researchers do not have a credible identification strategy to address two significant problems: attrition from the sample and the endogeneity of program participation.

The attrition from the sample is significant. Although it is extremely difficult to accurately assess the extent of the attrition problem, it is somewhere near 50%. To address this issue the researchers use two approaches: Heckman's sample selection model and a weighted least squares correction. In Ozler (2001), the conclusion is that sample attrition is not important. Interestingly, in King and Ozler (1998) the authors conclude: "The selection bias is significant in all cases except for primary school students in the language test. This justifies the correction with a selection equation" (p. 22 King and Ozler 1998). The inconsistency of the findings is indicative of the problems with the statistical approaches chosen to correct for sample attrition. As noted above, under ideal conditions, the Heckman sample selection model or the Fitzgerald, Gottschalk and Moffitt (FGM) (1998) weighted least squares procedure may effectively address the attrition problem. But the ideal conditions are rarely met. In the case of the Heckman

procedure, the efficacy of the procedure depends on the presence of excluded variables that can predict attrition, but that do not affect test scores. NRET researchers use three variables: monthly fees, negative school actions, and whether or not school lends books. It is difficult to believe that these actions, all taken at the school level, are not correlated with test scores. As such they are not valid exclusion restrictions. The FGM procedure depends on there being an observed variable that is jointly determined with test scores, but which does not influence test scores. Ozler (2001) uses fees as such a variable. While fees and test scores may be jointly determined, fees are likely to determine test scores since they are correlated with resources available in the school. They are also likely to be correlated with the financial capacity of the parents. More generally, since there are undoubtedly unmeasured variables correlated with both fees and test scores, fees will be correlated with test scores. In short, fees are not a valid choice to implement the FGM procedure. In sum, the solutions to the attrition problem offered by NRET researchers are indefensible and unreliable.

Even in the absence of any attrition, the estimates presented by the NRET of the effect of school reform on tests scores are not credible because of the endogeneity of program participation. As the researchers demonstrate so forcefully, program schools and non-program schools are not alike. There are significant observed differences between the two types of schools that strongly suggest similar unmeasured differences. Therefore, school participation in reform is hardly random. To address this problem, the NRET uses an instrumental variables procedure. Unfortunately, this procedure is used almost indiscriminately, as if using it incorrectly is better than not using it at all. This is incorrect. To reiterate, the IV procedure requires the fulfillment of two conditions:

- the existence of variables that predict school participation, but which do not affect test scores,
- and variables that are sufficiently correlated with school participation to produce enough variation in school participation to detect reliably differences in test scores.

These two conditions have not been met. The most important condition is the first—the validity of the exclusion restrictions. For the analysis of primary school test scores, school-level and community-level variables are used as instruments, and excluded from the test score equation. For the secondary school analysis, just school-level variables are used as instruments. In neither case is there a compelling reason to believe these are valid instruments. Indeed, the researchers themselves use school and community characteristics to explain test scores. It stretches the imagination to think that some school- and community-level variables are important predictors of test scores and others are not, and the variables that are not predict school participation. The NRET is not unaware of these problems as illustrated by the following quote from Ozler (p. 108, 2001):

> "It is possible that some of my readers, or even myself, are not satisfied with the identification. Some may disagree with the specification—some of the instruments may belong in the primary equation. Some others may argue that the F-tests and the partial R-squares that I report for the validity of the instruments indicate some bias. Some may think that our identification for secondary schools is weak, as I use school size as our main instrument.

All of these concerns are legitimate."

In sum, the instrumental variables analysis of student test scores and the relationship between school reform in Nicaragua and these scores is not credible. The NRET has provided virtually no reliable information as to whether or not school reform, and greater local control of schools, affects student test scores.

Ozler (2001) also uses a school fixed-effects estimation strategy and data aggregated to the school level to estimate the effect of school reform (de facto autonomy as measured by percentage of decisions made locally) on test scores. He argues that this strategy is in some ways superior to the IV procedure since it does not rely on exclusion restrictions. However, as he notes, it does assume that there are no unmeasured, time-varying characteristics that are correlated with program participation (in this case de facto autonomy) and test scores. This assumption strains credibility since the key independent variable is the degree of autonomy as measured by the percentage of local decisions made at the local level. Why does the degree of local control change over time? It is clearly not solely due to de jure autonomy since, as demonstrated in other parts of the NERT research, local control increased for many non-program schools. In short, de facto autonomy is unlikely to be exogenous even in the fixed-effects model.

There are several other shortcomings associated with the fixed-effects (FE) model. First, it is not clear that Ozler (2001) used weighted least squares, which is the appropriate estimation methodology given the aggregate data. The weight should be proportional to the number of students in each school. Second, the sample of schools included in the FE analysis is much smaller than the original sample—approximately 60 primary and 60 secondary schools. So, the reader (nor the researcher for that matter) does not know whether or not the estimates obtained from the FE model are different (or the same) because of the sample or because of the estimation strategy. The NRET should have replicated the OLS analysis using the smaller sample to assess this possibility. Moreover, the students in these schools are not the same from year to year. As noted, the student samples are refreshed using students in the same schools in the second year (1997). It is unclear how much migration there is between schools in Nicaragua, but if there is a general migration from rural to urban areas, and schools in urban areas were the most likely to participate, this raises the possibility that students in the replacement sample are non-randomly distributed across schools. Finally, aggregating the data was unnecessary, although it may be useful to reduce measurement error. Aggregating the data eliminates within school covariance, which is valid information, and estimates of the effect of school reform (de facto autonomy) may be substantially affected by this choice. In conclusion, I do not believe that the FE analysis is sufficient to lend much more credibility to the findings of this research.