

MAY/JUNE 1995

# ECONOMIC PERSPECTIVES

A review from the  
Federal Reserve Bank  
of Chicago

**Temporal instability of the  
unemployment-inflation  
relationship**

**An analysis of the effect of  
Chicago school reform on  
student performance**

FEDERAL RESERVE BANK  
OF CHICAGO

# Contents

**Temporal instability of the unemployment–inflation relationship. . . . . 2**

**Robert G. King, James H. Stock, and Mark W. Watson**

This article investigates three aspects of temporal instability in the unemployment–inflation relationship.

**An analysis of the effect of Chicago school reform on student performance. . . . . 13**

**Thomas A. Downes and Jacquelyn L. Horowitz**

This article analyzes pre- and post-reform student test scores and graduation rates to evaluate the success of the Chicago public schools in achieving the goals of the 1988 Chicago School Reform Act.

## ECONOMIC PERSPECTIVES

May/June 1995 Volume XIX, Issue 3

William C. Hunter, *Senior Vice President and Director of Research*

**Editorial direction**

David R. Allardice, *regional studies*  
Douglas Evanoff, *financial studies*  
Charles Evans, *macroeconomic policy research*  
Kenneth Kuttner, *macroeconomic policy research*  
Daniel Sullivan, *microeconomic policy research*  
Anne Weaver, *administration*  
Janice Weiss, *editor*

**Production**

Christine Berta, *typesetting coordinator*  
Rita Molloy, Yvonne Peeples, *typesetters*  
Kathleen Solotroff, *graphics coordinator*  
Roger Thryselius, Thomas O'Connell,  
Lynn Busby-Ward, John Dixon, *graphics*  
Kathryn Moran, *assistant editor*

**ECONOMIC PERSPECTIVES** is published by the Research Department of the Federal Reserve Bank of Chicago. The views expressed are the authors' and do not necessarily reflect the views of the management of the Federal Reserve Bank.

Single-copy subscriptions are available free of charge. Please send requests for single- and multiple-copy subscriptions, back issues, and address changes to the Public Information Center, Federal Reserve Bank of Chicago, P.O. Box 834, Chicago, Illinois 60690-0834, or telephone (312) 322-5111.

Articles may be reprinted provided the source is credited and the Public Information Center is sent a copy of the published material.

ISSN 0164-0682

# Temporal instability of the unemployment–inflation relationship

Robert G. King, James H. Stock,  
and Mark W. Watson



Econometric modeling of the relationship between inflation and unemployment has been a central topic in macroeconomics since the investigation of Phillips (1958), who documented a negative correlation between these variables in a half-century of U.K. data. Since the simultaneous occurrence of high inflation and high unemployment in the United States and other countries during the 1970s, there has been general agreement that this econometric relationship is unstable. Indeed, the instability has been so great that Lucas and Sargent characterized it as “econometric failure on a grand scale.”<sup>1</sup> This article summarizes some results from our recent work that documents various dimensions of this instability.<sup>2</sup>

We display econometric instability in three alternative and complementary ways. First we look at the simple correlation coefficient linking the unemployment rate and inflation, which initially attracted the attention of Phillips (1958) in U.K. data and Samuelson and Solow (1960) in U.S. data. We show that this correlation has changed in an important way since World War II, so that over the entire 1954–94 period the correlation is essentially zero. However, we also show that this largely reflects the changing trend behavior of the two series: When we eliminate trends and high-frequency components of inflation and unemployment so as to focus on the business cycle behavior of the two series, we find that there has been a remarkably stable negative correlation.

This combined pattern of stability and instability suggests the value of investigating the stability and performance of inflation and unemployment forecasting rules over various sample periods and horizons. Accordingly, our second approach is to investigate instability in the parameters of a reduced-form bivariate forecasting model—a vector autoregression (VAR)—for the two series. We document instability in the parameters of the forecasting model, particularly for the coefficients in the inflation equation. However, a closer examination suggests that this statistically significant time variation in parameters has a relatively small effect on the accuracy of forecasts.<sup>3</sup>

The inflation-unemployment relationship in major macroeconomic models is governed in large part by the econometrics of the “price equation” or “wage-price block,” which specifies that the inflation rate is a negative function of the level of unemployment. Typically, these specifications imply that there is a rate of unemployment at which inflation is stable—a nonaccelerating inflation rate of unemployment (NAIRU), or “natural” rate of unemployment.<sup>4</sup> One potential

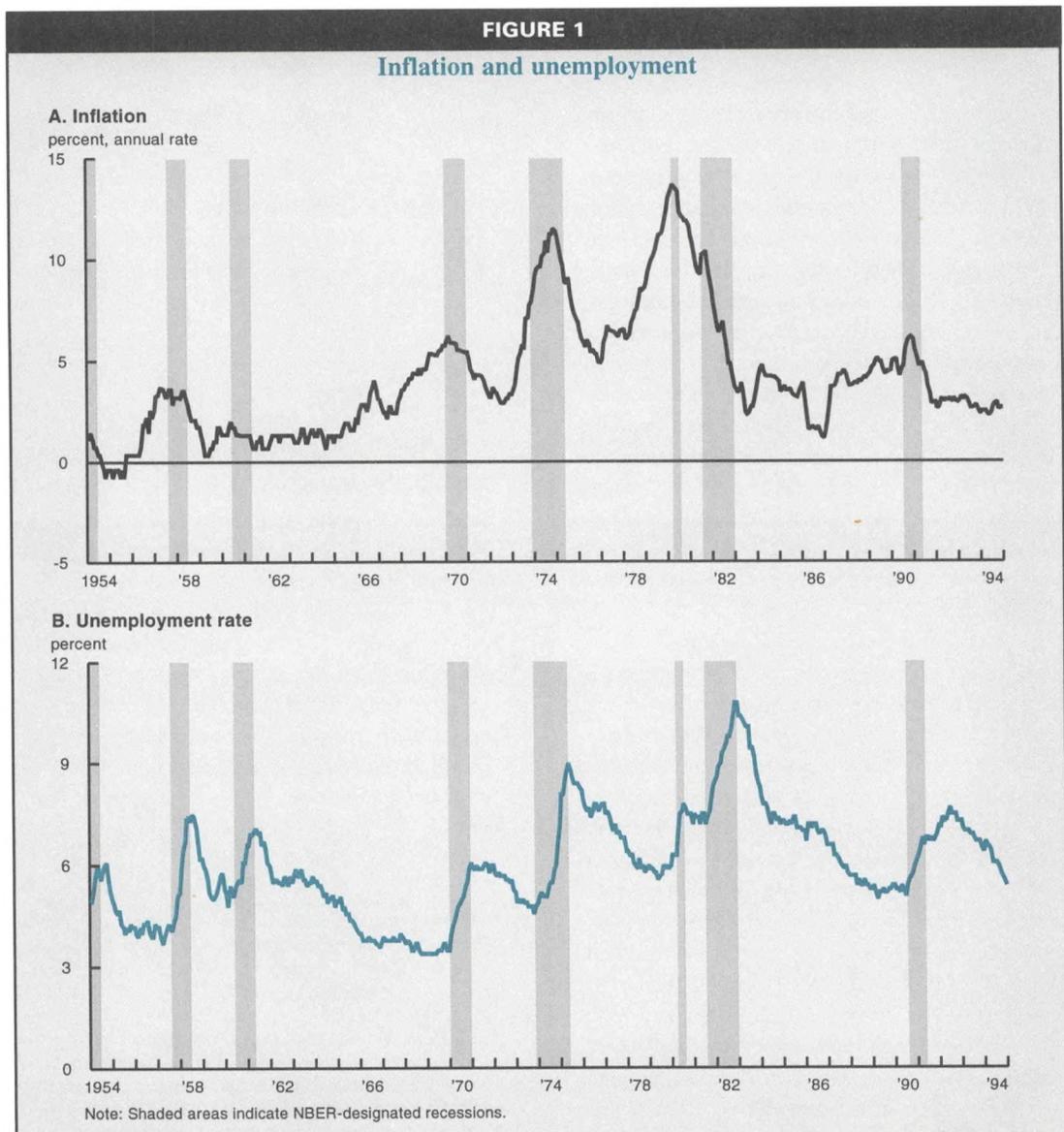
---

Robert G. King is a professor of economics at the University of Virginia and a consultant to the Federal Reserve Bank of Richmond. James H. Stock is a professor of public policy at the Kennedy School of Government, Harvard University. Mark W. Watson is a professor of economics at Northwestern University and a consultant to the Federal Reserve Bank of Chicago. The authors wish to thank Charles Evans, Robert Gordon, Ken Kuttner, and Doug Staiger for comments.

source of instability in the observed unemployment-inflation relationship in major econometric models is time variation in the NAIRU; our third approach is to study the extent of this form of instability in a conventional price equation. We find that while there may be time variation in the NAIRU, it is very hard to estimate precisely the extent of this time variation and the level of the NAIRU at any point in time. For example, in the econometric model with the most precise NAIRU estimate, the 95 percent confidence interval for the current value of the NAIRU ranges from 4.9 to 7.6 percentage points.<sup>5</sup>

### Instability in the correlation between the unemployment rate and inflation

Figure 1 plots the unemployment and inflation data used in this paper over the 1954-94 period.<sup>6</sup> The characteristics of the data evident in the figure suggest that a single correlation coefficient will do a poor job summarizing the relation between the two series. For example, both inflation and the unemployment rate are lower in the first half of the sample than in the second half. This suggests a positive correlation between the series. On the other hand, during business cycle recessions (shown as shaded areas in the figure), inflation tends to fall and unemployment increases; the opposite occurs



during business cycle expansions. This suggests a negative correlation between the series of the sort summarized in the Phillips curve.

Table 1 shows the correlation coefficient calculated over the entire sample period and over two subsamples. When computed over the entire sample period, the two forces discussed in the last paragraph essentially cancel one another, yielding a correlation coefficient of 0.08. However, when we crudely adjust the data for time-varying trends by splitting the sample, the negative business cycle correlation is apparent. The sample correlations over the first half and second half of the sample are  $-0.28$  and  $-0.26$ , respectively.

Figure 2 is a more careful attempt to extract the time-varying trends from the series. Panel A shows the results of filtering the data to isolate those components with cyclical periodicity greater than 8 years. These represent the slowly varying trend components of the data. Panel B shows the results of filtering the data to isolate those components with cyclical periodicity between 6 months and 8 years.<sup>7</sup> This isolates the components of the series associated with business cycle variability. As is evident from the figure, there is no apparent

Sample period	Raw data	Trend component	Cyclical component
1954-94	0.08	0.43	-0.61
1954-73	-0.28	-0.12	-0.60
1974-92	-0.26	0.01	-0.64

Notes: The raw data correspond to the unemployment rate and the monthly percentage change in prices. The trend and cyclical components are the bandpass-filtered values of the raw data using a trend (with periods > 96 months) and a business cycle (with periods between 6 and 96 months) filter. See text for additional details.

systematic relationship between the trend components, but there is a clear and apparently stable negative relationship between the business cycle components. These correlations are summarized in the last two columns of table 1. The sample correlation between the trend components of the table is unstable:  $-0.12$  in the first half of the sample,  $0.01$  in the second half, and  $0.43$  over the entire sample period. On the other hand, the correlation of the business cycle components is remarkably stable:  $-0.60$  in the first half,  $-0.64$  in the second half, and  $-0.61$  over the entire sample.

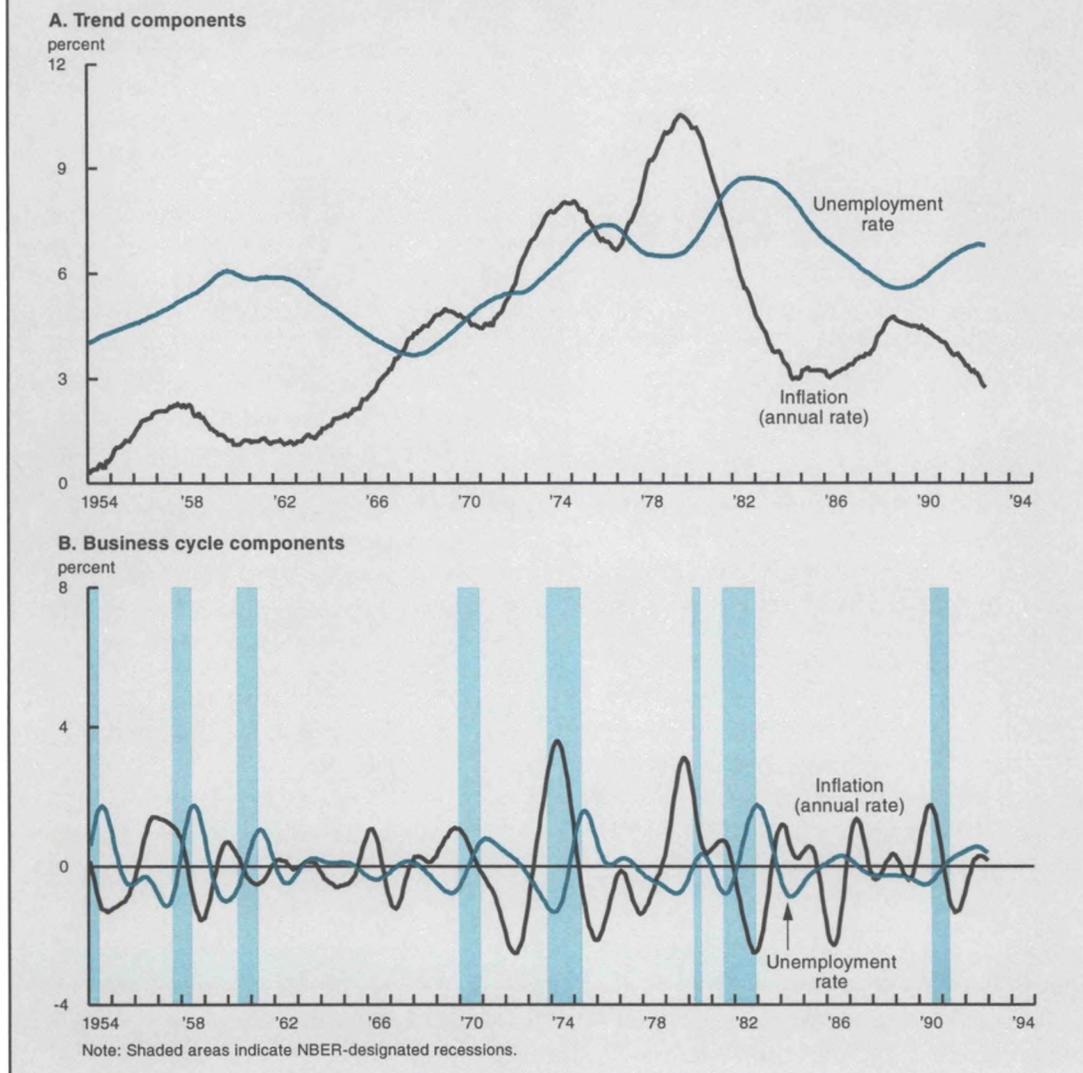
This analysis suggests that it may be possible to uncover a stable forecasting relationship linking inflation and the unemployment

<b>A. Lag lengths chosen by BIC</b>					
Forecasting equation	GC	Chow(73:M12)	QLR	QLR <sub>date</sub>	Lag length
$\Delta\pi$	3.74 (0.00)	3.46 (0.00)	3.66 (<0.01)	1974:M8	8
$\Delta u$	1.85 (0.16)	0.76 (0.58)	4.08 (0.03)	1960:M1	2
<b>B. Fixed lag length</b>					
Forecasting equation	GC	Chow(73:M12)	QLR	QLR <sub>date</sub>	Lag length
$\Delta\pi$	3.55 (0.00)	2.24 (0.00)	2.60 (<0.01)	1974:M8	12
$\Delta u$	1.37 (0.17)	1.81 (0.01)	2.46 (<0.01)	1960:M3	12

Notes: GC refers to the Granger-causality F-statistic testing the null hypothesis of no Granger-causality. Chow (73:M12) is the Chow test for instability in all of the coefficients with a single break allowed in 1973:M12. QLR is the Quandt likelihood ratio statistic (expressed here in F-statistic form) for a single break at an unknown time period between 1960:M1 and 1988:M12 (the middle 70 percent of the sample). QLR<sub>date</sub> is the date at which the QLR statistic is maximized. Lag length refers to the number of lags included in the regression. The numbers in parentheses are p-values for the statistics under the null hypothesis.

FIGURE 2

Trend and cyclical components



rate, but that the specification must focus on the shorter-run variation in the series and mask longer-run trend variation.

**Instability in the bivariate VAR**

Trending behavior in the series can be masked if the forecasting model is specified using the first differences of the variables.<sup>8</sup> Accordingly, we consider forecasts constructed from a bivariate VAR that incorporates  $\Delta\pi_t$  and  $\Delta u_t$ , where  $\pi_t$  is the inflation rate and  $u_t$  is the unemployment rate.

Table 2 summarizes a variety of F-statistics constructed from each of the equations in the model. Panel A of the table shows the results for a VAR with lag lengths chosen by

the Bayesian Information Criterion (BIC, a standard model-selection model), and panel B shows results when the lag lengths are set equal to 12 (a common specification for monthly models). The first column of the table identifies the equation considered, and the second column reports the Granger-causality F-statistic for the equation. This statistic tests the hypothesis that the other variable does not help predict the dependent variable in this equation. For example, from panel A, the Granger-causality F-statistic in the  $\Delta\pi$  equation is 3.74, and this tests the null hypothesis that lags of unemployment do not help predict future inflation. Since the p-value (shown in parentheses) is very small, the test suggests

that lags of unemployment do help predict future inflation. On the other hand, the Granger-causality statistic for the  $\Delta u$  equation is 1.85 with a p-value of 0.16. This suggests that lags of inflation are not statistically significant predictors of future changes in the unemployment rate. These results obtain for BIC and fixed-lag-length specifications. Thus, unemployment helps predict future inflation, but inflation does not help predict future unemployment.

The next three columns of table 2 investigate the temporal stability of the regressions. The column labeled Chow(73:M12) is the standard Chow test for the stability of the regression, allowing for a break date in 1973:M12. Evidently, the inflation equation is statistically unstable. The unemployment equation is stable if only a few lags are included (panel A), but there is some evidence of instability at longer lags (panel B). The Chow test can be criticized in this context because the choice of the break date (1973:M12) is not statistically independent of the data: Many have argued that the relationship between unemployment and inflation changed around 1973, and this “sample selection” invalidates the standard Chow testing procedure. The next statistic labeled QLR (for Quandt likelihood ratio) overcomes this problem by endogenizing the break date. This statistic is calculated as the largest of the standard Chow F-tests over all possible break dates between 1960:M1 and 1988:M12. The critical value for the statistic now explicitly accounts for the sample selection associated with choosing the largest of this sequence of statistics. Using the QLR test, one sees instability in both the  $\Delta \pi$  and  $\Delta u$  equations regardless of the lag-length specification. The column labeled  $QLR_{date}$  shows the break date that yielded the largest F-statistic. If there is only one break in the process, this serves as an estimate of that break date. Notice that the inflation process appears to have undergone a shift in the middle of the sample (in 1974), while the apparent shift in the unemployment process occurred much earlier (in 1960).

While the evidence suggests some instability in the inflation and unemployment processes, table 2 says little about the magnitude of the shift. This is addressed in table 3, which shows how forecasting performance is affected by three factors: the forecasting horizon, the sample period used to estimate the model, and

the sample period for the forecasts. For example, panel A shows results for 1-month-ahead forecasts. The first row of each panel shows results for the forecasting model estimated over the 1954-73 sample period, the second row shows results for the model estimated over the 1974-94 sample period, and the final row shows results for the model estimated over the entire sample period, 1954-94. For each of these models, root mean square forecast errors (RMSEs) are shown for forecasts constructed for the 1954-73 period (column 2), the 1974-94 period (column 3), and the 1954-94 period (column 4). Panel B of the table shows the same set of results for 6-month-ahead forecasts, and so on.

Two conclusions emerge from the table. First, using different coefficients over different forecast periods has relatively little effect on forecast accuracy. For example, consider an experiment in which the 1954-73 model is used to construct forecasts over the 1954-73 period, and the 1974-94 model is used to construct forecasts over the 1974-94 period. From panel A, this procedure produces one-month-ahead full-sample RMSEs of about 0.18 for the unemployment rate and 0.21 for the inflation rate. If the 1954-73 model is used for the entire sample period, these RMSEs increase only slightly to about 0.19 and 0.24, respectively. Similarly, if the 1974-94 model is used over the entire sample period, the RMSEs increase to about 0.19 and 0.22, respectively. From the other panels, this basic result holds for other forecast intervals as well. Thus there is only a small gain from changing the coefficients of the forecasting model over different forecast periods. The second conclusion that follows from table 3 is that inflation became more difficult to forecast over long horizons in the second half of the sample. For example, at the twelve-month horizon, the in-sample RMSE from 1954-73 is about 1.13, and this increases to 1.86 in the second period.

In summary, the results presented in tables 2 and 3 suggest that statistically significant changes occurred in the unemployment-inflation processes during the sample period. This change had little effect on the best choice of a bivariate forecasting model but did have an effect on the accuracy of inflation forecasts. Regardless of the forecasting model used, inflation became more difficult to forecast in the second half of the sample.

TABLE 3

**Root mean square error for models estimated  
over different sample periods**

**A. 1-month-ahead forecast error RMSE**

Estimation period	Forecasting period					
	1954:M1-73:M12		1974:M1-94:M12		1954:M1-94:M12	
	Unemp.	Infl.	Unemp.	Infl.	Unemp.	Infl.
1954:M1-73:M12	0.177	0.213	0.205	0.262	0.193	0.235
1974:M1-94:M12	0.208	0.250	0.174	0.213	0.192	0.224
1954:M1-94:M12	0.185	0.226	0.181	0.220	0.184	0.224

**B. 6-month-ahead forecast error RMSE**

Estimation period	Forecasting period					
	1954:M1-73:M12		1974:M1-94:M12		1954:M1-94:M12	
	Unemp.	Infl.	Unemp.	Infl.	Unemp.	Infl.
1954:M1-73:M12	0.566	0.554	0.725	1.007	0.658	0.811
1974:M1-94:M12	0.657	0.699	0.607	0.901	0.635	0.796
1954:M1-94:M12	0.591	0.589	0.632	0.929	0.617	0.796

**C. 12-month-ahead forecast error RMSE**

Estimation period	Forecasting period					
	1954:M1-73:M12		1974:M1-94:M12		1954:M1-94:M12	
	Unemp.	Infl.	Unemp.	Infl.	Unemp.	Infl.
1954:M1-73:M12	0.964	1.126	1.271	2.052	1.141	1.643
1974:M1-94:M12	1.120	1.328	1.095	1.864	1.116	1.585
1954:M1-94:M12	0.988	1.158	1.151	1.910	1.083	1.585

**D. 24-month-ahead forecast error RMSE**

Estimation period	Forecasting period					
	1954:M1-73:M12		1974:M1-94:M12		1954:M1-94:M12	
	Unemp.	Infl.	Unemp.	Infl.	Unemp.	Infl.
1954:M1-73:M12	1.196	2.287	1.754	5.214	1.525	4.126
1974:M1-94:M12	1.300	2.750	1.661	4.680	1.512	3.904
1954:M1-94:M12	1.199	2.270	1.705	4.844	1.498	3.904

Notes: The entries in the table refer to the root mean square forecast error for unemployment and inflation for the forecasting period shown. The inflation forecast corresponds to price inflation over the forecast period and is expressed as the annual percentage rate. For example, the 6-month forecast error for inflation dated 1954:M1 is the forecast error for the rate of change in prices over the period 1953:M7-54:M1 using the forecast constructed in 1953:M7 and expressed in percent at an annual rate. The forecasts were formed using VAR(12) models (including a constant) estimated over the periods given in the first column of the table.

**Instability in estimates of the NAIRU**

One important characteristic of the forecasting relation linking unemployment and inflation is the NAIRU—that value of the unemployment rate, which if maintained, would forecast no long-run changes in the inflation rate. This NAIRU can be estimated

as the parameter  $\bar{u}$  in a regression specification of the form

$$(1) \Delta\pi_t = \sum_{i=1}^p \beta_i (u_{t-i} - \bar{u}) + \sum_{i=1}^k \gamma_i \Delta\pi_{t-i} + a_t,$$

where  $a_t$  is a regression error. If it is postulated that  $u_t = \bar{u}$  for  $\tau > t$ , and if the lag polynomial

$1 - \gamma_1 L - \dots - \gamma_k L^k$  is stable, then this equation produces long-run forecasts of  $\Delta\pi$  that are equal to zero, so that inflation is unchanging.<sup>9</sup>

Equation 1 differs from the VAR used above in one important way: The level of  $u$  enters equation 1, while the VAR is specified using the first difference of  $u$ . If the VAR is correctly specified (and we argue that it is, given the trend behavior in the unemployment rate evident in figure 2), then equation 1 can be correct only if the distributed lag of  $u$ 's entering the equation can be written entirely in terms of first differences of  $u$ . This is possible only if  $\sum_{i=1}^p \beta_i = 0$ . This constraint has important implications for estimation of the NAIRU. Notice that the NAIRU,  $\bar{u}$ , enters equation 1 only as  $\bar{u} \sum_{i=1}^p \beta_i$ . Thus, if  $\sum_{i=1}^p \beta_i = 0$ , then  $\bar{u}$  does not enter the equation, so that the inflation equation contains no information about  $\bar{u}$ . This implies that the value of the NAIRU is econometrically unidentified from equation 1. Alternatively, the NAIRU has no meaning in an equation when only changes in the unemployment rate help predict future inflation.

There are two ways around this criticism. The first is simply to assume that  $\sum_{i=1}^p \beta_i \neq 0$ , and use equation 1 to estimate  $\bar{u}$ . Since the  $\beta_i$ 's are estimated as part of this process, if  $\sum_{i=1}^p \beta_i = 0$ , then this will be true approximately for the estimates as well. This in turn will lead to estimates of  $\bar{u}$  that are very imprecise, which should be apparent from large standard errors for the estimate of  $\bar{u}$ . Equivalently, the problem should surface as wide confidence intervals for  $\bar{u}$ . An alternative is to specify equation 1 allowing the parameter  $\bar{u}$  to vary through time, capturing the time-varying trend in the unemployment data. This will obviate the need to first-difference  $u$  in the equation.

Here we use a model that incorporates both of these possibilities. Specifically we estimate a model of the form

$$(2) \Delta\pi_t = \sum_{i=1}^p \beta_i (u_{t-i} - \bar{u}_t) + \sum_{i=1}^k \gamma_i \Delta\pi_{t-i} + a_t$$

$$(3) \bar{u}_t = \bar{u}_{t-1} + e_t,$$

**TABLE 4**  
**Estimated parameters for time-varying NAIRU model**

Parameter	$\sigma_e$			
	0.00	0.05	0.10	0.15
$u(-1)$	-0.724 (0.659)	-0.723 (0.638)	-0.728 (0.635)	-0.727 (0.630)
$u(-2)$	-0.377 (1.433)	-0.373 (1.375)	-0.361 (1.362)	-0.341 (1.342)
$u(-3)$	0.729 (1.585)	0.721 (1.515)	0.704 (1.500)	0.679 (1.483)
$u(-4)$	0.115 (1.481)	0.110 (1.425)	0.102 (1.429)	0.079 (1.433)
$u(-5)$	0.190 (1.585)	0.194 (1.521)	0.223 (1.521)	0.267 (1.521)
$u(-6)$	1.529 (1.617)	1.532 (1.557)	1.525 (1.558)	1.493 (1.554)
$u(-7)$	-0.831 (1.538)	-0.822 (1.481)	-0.816 (1.482)	-0.792 (1.478)
$u(-8)$	-0.478 (1.522)	-0.481 (1.465)	-0.480 (1.468)	-0.452 (1.466)
$u(-9)$	-2.215 (1.620)	-2.198 (1.549)	-2.191 (1.546)	-2.189 (1.539)
$u(-10)$	2.725 (1.532)	2.701 (1.492)	2.669 (1.490)	2.605 (1.484)
$u(-11)$	-1.267 (1.367)	-1.309 (1.321)	-1.321 (1.319)	-1.320 (1.319)
$u(-12)$	0.560 (0.652)	0.605 (0.626)	0.633 (0.628)	0.663 (0.634)
$\Delta\pi(-1)$	-0.009 (0.049)	-0.015 (0.048)	-0.020 (0.051)	-0.029 (0.057)
$\Delta\pi(-2)$	0.024 (0.057)	0.020 (0.055)	0.018 (0.057)	0.014 (0.060)
$\Delta\pi(-3)$	-0.124 (0.055)	-0.124 (0.053)	-0.120 (0.054)	-0.114 (0.056)
$\Delta\pi(-4)$	-0.010 (0.055)	-0.101 (0.053)	-0.095 (0.054)	-0.086 (0.055)
MA(1)	-0.831 (0.037)	-0.827 (0.039)	-0.838 (0.046)	-0.854 (0.060)
$\sigma_e$	0.278 (0.104)	0.260 (0.123)	0.247 (0.136)	0.226 (0.136)
Log-likelihood	-711.73	-712.19	-712.74	-713.20

Notes: These are estimates of the parameters in equations 2 through 4 of the text. They are estimated by Gaussian maximum likelihood using data from 1953:M1-94:M12.

where  $e_t$  is an *iid* error term with a mean of zero and standard deviation of  $\sigma_e$ . Since  $\bar{u}_t = \bar{u}_0 + \sum_{s=1}^t e_s$ , then  $\bar{u}_t$  is constant when  $\sigma_e = 0$ , so that equation 2 collapses to equation 1. When  $\sigma_e \neq 0$ , the model allows the NAIRU to change by  $e_t$  in each period. To complete the model, we set the lag lengths as  $p=12$  and  $k=4$ , and allow the error term  $a_t$  to follow an MA(1) process:

$$(4) \quad a_t = \varepsilon_t - \Theta \varepsilon_{t-1},$$

with  $\varepsilon_t$  an *iid* error term with mean zero, variance  $\sigma_\varepsilon^2$ , and uncorrelated with  $e_t$  at all leads and lags. (The MA(1) specification turned out to be a parsimonious way to model persistence in the inflation process.) The model summarized by equations 2-4 is a standard stochastic time-varying parameter regression model that can be estimated using Gaussian maximum likelihood methods as described in Harvey (1989) or Hamilton (1994).<sup>10</sup>

Table 4 shows the results obtained by estimating the model over a range of values of  $\sigma_e$ , the time variation allowed in  $\bar{u}$ . Notice that the model with  $\sigma_e = 0$  results in the highest value of the log-likelihood and hence corresponds to the maximum likelihood estimate. However, models with larger values of  $\sigma_e$  produce log-likelihoods that are not significantly larger, at least using conventional rules of thumb.<sup>11</sup>

Figure 3 plots the estimates of  $\bar{u}$  produced by each of the models, together with the actual unemployment rate.<sup>12</sup> When  $\sigma_e = 0$ , the NAIRU is constant with an estimated value of 6.26 percent. As  $\sigma_e$  increases, more variation is apparent in the estimated values of  $\bar{u}$ . For example, when  $\sigma_e$  is 0.15 (the largest value considered), the estimates of  $\bar{u}$  vary from a high of 7.87 percent (in 1980:M1) to a low of 5.62 percent (in 1967:M2 and 1994:M12).

Table 5 presents estimates of  $\bar{u}$  at five intervals for the each of the models. Also shown are the estimated standard errors of the estimates.<sup>13</sup> The most striking feature of this table is the size of these standard errors. For example, if it assumed that the NAIRU is constant, then the 95 percent confidence interval is 4.9 to 7.6 percentage points. If it is assumed that the NAIRU has significant time variation ( $\sigma_e = 0.15$ ), then the 95 percent confidence interval for the NAIRU in 1994:M12 is 2.7 to 8.5 percentage points. The source of this uncertainty in the estimated value of  $\bar{u}$  is the very small estimated value of  $\sum_{i=1}^p \beta_i$ . This is estimated as  $-0.04$  in the  $\sigma_e = 0$  model, and does not change appreciably as  $\sigma_e$  is allowed to take on non-zero values.<sup>14</sup>

In summary, while the data may be characterized by a model with a time-varying NAIRU, the value of this NAIRU is estimated very imprecisely from the data.

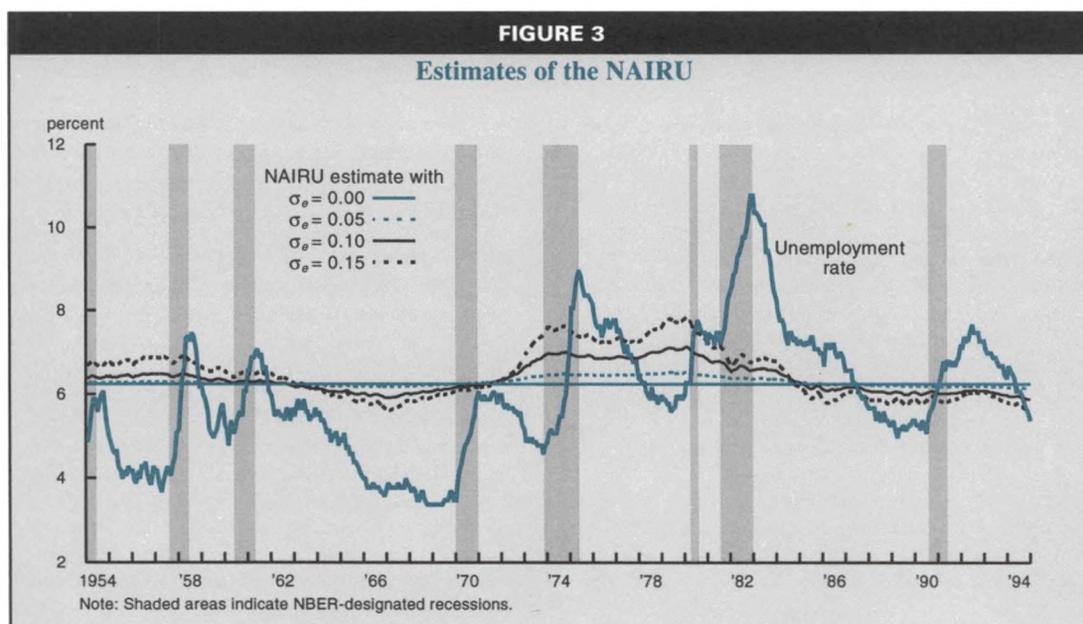


TABLE 5

## Selected values of the NAIRU

Date	Unemployment rate	NAIRU with $\sigma_e$			
		0.00	0.05	0.10	0.15
1954:M1	4.90	6.26 (0.67)	6.29 (1.28)	6.40 (1.99)	6.68 (2.63)
1960:M1	5.20	6.26 (0.67)	6.28 (1.17)	6.33 (1.72)	6.53 (2.07)
1965:M1	4.90	6.26 (0.67)	6.22 (1.10)	6.07 (1.58)	5.97 (1.80)
1970:M1	3.90	6.26 (0.67)	6.27 (1.04)	6.18 (1.46)	6.09 (1.64)
1975:M1	8.10	6.26 (0.67)	6.50 (0.98)	6.96 (1.30)	7.51 (1.47)
1980:M1	6.30	6.26 (0.67)	6.55 (0.95)	7.13 (1.23)	7.87 (1.38)
1985:M1	7.30	6.26 (0.67)	6.31 (0.94)	6.29 (1.17)	6.17 (1.30)
1990:M1	5.30	6.26 (0.67)	6.22 (0.95)	6.06 (1.17)	5.88 (1.33)
1994:M12	5.40	6.26 (0.67)	6.16 (0.98)	5.89 (1.24)	5.62 (1.48)

Notes: These are estimates of the NAIRU computed using the Kalman smoother applied to the model 2 through 4 with parameter values taken from table 4. The standard errors (in parentheses) were computed following Hamilton (1986).

## Conclusion

In this article we investigated the temporal stability of the relationship between unemployment and inflation. We documented both stable and unstable characteristics of the relationship. The correlation between the two series over the business cycle is remarkably stable, but there appears to be no stable relationship over long horizons. We uncovered statistically significant changes in the forecasting relationship between the variables. However, splitting the sample to allow changes in

the coefficients did little to improve the forecasts. The major unstable characteristic of the forecasting relationship is an increase in the long-horizon variance of inflation. Finally, we constructed models that allowed time variation in the NAIRU. The resulting estimates of the NAIRU were very imprecise, which is consistent with the theory that future inflation is better predicted by changes in the unemployment rate than by the size of the unemployment gap (the difference between unemployment and the NAIRU).

## NOTES

<sup>1</sup>Lucas and Sargent (1979).

<sup>2</sup>King and Watson (1994) and Staiger, Stock, and Watson (1995).

<sup>3</sup>The results reported in the first two sections are abstracted from King and Watson (1994, sections 2 and 5.1–5.2), whose main focus is a different type of econometric instability: the stability of structural models of the unemployment-inflation trade-off with respect to specific econometric identifying assumptions. In addition, King and Watson discuss the stability of estimated structural trade-offs across different periods.

<sup>4</sup>The natural unemployment rate concept is intimately linked to the notion of a vertical long-run Phillips curve as explained in Phelps (1967) and Friedman (1968). However, as Modigliani and Papademos (1976) argue, NAIRU is an interesting concept even in models without a vertical long-run Phillips curve trade-off.

<sup>5</sup>The work described in this section reports preliminary results from Staiger, Stock, and Watson (1995).

<sup>6</sup>The unemployment rate is for all workers 16 years and older, seasonally adjusted. The inflation rate is computed from the all-items Consumer Price Index for urban consumers. Letting  $P_t$  denote the value of this price index at time  $t$ , the inflation rate plotted in figure 1A is the annual inflation rate:  $100 \cdot \ln(P_t/P_{t-12})$ . Much of the analysis in this article is carried out using the monthly inflation rate (expressed in percent at an annual rate) defined as  $\pi_t = 1,200 \cdot \ln(P_t/P_{t-1})$ . Unemployment and Consumer Price Index data are from Citibank (1994).

<sup>7</sup>Specifically letting  $x_t$  denote the raw series, panel A of figure 2 plots  $y_t = A(L)x_t$ , where the spectral gain of  $A(L)$  is approximately equal to 1 for periods greater than 96 months and approximately equal to 0 for other periods; the spectral phase of  $A(L)$  is 0.  $A(L)$  is a two-sided 24-term lag polynomial constructed as the optimal approximate bandpass filter using the procedure developed in Baxter and King (1994). The filter for panel B is constructed analogously as the optimal approximate 6-month to 96-month bandpass filter.

<sup>8</sup>Formally, this amounts to modeling the data as “integrated processes,” so that they exhibit stochastic growth, but not “co-integrated,” so that each series has its own distinct long-run trend.

<sup>9</sup>The essentials of this method for estimating the NAIRU or the natural unemployment rate can be traced back to Gordon (1972). In that original work and much subsequent work, Gordon has investigated nonlinearities, demographic and other shifts in the relationship, and their implied effect on estimates of the natural rate. Equation 1, however, captures the essential features of the relationship between the NAIRU and inflation.

<sup>10</sup>The econometric model represented by equations 2 through 4 is one version of Cooley and Prescott’s (1973) adaptive regression model. Its use as a forecasting tool is surveyed in Engle and Watson (1988), and both Gordon (1994) and Staiger and Stock (1994) discuss the model’s potential for estimating the natural unemployment rate.

<sup>11</sup>In typical situation, 2 times the log-likelihood ratio is approximately distributed as a  $\chi^2$  random variable. Since the models in table 4 differ by the choice of one parameter, and since the 95 percent critical value for the  $\chi^2_1$  is 3.84, this suggests log-likelihoods must differ by more than 1.92 to be statistically significant.

<sup>12</sup>These estimates are computed from a Kalman smoother, conditional on the parameter estimates shown in table 4.

<sup>13</sup>These standard errors were calculated using the procedure developed in Hamilton (1986).

<sup>14</sup>One possible modification to increase the precision of  $\bar{u}_t$  is to use the unemployment equation in addition to the inflation equation to identify the NAIRU. This would explicitly link the NAIRU to the stochastic trend in the observed unemployment rate. Kuttner (1994) estimates “potential output” in such a framework.

---

## REFERENCES

**Baxter, M., and R. G. King**, “Measuring business cycles: Approximate bandpass filters for economic time series,” manuscript, University of Virginia, 1994.

**Citibank**, “CITIBASE: Citibank Economic Database (machine-readable magnetic data file), 1946-Present,” New York: Citibank, N.A., 1994.

**Cooley, T., and E. Prescott**, “An adaptive regression model,” *International Economic Review*, Vol. 14, 1973, pp. 364-371.

**Engle, R. G., and M. W. Watson**, “The Kalman filter: Applications to forecasting and rational expectations models,” in *Advances in Econometrics, Fifth World Congress*, T. Bewley (ed.), Cambridge, U.K.: Cambridge University Press, 1988.

**Friedman, M.**, “The role of monetary policy,” *American Economic Review*, Vol. 58, No. 1, March 1968, pp. 1-17.

**Gordon, R. J.**, “Wage-price controls and the shifting Phillips curve,” *Brookings Papers on Economic Activity*, Vol. 2, 1972, pp. 385-430.

\_\_\_\_\_, personal communication to Mark Watson, November 17, 1994.

**Hamilton, J. D.**, “A standard error for the estimated state vector of a state-space model,”

*Journal of Econometrics*, Vol. 33, No. 3, 1986, pp. 387-397.

\_\_\_\_\_, *Time-Series Analysis*, Princeton, NJ: Princeton University Press, 1994.

**Harvey, A. C.**, *Forecasting, Structural Time-Series Models, and the Kalman Filter*, Cambridge, U.K.: Cambridge University Press, 1989.

**King, R. G., and M. W. Watson**, “The postwar U.S. Phillips curve: A revisionist econometric history,” *Carnegie-Rochester Conference on Public Policy*, Vol. 41, December 1994, pp. 157-219.

**Kuttner, K. N.**, “Estimating potential output as a latent variable,” *Journal of Business and Economic Statistics*, Vol. 12, No. 3, July 1994, pp. 361-368.

**Lucas, R. E., Jr., and T. J. Sargent**, “After Keynesian macroeconometrics,” in *After the Phillips Curve: Persistence of High Inflation and Unemployment*, Federal Reserve Bank of Boston, Conference Series, No. 19, 1979.

**Modigliani, F., and L. Papademos**, “Monetary policy for the coming quarters: The conflicting views,” *The New England Economic Review*, Federal Reserve Bank of Boston, Vol. 76, March/April 1976, pp. 2-35.

**Phelps, E.**, "Phillips curves, expectations of inflation, and optimal inflation over time," *Economica*, Vol. 34, No. 135, August 1967, pp. 254-281.

**Phillips, A. W.**, "The relation between unemployment and the rate of change of money wages in the United Kingdom, 1861-1957," *Economica*, Vol. 25, No. 100, November 1958, pp. 283-299.

**Samuelson, Paul A., and Robert M. Solow**, "Analytical aspects of anti-inflation policy," in

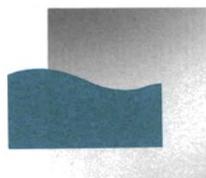
*American Economic Review, Papers and Proceedings*, Vol. 50, 1960, pp. 177-194.

**Staiger, D., and J. H. Stock**, "Time-varying parameter estimates of the natural rate of unemployment," research memorandum, Kennedy School of Government, Harvard University, 1994.

**Staiger, D., J. H. Stock, and M. W. Watson**, "What is the natural rate of unemployment?" manuscript in progress, Kennedy School of Government, Harvard University, 1995.

# An analysis of the effect of Chicago school reform on student performance

Thomas A. Downes and  
Jacquelyn L. Horowitz



In the past two decades, policymakers around the country have responded to growing public dissatisfaction with the quality of America's public schools by implementing various reforms of the delivery of schooling services. Some of the most dramatic of these reforms have been implemented in Chicago. The need for change was clear. A 1985 report on the Chicago public schools showed that less than two-thirds of all entering ninth graders would graduate from a Chicago high school, and only about one-third of those who did would read at the national average.<sup>1</sup> In 1987, on the American College Test (ACT) Chicago students averaged a score of 13.9 out of a possible 36.0, compared to an Illinois average of 18.9. These scores placed Chicago's high schools in the lowest 5 percent in the nation.<sup>2</sup> Chicago schools desperately needed help. The attempt at change came on December 2, 1988, with the passage of the Chicago School Reform Act.

The Chicago reform presents an interesting case for many reasons. Primary is the fact that the centerpiece of the reform is the shift of control of much of the governance of individual schools from the central administration to local governing bodies. Never before has such extensive decentralization been attempted. One observer called it "perhaps the most significant and potentially far-reaching school improvement strategy yet attempted in a large school district in the United States."<sup>3</sup> The national attention Chicago's school reform has received undoubtedly means that it will

influence reform efforts elsewhere. For this reason, it is especially important to sort out the lessons to be learned from it.

Chicago's school reform has been studied extensively by such groups as the Consortium on Chicago School Research, the Chicago Panel on Public School Policy and Finance, and Designs for Change (a Chicago educational consulting firm). However, these groups have focused on a limited number of elementary schools, on a certain year, or on specific issues such as dropout rates or local school council meetings.<sup>4</sup> Bryk et al. analyzed the changes in student performance between pre-reform and post-reform years, but their analysis neither considered what attributes of schools and communities influenced the success of reform efforts nor controlled for secular trends in student performance that might result in overstatement or understatement of reform effects.<sup>5</sup>

In this article, we supplement these earlier analyses by examining student performance in Chicago public schools for the school years 1987-88 through 1992-93, two pre-reform years and four post-reform years. We relate post-reform changes to attributes of each

---

Thomas A. Downes is an assistant professor of economics at Tufts University. When the bulk of the work on this paper was completed, Jacquelyn L. Horowitz was a student in Northwestern University's program in the Mathematical Methods in the Social Sciences. A slightly different version of this article was originally presented at the Chicago Fed's Conference on Midwest Approaches to School Reform in October 1994.

school and its surrounding community. We also control for trends in student performance common to all public schools in Illinois.

The benefits of the Chicago reform can be assessed in many ways. The most common means of evaluating the success of any educational program is to examine its effects on specific student outcomes. This article focuses on students' cognitive outcomes and graduation rates. It attempts to answer two broad questions: 1) Has reform affected student outcomes and graduation rates positively or negatively to date? and 2) Are particular features of individual schools or community areas associated with differential effects of reform on student outcomes? For measures of student outcomes, we chose the Illinois Goals Assessment Program (IGAP) reading test scores, ACT scores, and graduation rates.

In the next section, we present some background on the arguments for decentralization of governance as a basis for school reform. We also discuss evidence from other cities that have implemented similar though more limited reforms. The third section sketches a history of school problems in Chicago in the 1980s and gives an overview of the Chicago School Reform Act. In this section, we also consider whether it will be possible for reformers in other cities to be guided by the Chicago experience. The fourth section presents the data and methods we used in our analysis.

In section five, we discuss the results of two modes of analysis, assessment of means and regression analysis. The data suggest that at the elementary level, mean IGAP scores in Chicago have fallen since reform was implemented, even though statewide scores have changed little. Post-reform changes at the high school level have been more promising, with graduation rates and mean ACT scores showing relative increases. The results also indicate that changes in the performance of a school's students depend critically on the school's attributes. Moreover, a simple examination of changes in mean performance may understate post-reform success, since it ignores general trends in student performance. In addition, the regression results indicate that schools with a larger fraction of students with limited English proficiency or a larger fraction who are eligible for subsidized school lunch generally have seen decreases in student performance. In

light of the evidence from other jurisdictions on those factors that promote successful decentralization, one possible conclusion is that Chicago's schools could have improved more substantially if they had received additional resources. The results also indicate that, other things being equal, schools with large Hispanic populations have experienced greater post-reform success. Since Hispanic communities aggressively supported reform, this result is consistent with other research that has found that the gains from school-based management are directly related to the extent of community involvement. The final section of the paper outlines our conclusions and reviews the implications for future analyses and other reform efforts.

### **Decentralization and school-based management**

Most recent discussions about the reform of America's schools have involved the words *decentralization* and/or *school-based management*. These concepts have become the most popular ideas among proponents of school change.<sup>6</sup> A wide variety of decentralization and school-based management plans has been implemented in many school districts during the past few decades; among the most notable are the reforms in New York City, Miami, and Chicago.

From an economic perspective, decentralization is defined as the distribution of functions and powers from a single central authority to a number of local authorities.<sup>7</sup> One goal of this dispersion is to enable governments to adapt better to individual preferences. Another goal is to introduce more competition among governments, thereby lessening the threat of monopoly rule.<sup>8</sup> These goals make decentralization increasingly popular with the public. Not surprisingly, however, decentralization also carries potential costs. Among these are less equality in the distribution of public services and the inability of governments to exhaust all scale economies.<sup>9</sup>

The view that the benefits of decentralization exceed the costs has implicitly or explicitly motivated many of the recent reforms in the delivery of schooling services. "Underlying much current policy talk about school governance is a set of assumptions: that ineffective schools are to blame for the perceived lack of competitiveness of the U.S. economy; that

faulty governance produces faulty education. . . .<sup>10</sup> In the case of schools, decentralization can take a variety of forms. A large district can be divided into several smaller districts or effective control can be shifted to subdistricts

or to individual schools. School-based management is then decentralization taken to the extreme: It favors the individual school as the basic unit of management in a school district.<sup>11</sup> It has the added benefit of more effectively

## PROCEEDINGS NOW AVAILABLE

### Conference on Midwest Approaches to School Reform

On October 26 and 27, 1994, the Federal Reserve Bank of Chicago held a conference focusing on innovations in the finance and organization of elementary and secondary education. Researchers and practitioners came together to present and discuss papers evaluating the many reform experiments now underway in the Midwest and elsewhere. Proceedings from this conference are now available in book form or on-line and include the following contents:

#### Introduction

Thomas A. Downes and William A. Testa

#### Opening Remarks

Michael H. Moskow

#### Issues and Experience in Finance Reform

##### A Report on School Finance and Educational Reform in Michigan

Paul N. Courant, Edward Gramlich, and Susanna Loeb

##### School Finance Reform: The View from Wisconsin

Andrew Reschovsky and Michael Wiseman

##### Discussions:

##### *The Illinois Perspective*

Therese J. McGuire

##### *Issues in Indiana Public School Funding*

Larry DeBoer

##### *Comments and the Iowa Experience*

Thomas F. Pogue

#### The South Carolina Experience with Incentives

Garrett K. Mandeville

##### Discussion

Helen F. Ladd

#### Experiments and Experience with Choice

##### Chartered Public Schools:

##### A Brief History and Preliminary Lessons

Joe Nathan

##### Public School Choice in Minneapolis

Steven Glazerman and Robert H. Meyer

##### Who Chooses? Voucher and Interdistrict Choice Programs in Milwaukee

John F. Witte and Christopher A. Thorn

##### Discussion: School Choice and the Creaming Problem

Terry M. Moe

#### Directions in Management and Delivery in Urban School Systems

##### Redesigning Accountability at the Systemwide Level: The Politics of School Reform in Chicago

Kenneth K. Wong and Gail L. Sunderman

##### Discussion

James H. Lewis

##### An Analysis of the Effect of Chicago School Reform on Student Performance

Thomas A. Downes and Jacquelyn L. Horowitz

##### Discussion

John Q. Easton

#### Baltimore City Public Schools:

##### Experimenting with Private Operation

Sammis B. White

##### Discussion

Elaine Salinas

To request a copy of the Proceedings, call the Public Information Center at 312-322-5111 or write to

**Federal Reserve Bank of Chicago  
Public Information Center  
P.O. Box 834  
Chicago, IL 60690-0834**

These Proceedings are also available on Internet in Adobe Acrobat format via anonymous FTP at ftp.frbchi.org. See pub/edcon94/readme for further information.

involving individuals from outside the school system in the reform, since most school-based management plans include sharing power with teachers, principals, parents, community members, and students.<sup>12</sup> Usually, this sharing of power is accomplished through the creation of a decisionmaking council comprised of members of the different groups.

Increasing the sense of community is a central tenet of school-based management. "School-based management is more likely to produce a climate amenable to the creation of community than most other organizational designs. . . . School-based management is designed to bring community and school people to the same forum to talk about what they value most in education."<sup>13</sup> In school-based management, the sharing of responsibilities in decisionmaking by everyone affected raises the quality of the school. In other words, an effective school can be seen as a distinct community. Further, the success of the reform effort depends critically on the attributes of the community and on the extent of community involvement.

However, implementing school-based management plans can be problematic. Among the factors affecting success are the quality of leadership, the speed of change, the intensity of the opposition to reform, district size, level of support, and extent of preparations by the district.<sup>14</sup> School-based management requires considerable additional time, effort, and responsibility from school employees and community members, who may end up resenting the increased demand on their time. This can result in their abandoning some aspects of the original plan. Further, the nature of the cooperative decision-making process can delay the execution of some ideas. Finally, dissatisfaction may result from the "lack of clarity as to what school-based management is, how it is to be implemented, and who can do what."<sup>15</sup>

All of this discussion of possible constraints on the success of reform highlights the importance of community attributes and of the extent of community involvement. Since case studies indicate that a sense of community is essential for school-based management to succeed,<sup>16</sup> we included in our analysis of the Chicago reform characteristics that seem related to a sense of community, such as racial and

ethnic diversity and mobility between community areas.<sup>17</sup> The level of educational attainment of adults in the community areas was also included, since parents who have graduated from high school or college are more likely to be involved in and supportive of their children's education. Empirical evaluations of reform efforts that fail to control for community characteristics will provide little guidance on how to customize reform efforts in communities that might otherwise be less hospitable to school-based management.

Evidence from decentralization and school-based management plans in other large urban school systems also helped guide our analysis of the Chicago experience. New York City was among the first districts to enact a decentralization plan, in 1969. More recently, Miami has experimented with decentralization and school-based management. In both of these cities, as in Chicago, the central administration retained significant powers after decentralization. The major change imposed by New York's 1969 Decentralization Act was the creation of 32 community district boards responsible for curriculum and some personnel decisions. The central school board retained responsibility for tenure decisions and could suspend or remove a community board if its district failed to meet state or city education standards. Observers of this school reform and New York's scandal-plagued schools sometimes believe—mistakenly—that the experience has discredited the concept of school-based management. In fact, New York's reorganization moved power down only to the community level. This merely enlarged an intervening layer of the bureaucracy and did not redistribute power to principals, teachers, parents, or community members at the school level.<sup>18</sup>

The Miami plan was more successful because, while control over schools remained with the central district board, significant powers were also delegated to teachers and administrators of individual schools.<sup>19</sup> The experiences in New York and Miami suggest that a reform is more likely to institute positive change in systems that allow individual schools the flexibility to tailor resources, personnel, curriculum, and finances to their specific needs. To explore the validity of this conclusion in the Chicago context, we assessed

whether the relationship between indicators of student performance and indicators of flexibility in resource allocation changed after reform. Specifically, we tested the hypothesis that indicators of flexibility, such as average class size and minutes per day in academic subjects, explain more of the variation in performance in the post-reform period.

Another critical difference between the reform efforts in New York and Miami was the extent of local support for these efforts. In New York, reform was essentially forced on school personnel and community members by an external central government. This led to a high level of resentment within the school system. In contrast, teachers and administrators in Miami were the prime movers in formulating the reform plan.<sup>20</sup> The fact that Miami has had more success with reform than New York suggests that variables that correlate with the extent of community support for reform could be strongly related to the success of the reform. In the next section, we consider what these indicators of community support might be in the Chicago context.

### **School reform in Chicago: Antecedents and attributes**

Chicago school reform was the product of more than a decade of political and financial chaos within the Chicago public school system. A number of key factors profoundly influenced the nature of the reform. These included demographic change, racial tensions, political changes in Chicago, financial difficulties in the school system, and the growing power of the Chicago school board. The Chicago School Reform Act of 1988 completely altered the organization and governance of the city's public schools.

Many of the problems of urban public education in the U.S., and in Chicago specifically, stem from demographic changes within cities over the past few decades. As in most other major U.S. cities, the population of Chicago decreased between 1980 and 1990, from 3,005,072 to 2,783,726. The 7.4 percent decline in population was not uniform across all racial groups. The white population fell by 18.7 percent, while the African-American population decreased only 9.6 percent. On the other hand, the Hispanic population grew 29.3 percent, and the number of Asian and Pacific Islanders increased 42.8 percent.

Enrollment in the Chicago public schools followed a similar pattern. The total number of students fell from 458,497 to 408,714. White students decreased from 18.6 percent of the total student body to 11.8 percent, and African-Americans from 60.8 percent to 58 percent. Hispanics and Asian-Americans increased from 18.4 percent to 27.7 percent and from 2.1 percent to 2.9 percent of the student population, respectively.

These demographic changes influenced Chicago school reform in several ways. First, desegregation of schools, a major focus of the federal government during the late 1960s and early 1970s, was not a factor in the Chicago reform because the number of white students had decreased so significantly.<sup>21</sup> Second, Hispanics became active participants in the campaign for reform because of the overcrowding and general inattention of authorities to schools in primarily Hispanic neighborhoods. Finally, racial and ethnic struggles in Chicago changed from conflicts between African-Americans and whites to conflicts among African-Americans, whites, and Hispanics, thus altering the structure of politics in the city.

Changes in Chicago's political environment during the late 1970s and 1980s were also instrumental in the shaping of school reform. Racial tensions in the aftermath of Mayor Richard J. Daley's death resulted in increasing political involvement by the African-American population of Chicago. Many African-American leaders focused their attentions on the struggle for political power within the city. Because they were, therefore, less involved in the debate at the state level concerning the direction of the Chicago public schools, the way was paved for the growing involvement of the Hispanic population. The business community, the media, philanthropic organizations, and political leaders from outside Chicago also played a direct part in the development and passage of the Chicago School Reform Act. Because they were not major participants in structuring the reform, many African-American leaders and school personnel felt the reform was imposed, not implemented voluntarily.

The Chicago School Reform Act consisted of three main components: a set of goals to be met by each school by 1994, a reallocation of resources to the individual schools,

and the establishment of local school councils at every school. The ten explicit goals of the reform were:

- To assure that students achieve proficiency in reading, writing, mathematics, and higher-order thinking that equals or surpasses national norms;
- To assure that students attend school regularly and graduate from high school at rates that equal or surpass national norms;
- To assure that students are adequately prepared for further education and to aid students in making a successful transition to further education;
- To assure that students are adequately prepared for successful entry into employment and to aid students in making a successful transition to employment;
- To assure that students are, to the maximum extent possible, provided with a common learning experience that is of high academic quality and that reflects high expectations for all students' capacities to learn;
- To assure that students are better prepared to compete in the international market place by having foreign language proficiency and stronger international studies;
- To assure that students are encouraged in exploring potential interests in fields such as journalism, drama, art, and music;
- To assure that individual teachers are granted the professional authority to make decisions about instruction and the method of teaching;
- To assure that students are provided the means to express themselves creatively and to respond to the artistic expression of others through the visual arts, music, drama, and dance; and
- To assure that students are provided adequate athletic programs that encourage pride and positive identification with the attendance center and that reduce the number of dropouts and teenage delinquents.<sup>22</sup>

The primary goals among these were raising student achievement levels, attendance rates, and graduation rates to national norms. This article focuses on these measures as criteria for evaluating the reform's success.

The reform included two provisions that reallocated the resources of the school system

from the central bureaucracy to the individual schools. First, a cap was placed on administrative costs, and the central administration was downsized. Second, there were changes in the distribution of state Chapter 1 aid.<sup>23</sup> The reform act prohibited the use of state funds for any purposes outside local schools; it shifted the percentage of Chapter 1 aid targeted directly to the local schools from 60 percent to 100 percent over four years; and it freed funds by requiring that no Chapter 1 moneys could be used to support basic programs present in all schools.<sup>24</sup> As a result of this reallocation, each school received a considerable amount of new resources.

Most importantly, the Chicago School Reform Act created local school councils (LSCs) at each school in the district as the primary means to achieve the goals of the reform and to make use of the resources that were now under the control of the schools.<sup>25</sup> Each LSC is composed of six parents, two community members, two teachers, and the principal. In high schools, it also includes a non-voting student. All the members are elected every two years by the groups they represent. The three main responsibilities of the LSCs are to adopt a School Improvement Plan,<sup>26</sup> to adopt a budget based on a lump sum from the central office to implement that plan, and to decide whether to retain the current principal or hire a new one. The primary goal of the LSCs was to involve parents and community members in school governance. While other school-based management efforts elsewhere had increased the decisionmaking powers of teachers, Chicago's plan increased the powers of parents and community representatives.<sup>27</sup>

Within individual schools, principals were given a number of new responsibilities and duties. The principal became responsible for selecting all teachers and staff and for outlining a School Improvement Plan and a school budget for consideration by the LSC. A Professional Personnel Advisory Committee, consisting of classroom teachers and other school staff, was created to specifically advise the principal on the School Improvement Plan and budget. However, along with all this additional power, principals were made directly accountable to the LSCs. No longer enjoying job tenure, they became dependent on the LSC

for their employment, rather than the central office.<sup>28</sup> By the end of the second year of LSC principal selection, 38 percent of the schools had principals who had not been principals at their schools before the reform act was passed. That represented a significant turnover in local educational leadership.<sup>29</sup>

Because the scope of Chicago's school reform was limited in several ways, any lack of success should not be viewed as a general indictment of school-based management. The reform had very little impact on four key areas. First, it provided no additional funding even though the school system desperately needed more money to increase salaries, pay for capital improvements, and implement the reform.<sup>30</sup> Second, the reform allowed the Chicago School Finance Authority to retain and even increase its power.<sup>31</sup> Third, the mayor retained the right to appoint members of the central school board, thus continuing the tie between the public schools and mayoral politics. Fourth, the teachers' union still negotiated its contract with the central board rather than with the LSCs, thereby leaving the teacher hiring process essentially unchanged. Thus while the reform act radically changed a number of aspects about the schools, it "did not substantially alter the fundamental relationship between the school system and the most powerful political, financial, and educational interests in the city."<sup>32</sup>

Political and legal challenges to reform, plus the turmoil caused by the continuing financial problems of the school system, resulted in a sharp decline in the number of voters and candidates participating in the second LSC election in 1991.<sup>33</sup> This decline could have hampered the reform in two ways. As Mirel describes it, "low turnout undercuts the moral authority of the reforms by demonstrating that few people care enough to participate in the process. Second, small numbers of voters make it easy for political clubs or well-organized groups to control elections."<sup>34</sup>

Another factor that needs to be accounted for in evaluating the Chicago School Reform Act is the feeling of African-American leaders that the reform had been imposed on their community. The main impetus for the reform undeniably came from whites and Hispanics. African-Americans remained distant from, and sometimes even opposed to, reform. "They

argued that reformers had excluded them from negotiations with the state legislature. Some contended it was no coincidence that reformers launched an attack on the central administration only when its color had changed to black."<sup>35</sup> Since, as our discussion of the Miami and New York reforms indicated, this absence of community support and involvement could hinder the reform effort, the lack of support in the African-American community could limit the success of reform in schools with sizable African-American populations. Similarly, the support for reform in the Hispanic community could translate into more post-reform success in schools with large fractions of students who are Hispanic. One of the goals of the empirical work that follows is to test these predictions.

### Data and methodology

Most of the data used in this study came from the school report cards submitted to the state of Illinois by each school in the state. Data from 524 Chicago elementary and high schools and 879 school districts elsewhere in Illinois were available for the pre-reform school years 1987-88 and 1988-89 and the post-reform years 1989-90, 1990-91, 1991-92, 1992-93, and 1993-94.<sup>36</sup> The key variables in the data set were student outcome measures, such as attendance rates, graduation rates, IGAP reading and mathematics scores, and ACT test scores, and student and school characteristics, such as racial and ethnic composition, total enrollment, fraction of students with limited English proficiency, and fraction of students eligible for subsidized school lunch. Demographic data for the community areas of the city of Chicago, including measures of racial and ethnic diversity, the level of educational attainment of adults, family income, and mobility between communities, come from the 1990 Census. These data are used to control for differences across Chicago community areas. Each school in Chicago was assumed to draw from a population with the demographic characteristics of the community area in which it was located.<sup>37</sup>

The success and effectiveness of the reform cannot be judged solely by examining changes in student test performance within the district. Three limitations associated with the use of students' cognitive outcomes warrant mention. First, there are different ways to

measure student performance: measures linked directly to the content of instruction (central measures); measures, such as test scores, of cognitive outcomes (proximal measures); and measures of more general standards of self-concept (distal measures). Proximal measures, as are considered here, “may be insensitive to some program effects,”<sup>38</sup> while distal measures are influenced by the changes that follow reform. Thus, test scores may fail to measure important gains attributable to reform.<sup>39</sup> Second, changes in test scores may not be a direct result of changes in the school system. They may be due to statewide or nationwide trends in scores, variations in student characteristics, or modifications in the tests themselves. Finally, in discussions about a reform’s benefits, it is not always apparent what is meant by “effective.” Such ambiguity is inherent in the use of any data, not only student outcomes, to judge the impact of the Chicago reform.

Even with these limitations, we chose test scores as one of the criteria for this study simply because they are the measure most widely recognized and easily understood by educational groups and the general public. Test scores are also viewed as broad measures of achievement and lend themselves well to cross-program and cross-year comparisons.<sup>40</sup> The other measure we used was the school-by-school graduation rate.

Our decision to use school-level rather than student-level data can also be questioned. Previous researchers have clearly stated the risk of assessing reform using school-level data: Student mobility and changes in the ability of cohorts can result in school-level data implying that a successful reform has failed.<sup>41</sup> Nevertheless, we felt there were compelling arguments for using school-level data to evaluate the Chicago reform. Individual student scores on the IGAP tests were unavailable for most of the period considered in this article. Bryk et al. used scores on the Iowa Test of Basic Skills (ITBS) to evaluate reform.<sup>42</sup> However, the ITBS test instruments changed from year to year, making it impossible to attribute cross-time changes in performance to reform. The availability of IGAP scores for schools within and outside Chicago allows us to distinguish between general trends in scores and changes attributable to reform. Further, if reforms induce changes in a school’s student body (for example, by attracting students previously served by private schools), the resultant changes in mean test scores at the school level are properly attributed to the reform. Omitting new entrants to the system from the analysis would then result in an incorrect measure of the effect of reform.

To mitigate the problems associated with school-level data, we used measures of the

**TABLE 1**  
Selected student outcome measures, Chicago public schools

	Pre-reform years		Post-reform years				
	1987-88	1988-89	1989-90	1990-91	1991-92	1992-93	1993-94
IGAP reading score—grade 3	199.0	197.2	204.2	187.2	176.1	161.6	172.4
Percent taking reading test—grade 3	87.1	87.6	85.5	79.9	78.1	80.1	79.8
IGAP reading score—grade 6	203.6	202.7	201.0	203.9	184.3	189.8	201.7
Percent taking reading test—grade 6	88.3	89.5	90.7	85.3	83.4	84.5	85.8
IGAP reading score—grade 8	223.1	216.8	217.1	214.9	205.3	210.9	208.9
Percent taking reading test—grade 8	88.1	89.5	91.0	85.8	84.4	84.8	86.3
ACT composite score	14.10	13.93	15.99	15.93	15.90	15.98	15.85
Percent taking ACT	51.5	47.6	52.3	55.8	55.5	55.9	53.7
Graduation rate (%)	48.5	45.6	47.1	43.2	47.6	48.5	49.3

Note: Values are means of school-level means.

TABLE 2

## Selected student outcome measures, all single-school districts in Illinois

	Pre-reform years		Post-reform years				
	1987-88	1988-89	1989-90	1990-91	1991-92	1992-93	1993-94
IGAP reading score—grade 3	271.6	275.8	276.8	266.0	267.2	270.2	281.0
Percent taking reading test—grade 3	93.1	93.1	92.6	90.8	89.7	91.3	92.2
IGAP reading score—grade 6	270.5	264.0	263.4	268.4	264.0	282.3	283.4
Percent taking reading test—grade 6	91.4	90.8	92.2	89.8	88.7	89.6	91.0
IGAP reading score—grade 8	264.0	271.7	273.2	267.9	262.0	274.7	279.9
Percent taking reading test—grade 8	91.5	90.6	91.4	90.0	88.0	89.6	91.1
ACT composite score	20.81	20.58	20.67	20.60	20.63	20.90	21.0
Percent taking ACT	57.0	57.9	57.3	57.8	60.3	61.2	62.1
Graduation rate (%)	90.7	90.9	89.4	90.8	91.0	90.8	85.7

Note: Single-school districts were defined as those elementary districts with one elementary school in 1992-93, those high school districts with one high school in 1992-93, and those unit districts with one elementary school and one high school in 1992-93. Values are means of school-level means.

characteristics of each school's student population to control for changes in the characteristics of the cohorts being tested. While these controls are imperfect, previous research indicates that variation in characteristics of the student population explains a significant portion of the variance in test performance.<sup>43</sup> In addition, trends in student performance in districts outside Chicago provide a control for statewide (and to some extent, nationwide) trends in student performance. Even with these controls, using school-level data to assess the efficacy of reform is as problematic as using student-level data. Therefore, we view this work as complementary to, and not substituting for, the work of Bryk et al.

The first analysis we performed to assess the success of the Chicago reform was a simple comparison of the pre- and post-reform means of IGAP reading test scores, attendance rates, and graduation rates. If changes in the student population and in the nature of the test instrument are, on average, unrelated to changes in these performance measures, the comparison of means provides a clear indication of the effect that the reform has had on schooling outcomes. The variables used in the analysis are means of IGAP reading test scores for grades 3, 6, and 8; composite ACT scores; the

percentage of students taking the tests; and graduation rates.<sup>44</sup>

However, means provide a less-than-perfect mechanism for judging the effects of the reform. First, it is impossible to distinguish reform effects from secular changes in test scores or graduation rates unattributable to reform. Second, means do not distinguish specific areas or schools where reform is successful or unsuccessful. Finally, analysis of means fails to account for changes in cohorts or for above-average mobility in the population served by a school.

To cope with these limitations, we estimated a number of linear regressions with student outcomes, test scores, and graduation rates as the dependent variables and measures of exogenous school and student characteristics from the school report cards, as well as characteristics of the communities surrounding the schools, as the independent variables.<sup>45</sup> In some of the specifications, we used schools outside of Chicago as a control group to allow us to account for secular trends not attributable to reform.

We created dummy variables to distinguish between pre- and post-reform years and between Chicago and the rest of Illinois. The reform dummy took the value of one in the post-reform years and zero in the pre-reform

years. Similarly, the Chicago dummy took the value of one if the school was located in Chicago, and zero otherwise.

We also created interactions between the reform dummy, the Chicago dummy, and student characteristics.<sup>46</sup> These interactions allowed us to determine the characteristics of individual schools and the surrounding communities that were most strongly correlated with reform success. For example, reform would be expected to be less successful in communities with a higher percentage of students with limited English proficiency or a higher percentage eligible for subsidized school lunch. Evidence from East Harlem, New York, supports the argument that additional resources are necessary in schools where larger shares of the student population are disadvantaged.<sup>47</sup> We also expected reform would be more successful in communities that are predominantly Hispanic than those that are predominantly African-American because Hispanics were more supportive of reform and more actively involved in the shaping of it.

## Results

### Means

Table 1 reports the means of the student outcome measures for schools within Chicago. Also included in the table are the mean percentage of students taking the tests each year. These latter statistics serve as an indication of the consistency in test administration.

Table 2 gives the same summary statistics for districts elsewhere in the state that had, at most, one elementary school and one high school in the 1992-93 academic year. Focusing on single-school districts allows us to choose a control group in which questions of district management are not a concern. As a result, the single-school districts elsewhere in Illinois provide a natural control group that allows us to distinguish between general trends in student outcomes and changes that could be reform-induced.<sup>48</sup>

In the Chicago public schools, scores on the IGAP tests exhibited considerable change post-reform. Third-grade reading scores increased in the first year after reform, decreased precipitously in the following three years, and recovered slightly in 1993-94.<sup>49</sup> Sixth-grade reading scores changed little in the first two years after reform and then fell noticeably in 1991-92. Improved performance in the final

two academic years still left sixth-grade scores below their pre-reform level.

Unlike the third- and sixth-grade scores, the eighth-grade reading scores showed little overall change, but the mean eighth-grade reading score in 1993-94 was below both the 1987-88 and 1988-89 means.<sup>50</sup> The percentage of students taking the tests decreased slightly for each grade following the reform, which may cause the drop in test scores to be understated if the group of test takers was becoming more select.

The post-reform changes in the high schools' performance measures painted a more positive picture of reform. As can be seen in table 1, the graduation rate decreased slightly in the second year after reform but rebounded completely in the next three years. By 1993-94, the graduation rate was above the 1987-88 pre-reform level. The ACT composite score displayed even more substantial improvement. Even with adjustment made for the change in the test instrument, the mean composite ACT score increased by almost 18 percent in the first year after reform. This gain was largely sustained over the remaining four post-reform years.

Single-school districts outside Chicago showed a pattern of improved student performance at the elementary level. In all grades, IGAP reading scores were lowest in 1990-91 and 1991-92, but even in these years, the changes relative to the pre-reform years were far less dramatic than in Chicago. When the results from 1992-93 and 1993-94 were factored in, the mean performance of Chicago public school students on the IGAP reading test in grades 3, 6, and 8 had clearly fallen relative to mean performance elsewhere in Illinois.

A more heartening picture emerged from a comparison of high school student performance measures inside and outside Chicago. The graduation rate in Chicago improved noticeably relative to the rate in districts outside the city. Even more substantial relative improvement was evident in the mean ACT composite score. While this gain may be taken as an encouraging signal for reform advocates, the change in the test instrument in 1989-90 and the select nature of the students taking this test means that these results cannot be treated as an unequivocal signal of positive reform

effects.<sup>51</sup> The coincidence in timing of the Chicago reform and the change in ACT test instruments, combined with the absence of any clear guidelines for renorming the ACT for specific subpopulations, make it impossible to

identify separately the impact of reforms on performance on the ACT. National data indicate that after the change in the test instrument, scores tended to cluster more tightly about the national mean. If true, this would imply that

<b>TABLE 3</b>					
<b>Analysis of schooling outcomes, Chicago public schools, 1987-88 to 1992-93</b>					
<b>Independent variables</b>	<b>Dependent variables</b>				
	<b>Mean IGAP reading score-grade 3</b>	<b>Mean IGAP reading score-grade 6</b>	<b>Mean IGAP reading score-grade 8</b>	<b>Graduation rate-log of odds ratio</b>	<b>Mean ACT score</b>
Constant	283.997 (20.270)	270.873 (16.419)	266.977 (15.106)	-1.126 (1.160)	26.196 (2.565)
Reform dummy	27.610 (9.309)	35.167 (7.471)	21.433 (6.805)	0.278 (0.596)	2.869 (1.435)
<b>Characteristics of school's student population</b>					
Fraction African-American	-28.162 (12.831)	-7.755 (10.388)	-24.909 (10.172)	0.003 (0.696)	-6.906 (1.568)
Fraction Hispanic	-21.678 (18.855)	13.846 (15.021)	-2.012 (14.034)	1.398 (0.988)	-1.969 (2.247)
Fraction Asian-American	91.303 (34.046)	67.027 (27.175)	25.717 (24.535)	4.334 (1.852)	9.447 (4.389)
Fraction limited-English-proficient	-49.492 (21.711)	-52.569 (18.897)	-49.829 (18.259)	-12.249 (2.612)	-32.133 (6.222)
Fraction eligible for school lunch	-72.860 (9.289)	-79.259 (7.553)	-64.852 (7.300)	-0.558 (0.512)	-2.888 (1.216)
School enrollment	-0.013 (0.006)	-0.017 (0.005)	-0.002 (0.004)	0.0002 (0.0001)	0.0008 (0.0002)
<b>Characteristics of community area in which school is located</b>					
Fraction college graduate	133.378 (16.225)	115.948 (13.348)	121.760 (12.404)	0.465 (0.724)	3.934 (1.751)
Fraction high school graduate	-7.974 (24.052)	15.627 (19.731)	12.740 (18.235)	1.300 (1.114)	-7.198 (2.594)
Median income	-0.0013 (0.0004)	-0.0006 (0.0003)	-0.0005 (0.0003)	0.00001 (0.00001)	-0.00005 (0.00003)
Fraction with income less than \$15,000	-60.019 (17.807)	-38.831 (14.475)	-12.164 (13.175)	0.034 (0.834)	-5.653 (1.886)
Fraction with income greater than \$75,000	1.691 (32.797)	-72.957 (26.691)	-55.041 (24.820)	-1.512 (1.304)	-6.867 (3.022)
Fraction African-American	-187.887 (149.358)	307.622 (123.084)	259.591 (119.233)	3.175 (6.647)	17.672 (16.004)
Fraction Hispanic	28.539 (8.792)	12.391 (7.180)	7.616 (7.024)	-0.547 (0.465)	-3.628 (1.106)
Fraction Asian-American	-2,423.256 (859.066)	-1,892.554 (723.738)	-2,030.889 (655.759)	21.566 (35.876)	-169.66 (85.416)
Fraction nonmovers	68.141 (13.179)	43.770 (10.932)	46.479 (10.296)	0.337 (0.542)	-1.951 (1.302)
<b>Interactions of reform dummy and student characteristics</b>					
Fraction African-American	-67.523 (13.781)	-63.762 (11.171)	-38.483 (10.881)	-0.550 (0.709)	-1.014 (1.681)
Fraction Hispanic	-33.554 (20.594)	-35.625 (16.344)	-17.636 (15.300)	-1.369 (1.041)	-3.848 (2.430)
Fraction Asian-American	-16.989 (39.091)	27.251 (31.379)	44.881 (28.352)	-0.652 (2.017)	-8.417 (4.830)
Fraction limited-English-proficient	-25.861 (25.142)	-58.373 (21.914)	-46.935 (21.223)	8.525 (2.678)	20.654 (6.355)
Fraction eligible for school lunch	14.857 (10.514)	13.473 (8.580)	2.372 (8.296)	0.510 (0.539)	2.468 (1.280)
School enrollment	-0.004 (0.007)	0.003 (0.006)	0.002 (0.005)	0.00002 (0.00010)	-0.0002 (0.0002)
Number of observations	2,384	2,297	2,114	335	343
R <sup>2</sup>	0.4993	0.5429	0.5319	0.4708	0.7123

Note: Standard errors in parentheses.

TABLE 4

**Analysis of schooling outcomes, measures of increased flexibility included,  
Chicago public schools, 1987-88 to 1992-93**

Independent variables	Dependent variables				
	Mean IGAP reading score—grade 3	Mean IGAP reading score—grade 6	Mean IGAP reading score—grade 8	Graduation rate—log of odds ratio	Mean ACT score
Constant	286.074 (24.959)	256.441 (20.146)	281.733 (18.158)	-1.841 (1.349)	24.346 (2.857)
Reform dummy	4.039 (19.719)	34.723 (15.248)	20.834 (13.261)	0.116 (0.994)	2.317 (2.277)
<b>Characteristics of school's student population</b>					
Fraction African-American	-28.254 (12.808)	-7.350 (10.366)	-25.169 (10.176)	0.038 (0.693)	-7.151 (1.530)
Fraction Hispanic	-21.891 (18.820)	14.265 (14.994)	-3.115 (14.029)	1.405 (0.978)	-2.429 (2.198)
Fraction Asian-American	90.459 (34.018)	67.913 (27.105)	26.048 (24.517)	4.503 (1.856)	9.482 (4.300)
Fraction limited-English-proficient	-49.431 (21.689)	-54.899 (18.883)	-48.189 (18.280)	-12.061 (2.598)	-30.991 (6.127)
Fraction eligible for school lunch	-73.581 (9.404)	-77.903 (7.582)	-65.722 (7.309)	-0.528 (0.508)	-2.652 (1.192)
Minutes per day in English class	-0.022 (0.080)	0.006 (0.087)	-0.034 (0.073)	—	—
Average class size	-0.040 (0.317)	0.378 (0.224)	-0.355 (0.221)	0.038 (0.032)	0.128 (0.073)
School enrollment	-0.012 (0.006)	-0.019 (0.005)	-0.001 (0.004)	0.0002 (0.0001)	0.0006 (0.0002)
<b>Characteristics of community area in which school is located</b>					
Fraction college graduate	136.209 (16.223)	119.114 (13.328)	120.552 (12.432)	0.654 (0.720)	4.552 (1.714)
Fraction high school graduate	-5.572 (24.054)	15.420 (19.666)	10.969 (18.248)	1.209 (1.107)	-7.879 (2.542)
Median income	-0.0012 (0.0004)	-0.0006 (0.0003)	-0.0005 (0.0003)	0.00001 (0.00001)	-0.00006 (0.00003)
Fraction with income less than \$15,000	-57.000 (17.863)	-35.596 (14.449)	-12.947 (13.194)	-0.032 (0.825)	-6.196 (1.846)
Fraction with income greater than \$75,000	-3.553 (32.854)	-81.233 (26.682)	-56.782 (24.836)	-1.426 (1.295)	-7.035 (2.950)
Fraction African-American	-195.363 (149.095)	325.243 (122.844)	251.165 (119.312)	4.531 (6.637)	23.378 (15.757)
Fraction Hispanic	29.745 (8.795)	13.398 (7.176)	8.364 (7.044)	-0.621 (0.462)	-3.995 (1.084)
Fraction Asian-American	-2,330.619 (858.667)	-1,813.274 (721.851)	-2,029.845 (656.919)	26.927 (35.597)	-151.085 (83.450)
Fraction nonmovers	68.802 (13.161)	45.171 (10.905)	47.101 (10.297)	0.344 (0.536)	-1.874 (1.269)
<b>Interactions of reform dummy and student characteristics</b>					
Fraction African-American	-65.599 (13.775)	-61.171 (11.190)	-38.983 (10.910)	-0.430 (0.711)	-0.486 (1.652)
Fraction Hispanic	-31.611 (20.567)	-33.726 (16.341)	-17.661 (15.321)	-1.259 (1.033)	-3.212 (2.377)
Fraction Asian-American	-15.294 (39.052)	26.527 (31.312)	44.493 (28.342)	-0.776 (2.029)	-8.556 (4.750)
Fraction limited-English-proficient	-27.550 (25.115)	-55.610 (21.889)	-48.284 (21.247)	8.597 (2.667)	20.324 (6.274)
Fraction eligible for school lunch	18.408 (10.654)	13.092 (8.602)	4.014 (8.310)	0.448 (0.533)	2.135 (1.254)
Minutes per day in English class	0.021 (0.102)	-0.052 (0.104)	-0.081 (0.086)	—	—
Average class size	0.702 (0.366)	0.187 (0.264)	0.310 (0.259)	0.003 (0.035)	0.009 (0.079)
School enrollment	-0.006 (0.007)	0.002 (0.006)	0.001 (0.005)	0.00002 (0.0001)	-0.0002 (0.0002)
Number of observations	2,384	2,297	2,112	334	342
R <sup>2</sup>	0.5020	0.5467	0.5338	0.4869	0.7295

Note: Standard errors in parentheses.

mean scores in districts like Chicago, in which many students had ACT scores well below the national mean, would have risen relatively even in the absence of any reform effects.

***Within-Chicago regressions: Controlling for variation in student and community-area characteristics***

Evaluating reform on the basis of a comparison of pre- and post-reform mean student outcomes may be misleading. Changes in the attributes of cohorts, particularly changes that are unique to either Chicago schools or schools in the control group, can result in fluctuations in mean performance and in relative performance even if there are no reform effects.<sup>52</sup> We control for potential changes in the student population in two ways. In this subsection, we present results that utilize data only on schools within Chicago. We use observable characteristics of the student body of each school and of the community area in which the school is located to account for variations in student ability and in the environment in which education is provided. In the next subsection, we analyze the relationship between student performance and student characteristics using both the schools within Chicago and the single-school districts elsewhere in Illinois. The schools outside Chicago provide us with the control group needed to distinguish general trends in performance from reform-induced effects.

Tables 3 and 4 give estimates of the parameters of the linear regression models that account for differences across schools and community areas.<sup>53</sup> These estimates are calculated using only schools within Chicago. Table 3 presents pure reduced-form estimates. In table 4 minutes per day in English classes and average class size in the relevant grade are added to the independent variables. These are added because post-reform variation should reflect the increased flexibility of local schools. The School Improvement Plan provides each LSC with the opportunity to mandate changes in average class size and minutes per day in the core academic subjects; the shift of control of the Chapter 1 funds to the schools potentially gives the LSCs the resources to implement these changes.

As is apparent from table 4, we found no evidence that schooling outcomes depended in any systematic way on minutes per day in

English classes or on average class size. Further, the relationship between the student performance measures and these indicators of flexibility was not statistically altered by the reform. In short, there was no indication that reform success was related to these proxies for local flexibility. This absence of any relationship between post-reform performance and these measures of flexibility could be taken as a signal that responses to increases in local flexibility did little to promote post-reform success.<sup>54</sup> Since a central premise of the Chicago experiment with school-based management was the idea that allowing increased local flexibility could promote achievement gains, the estimates in table 4 could be interpreted as evidence that the experiment failed. We were loath to draw this conclusion for two reasons. First, the financial crisis in the Chicago public schools limited the ability of the LSCs to use the Chapter 1 moneys to implement the School Improvement Plans. Second, discussions with individuals in the Department of Research, Evaluation, and Planning of the Chicago Public School System indicated that the data on average class size and minutes per day were particularly unreliable. Therefore, the discussion that follows concerning the effects of reform focuses on the estimates in table 3.

The specifications all included interactions of a dummy variable indicating whether or not the year is post-reform with the characteristics of the student population and of the school. As the discussion above indicated, these interactions allowed us to test the hypothesis that the success or failure of reform depended upon the environment in which that reform was implemented.

The level of student performance depended critically on the characteristics of the student population, as would be expected given previous findings. The fact that larger fractions of students with limited English proficiency and eligible for subsidized school lunch correlate with lower student performance and higher student costs was well established.<sup>55</sup> Similarly, the less consistent relationship between student performance and the fractions of students who were African-American and Hispanic paralleled findings of other authors.<sup>56</sup> The relationship between the fraction of students who were Asian-American and schooling outcomes, which we found to be consistently positive, has

not been explored as extensively. Nevertheless, these latter results were not surprising.

Smaller elementary schools had higher IGAP reading scores, though smaller high schools had, on average, lower ACT scores and graduation rates. The negative relationship between school size and reading scores was expected because students tend to get more personal attention and teaching in smaller schools.<sup>57</sup> The positive relationship between size and high school outcomes was more difficult to explain, though such a result is consistent with the existence of scale economies in high schools.<sup>58</sup>

Results showing that several community-area characteristics correlated significantly with schooling outcomes also replicated a common finding in the literature on education production. Student performance was better in schools where a larger fraction of the adult population had graduated from college. This result, and our finding that community areas with less mobility generally had better performance outcomes, duplicated Ferguson's results.<sup>59</sup> Like Sander, we found that, except for the graduation rate, there was a negative relationship between community median family income and student performance.<sup>60</sup> Student performance was also lower in those community areas where the variance in family income was greater. The racial/ethnic composition of community areas exhibited no systematic relationship with student outcomes, though somewhat surprisingly, community areas with larger Asian-American populations had, *ceteris paribus*, lower standardized test scores.

The fact that student performance depended on the characteristics of the students and of the communities in which the schools are located was not surprising. Of far more immediate interest was the relationship between reform success and the attributes of individual schools. For all of the performance measures, we were able to reject the null hypothesis that the relationship between schooling outcomes and the characteristics of the students and the schools was unchanged after reform. Further, for all of the performance measures, the coefficient on the reform dummy variable was positive and, in most cases of interest, statistically significant. In combination with the downward trend in mean performance on the IGAP reading test, this result implied that changes in

the structure of the relationship between the characteristics of the student population and schooling outcomes explained the preponderance of the post-reform changes in those outcomes.

Schools with larger fractions of African-American students were more likely to exhibit a post-reform slippage in student performance, a not-unexpected result given the opposition to reform in the African-American community. More surprising, given the Hispanic community's support for reform, were the negative coefficients on the interaction between the reform dummy and the fraction of students who were Hispanic. However, only one of these coefficients differed significantly from zero.

The estimates also indicated that after reform, schools with larger fractions of students eligible for subsidized school lunch exhibited relatively better student performance, a possible indication that schools have been able to make effective use of increased discretion over Chapter 1 moneys. The relationship between post-reform success and another important determinant of schooling costs—the fraction of students with limited English proficiency—was less clear. At the elementary level, schools with larger fractions with limited English proficiency had further diminished in relative performance after reform. There was no obvious explanation for this result, though it could have been attributable to increased difficulty in operating successful bilingual programs in the face of budgetary limitations and central administration cutbacks. If these factors did create problems at the elementary level, they failed to have the same effect at the high school level, where schools with a larger fraction of students with limited English proficiency had relatively greater post-reform success, holding all else equal.

Simply reviewing the estimates in table 3 provided an incomplete picture of what the estimates allowed us to conclude about the relationship between the characteristics of individual schools in Chicago and the post-reform success of these schools. To complete this picture, we used the estimates in table 3 to calculate predicted IGAP reading scores under two alternative scenarios. The first scenario assumed that, in the absence of reform, the pre-reform relationship between student performance and student and school characteristics

would have continued to be the relationship between these variables in 1993-94. In other words, we calculated predicted scores using the estimates in table 3 with the reform dummy set to zero. Under the second scenario, we calculated the predicted post-reform scores implied by the estimates in table 3. The difference between the predicted scores under the second scenario and under the first scenario provided a measure of the effect of reform on achievement in a particular school. The first row of table 5 gives these predicted effects for grade 3 IGAP scores for a hypothetical mean school and several actual schools.<sup>61</sup>

In the 1993-94 academic year, an elementary school with the mean student body would have had 623 students, 61 percent African-American, 24 percent Hispanic, and 3 percent Asian-American. Of these students, 13 percent would have been limited-English-proficient and 81 percent would have been eligible for subsidized school lunch. Given the estimates in table 3, the third-grade IGAP reading score in this school was 15.42 points below what it would have been if no reform had been implemented.

Beethoven Elementary School, located in the Fuller Park community area, is an example of a school that has seen performance fall in the post-reform period relative to what it

would have been in the absence of reform. In 1993-94 Beethoven's student population was 838, all of whom were African-American and 93 percent of whom were eligible for subsidized school lunch. None of the students had limited English proficiency. The predicted post-reform third-grade IGAP reading score for Beethoven was 29.15 points below the score the estimates in table 3 indicate that a school with these attributes would have had if no reform had been implemented. Marsh Elementary School, in the South Deering community area, is an example of a school that, the estimates imply, was better situated to benefit from reform. Of the students at Marsh, 2 percent were African-American and 92 percent Hispanic; there were no Asian-Americans. Of the 378 students, 86 percent were eligible for subsidized school lunch and 18 percent had limited English proficiency. The implied post-reform third-grade IGAP reading score for Marsh was 2.08 points higher than the predicted score for a school with Marsh's attributes if no reform had been implemented.

*Using other districts in Illinois as a control group*

The predicted scores for the hypothetical mean schools and for selected individual schools point us toward two potential conclu-

TABLE 5						
Analysis of changes in 1993-94 grade 3 IGAP reading scores for a hypothetical mean school and selected actual schools						
	Mean elementary school <sup>a</sup>	Beethoven Elementary School	Clinton Elementary School	Marsh Elementary School	Clissold Elementary School	Irving Elementary School
Change in scores not explained by pre-reform within-Chicago relationships <sup>b</sup>	-15.42	-29.15	10.38	2.08	-6.32	-12.33
Change in scores not explained by statewide trends <sup>c</sup>	1.78	-40.20	-35.18	111.45	3.49	32.58
Change in scores predicted by statewide trends alone <sup>c</sup>	-16.66	7.48	21.89	-94.97	-4.57	-40.02

Note: The methods used to calculate these values are described in the text.

<sup>a</sup>In the 1993-94 academic year, an elementary school with the mean student body would have had 623 students, 61 percent African-American, 24 percent Hispanic, and 3 percent Asian-American. Of these students, 13 percent would have been limited-English-proficient and 81 percent would have been eligible for school lunch subsidies.

<sup>b</sup>Based on the estimates in table 3.

<sup>c</sup>Based on the estimates in table 5.

sions that can be reached on the basis of estimates that control for cross-school variation within Chicago in the determinants of performance. First, these estimates fail to alter the basic conclusion that, on average, performance fell post-reform. Second, the results of this analysis make clear the dependence of post-reform success on the characteristics of individual schools.

However, if we base an evaluation of reform success only on performance trends within Chicago, we may risk masking the true effects of reform on student performance. For example, if in the post-reform period student performance was falling throughout the state in schools like those in Chicago, this might actually obscure actual gains due to reform. We therefore chose single-school districts elsewhere in Illinois to be the control group that would enable us to distinguish reform effects from general trends. Table 6 gives estimates of the parameters of the linear regression models first presented in table 3, calculated using both the control group and schools within Chicago.<sup>62</sup>

The estimates in table 6 continue to imply that across-school variation in student performance depends critically on variation in the characteristics of the schools' student bodies. Schools with larger-than-average fractions of students who are limited-English-proficient, who are eligible for subsidized school lunch, or who are African-American have lower-than-average student performance. Schools with larger-than-average fractions of students who are Asian-American have higher-than-average student performance. The coefficients on the fraction of Hispanic students exhibit no consistent pattern across the performance measures. Student performance tends to be lower in schools with larger enrollments, though a notable exception to this pattern is the significant positive effect that school size has on ACT scores.

The more striking implication of the estimates in table 6 is that looking only within Chicago yields a mistaken picture of the relationship between reform success and a school's attributes. The coefficients on the interactions with the reform dummy indicate how post-reform performance throughout the state related to a school's attributes; the coefficients on the interactions with both the reform dummy

and the Chicago dummy tell us how the post-reform effects of these attributes differed for schools within Chicago. For example, since the coefficient of the interactions between the reform dummy and the fraction of Hispanic students is negative, we can conclude that in Illinois schools the relationship between student performance and the fraction of Hispanic students became less positive or more negative. However, since the coefficient on the interaction between the reform dummy, the Chicago dummy, and the fraction of Hispanic students is positive, student performance in schools in Chicago deteriorated less rapidly than did student performance in schools outside Chicago with comparable Hispanic fractions. The reform appears to have mitigated a downward trend in student performance in schools with large fractions of Hispanic students, a result that the within-Chicago estimates fail to reveal. These results lend support to the argument that the involvement of the Hispanic community apparently translated into greater success in heavily Hispanic schools after reform.

The within-Chicago results also led to what are apparently mistaken conclusions about the relationship between post-reform success and the fractions of students eligible for subsidized school lunch and with limited English proficiency. The results in table 6 indicate that, in Chicago elementary schools with larger fractions of students eligible for subsidized school lunch, student performance declined relatively after reform, a result that was obscured in the estimates in tables 3 and 4. Similarly, at the high school level, the within-Chicago results contrast with the estimates in table 6, which imply that Chicago schools with large percentages of limited-English-proficient students had relatively lower post-reform student performance.

The negative relationship between post-reform success and the fraction of students eligible for subsidized school lunch would be particularly disheartening if it implied an inability of schools to take advantage of increased discretion over Chapter 1 funds. However, such a conclusion may be unwarranted because the budgetary crisis in the Chicago schools has resulted in schools' using Chapter 1 dollars to pay for positions previously supported out of regular funds.<sup>63</sup> While some

TABLE 6

**Analysis of schooling outcomes, Chicago public schools,  
1987-88 to 1993-94**

Independent variables	Dependent variables				
	Mean IGAP reading score—grade 3	Mean IGAP reading score—grade 6	Mean IGAP reading score—grade 8	Graduation rate—log of odds ratio	Mean ACT score
Constant	290.404 (2.100)	285.906 (1.783)	278.788 (1.744)	4.051 (0.197)	20.777 (0.132)
Chicago dummy	-8.121 (4.884)	-5.122 (4.125)	4.137 (3.981)	-2.081 (0.634)	-4.085 (0.425)
Reform dummy	-3.560 (2.976)	-3.388 (2.517)	7.389 (2.425)	-0.084 (0.249)	0.036 (0.167)
Chicago reform interaction	37.600 (6.757)	39.261 (5.701)	15.246 (5.510)	-0.697 (1.311)	2.323 (0.877)
<b>Characteristics of school's student population</b>					
Fraction African-American	-29.594 (6.913)	-16.590 (5.846)	-22.090 (5.768)	-1.163 (0.751)	-3.913 (0.501)
Fraction Hispanic	-18.454 (11.439)	6.973 (9.485)	-2.364 (9.224)	-0.162 (1.598)	0.226 (1.055)
Fraction Asian-American	110.280 (20.637)	70.931 (17.481)	57.570 (16.955)	6.112 (3.621)	13.031 (2.423)
Fraction limited-English-proficient	-48.339 (18.613)	-56.644 (16.800)	-54.394 (16.955)	-12.716 (6.955)	-25.148 (4.612)
Fraction eligible for school lunch	-75.356 (5.794)	-80.886 (4.883)	-63.139 (4.784)	-0.630 (0.680)	-3.425 (0.455)
School enrollment	-0.008 (0.003)	-0.010 (0.003)	0.0006 (0.003)	-0.0005 (0.0002)	0.0008 (0.0001)
<b>Interactions of reform dummy with</b>					
Fraction African-American	-6.546 (10.524)	-48.573 (8.872)	-28.878 (8.557)	0.292 (0.968)	0.438 (0.653)
Fraction Hispanic	-123.441 (27.579)	-102.547 (23.205)	-75.017 (22.022)	-3.790 (2.478)	-3.220 (1.652)
Fraction Asian-American	3.401 (32.100)	74.019 (27.085)	83.147 (25.915)	22.476 (6.803)	8.617 (4.560)
Fraction limited-English-proficient	47.540 (50.744)	15.905 (43.134)	15.984 (41.387)	9.235 (7.857)	23.100 (5.223)
Fraction eligible for school lunch	13.122 (8.702)	34.705 (7.329)	3.942 (7.054)	-0.051 (0.851)	1.027 (0.570)
School enrollment	0.006 (0.004)	0.011 (0.004)	-0.004 (0.003)	-0.0005 (0.0002)	-0.0003 (0.0001)
<b>Interactions of reform dummy and Chicago dummy with</b>					
Fraction African-American	-41.137 (9.723)	8.777 (8.224)	3.831 (7.961)	-0.612 (1.315)	-0.441 (0.881)
Fraction Hispanic	125.642 (26.509)	101.357 (22.355)	80.146 (21.218)	2.701 (2.436)	1.055 (1.623)
Fraction Asian-American	-8.382 (29.749)	-29.081 (25.112)	-51.196 (24.123)	-26.785 (6.665)	-14.085 (4.463)
Fraction limited-English-proficient	-92.695 (48.580)	-82.058 (41.062)	-74.002 (39.112)	0.392 (4.381)	-9.853 (2.929)
Fraction eligible for school lunch	-17.245 (7.765)	-36.863 (6.579)	-16.457 (6.359)	0.187 (0.800)	1.005 (0.535)
School enrollment	-0.024 (0.004)	-0.023 (0.004)	0.005 (0.003)	0.001 (0.0002)	0.0001 (0.0002)
Number of observations	5,654	5,557	5,334	2,314	2,316
R <sup>2</sup>	0.6823	0.6897	0.6118	0.3405	0.7616

Note: Standard errors in parentheses. For the definition of single-school districts, see table 2.

discretionary hiring has still occurred, the budget problems have prevented us from observing a clear experiment that could be used to quantify the benefits of increased school-level control over budgetary decisions.

To better understand the implications of the estimates in table 6, we again calculated the estimated effect of the Chicago reform on the hypothetical school with the mean student body characteristics and on several other schools. The results of these calculations are given in the final two rows of table 5. Since the estimates in table 6 implied that it is inappropriate to assume that the relationship between student performance and student and school characteristics was unchanged after reform, we needed to change our method for calculating the predicted effect of reform on student achievement. In particular, we needed to use the fact that the control group allowed us to determine the trend in student performance in Illinois schools. Therefore, under the first scenario, we calculated predicted scores for Chicago schools assuming that, in the absence of reform, they would have followed the statewide trends. The third row in table 5 gives the predicted change in grade 3 IGAP reading scores for the case of no reform. Under the second scenario, we again calculated predicted post-reform scores. The effect of reform was then the amount by which the predicted post-reform score differed from the trend score. These predicted reform effects are given in the second row of table 5.<sup>64</sup>

The estimates implied that in the school with the mean characteristics, third-grade test scores rose 1.78 points above what they would have been in the absence of reform. Failing to control for universal trends obscures real post-reform gains in the Chicago schools.

Looking at individual schools again makes it clear that post-reform trends depended critically on the characteristics of the students served by those schools. Because of the large fraction eligible for subsidized school lunch in Beethoven Elementary School, the estimates indicate that third-grade IGAP reading scores fell by 40.20 points relative to trend. On the other hand, since Marsh Elementary had a large fraction of students who were Hispanic, the estimates in table 5 imply that student performance in Marsh was 111.45 points above trend.

## Conclusion

The goal of this article was to measure the effectiveness of the Chicago reform by comparing student outcomes from pre- and post-reform years. We also examined whether particular features of individual schools or communities contributed to the success of the reform. Single-school districts from the rest of Illinois were used to separate the effects of reform from secular trends independent of it.

In the Chicago public schools, mean performance on IGAP tests in grades 3, 6, and 8 has fallen in the post-reform period relative to mean performance elsewhere in the state. Post-reform results at the high school level have been more promising; graduation rates and ACT scores have exhibited relative increases. At the same time, examination of these post-reform changes in mean performance and the relationship within Chicago between student performance and school and community attributes fails to reveal gains across all grade levels in the Chicago schools. When we used single-school districts in Illinois as a control group, we found that the reform may have actually mitigated the downward trend in performance in the Chicago schools. In particular, we noted that for the elementary school with the mean characteristics, third-grade IGAP reading scores in 1992-93 were 1.78 points higher than they would have been in the absence of reform.

In addition, as Bryk et al. have suggested, post-reform changes in performance depended on the characteristics of a school's student body.<sup>65</sup> The results in table 6 indicate that schools with larger fractions of students with limited English proficiency and larger fractions of students eligible for subsidized school lunch have tended to fare worse in the post-reform period, while those with relatively larger Hispanic populations have fared better. The negative relationship between post-reform success and the relative size of a school's at-risk population appears to imply that decentralization may not be the answer for urban school systems, since individual Chicago schools appear not to have benefited from additional control over Chapter 1 funds. However, the budget crisis in Chicago has limited the benefits of this discretion. The Chicago case may just confirm a lesson demonstrated in the East Harlem school district of New York City;

decentralization of schooling governance is less likely to generate significant gains in the absence of funding increases or in the face of funding reductions.<sup>66</sup>

The relative improvement in outcomes after reform in schools where a larger percentage of the student population was Hispanic highlights the relationship between the amount of input a group has in the creation and implementation of reform and the effectiveness of reform in schools in which that group is dominant. Hispanic leaders were active in designing the Chicago reform; African-American leaders felt the reform was imposed on them. However, the absence of pre-reform support among African-American leaders has not meant that no schools with large fractions of students who are African-American were among the post-reform successes. Our results indicated post-reform success was unrelated to the fraction of a school's students who are African-American. Further, there were schools such as Washington Irving Elementary with large fractions of African-American students (49.2 percent) and considerable post-reform success (a 32.58 point increase in grade 3 IGAP reading score, relative to trend).

There are a number of important caveats to the results of this study, caveats that suggest directions for future research. First, focusing on changes in student test performance at the school level diminishes the strength of our conclusions. Changes in student performance on standardized tests provide an incomplete measure of the effectiveness of reform. In addition, changes in cohorts or in the test instruments can affect scores independent of reform. By accounting for student characteristics and by using other schools in Illinois as a control group, we lessened the impact of these

potential problems. Nevertheless, a more complete picture of the effects of reform can be created if a broader set of outcomes is considered, if better value-added measures are utilized, and if performance trends in other potential control groups are considered.

Second, the methodology and data used may not provide enough information for a complete analysis of the Chicago reform. Since the results of this and other research<sup>67</sup> show that reform affects schools differentially, the analysis could be strengthened by supplementing the data used with information on LSC elections and participation, school-by-school teacher and principal characteristics, measures of parental involvement, and indications of specific changes that have been made at each school since the reform. Obtaining complete demographic data for the rest of Illinois would also prove useful, since parallel regressions could then be estimated for both the within-Chicago case and the full-sample case.

Even at this preliminary stage, the Chicago reform has instructive implications for other cities. Change must come from within the school system and should be fully supported by the major constituencies in the city. Efforts should be concentrated on providing assistance to schools serving large populations of students from low-income backgrounds or who have limited English proficiency. Further, the strength of the relationships among student characteristics, student performance, and post-reform success further supports the argument that individual schools should be able to tailor their response to reforms as need arises. The Chicago reform, even if it does not seriously change the condition of the Chicago public schools, will provide useful information to reformers contemplating school-based management.

---

## NOTES

<sup>1</sup>Hess (1990).

<sup>2</sup>Mirel (1993).

<sup>3</sup>Pink (1992).

<sup>4</sup>In particular, see the articles in the May 1994 issue of *Education and Urban Society*.

<sup>5</sup>Bryk, Deabster, Easton, Luppescu, and Thum (1994).

<sup>6</sup>Hannaway and Carnoy (1993).

<sup>7</sup>For a more in-depth discussion of decentralization, see Oates (1985).

<sup>8</sup>Zax (1988).

<sup>9</sup>Oates (1972).

<sup>10</sup>Hannaway and Carnoy (1993).

<sup>11</sup>For further discussion of the relationship between school-based management and other reforms that decentralize management of school systems, see Brown (1990).

<sup>12</sup>Brown (1990).

<sup>13</sup>Lane (1991).

<sup>14</sup>Brown (1990).

<sup>15</sup>Hatry et al. (1993).

<sup>16</sup>Ibid.

<sup>17</sup>While it would be interesting to consider such measures of community involvement as the percentage of voters participating in local school council elections, we were not able to obtain complete local school council data for this study.

<sup>18</sup>Walberg, Bakalis, Bast, and Baer (1989).

<sup>19</sup>See Wohlstetter and McCurdy (1991). Hanson, Morris, and Collins (1992) argue that in Miami, few substantive powers have been delegated to teachers. Hess (1992b) notes that no significant improvement in student achievement has resulted from the broadening of school-based management in Miami. While these observations make clear that Miami's reform has not been an unreserved success, they are not inconsistent with the conclusion that it has been more successful than New York's.

<sup>20</sup>Wohlstetter and McCurdy (1991).

<sup>21</sup>Mirel (1993) states that if integration had been a goal, "civil rights organizations undoubtedly would have filed suit against any reform effort similar to the one that was passed because the local school councils would have further 'legitimized' segregated neighborhood schools."

<sup>22</sup>Hess (1990).

<sup>23</sup>State Chapter 1 aid is intended to compensate districts with large concentrations of economically disadvantaged students for the higher cost of serving them.

<sup>24</sup>Hess (1990).

<sup>25</sup>Wong and Sunderman (1995) note that the Chicago School Reform Act did not explicitly state the precise manner in which decisionmaking was to be divided between the school district's central administration and the individual schools.

<sup>26</sup>This plan represents the LSC's vision for the school and a statement of how the goals of the Chicago School Reform Act are going to be achieved (Flinspach and Ryan 1994). The act did not state in detail what should be in the plan.

<sup>27</sup>Hess (1990).

<sup>28</sup>Pink (1992).

<sup>29</sup>Hess (1992).

<sup>30</sup>Mirel (1993).

<sup>31</sup>The Chicago School Finance Authority (SFA) was established in 1979 after the city of Chicago school's district declared bankruptcy. Under the Chicago School Reform Act, the SFA continued to exercise oversight on the district's budget even after the district had satisfied the mandate of the act originally authorizing the SFA (had balanced its budget for six consecutive years). The power of the SFA was increased again after the district experienced a severe financial crisis in 1993 (Wong and Sunderman 1995).

<sup>32</sup>Mirel (1993).

<sup>33</sup>For a brief description of the legal challenges, see Grady and Thomas (1990).

<sup>34</sup>Mirel (1993).

<sup>35</sup>Katz (1992).

<sup>36</sup>State of Illinois (1988-1994). We omitted schools and school districts from the data set if their school report card data were unavailable for any one of the seven academic years. The omitted schools and districts serve only a small portion of the city's and state's student population and do not appear to differ from the schools or districts in the data set in any systematic way.

<sup>37</sup>Census data are from U. S. Department of Commerce (1994). Community areas are as defined in City of Chicago (1994). Community-area demographic data provide imperfect measures of the demographics of the population served by individual schools. Community areas are larger than the areas served by neighborhood schools; magnet schools serve populations that may have very different characteristics from those of the community areas in which the schools are located. Nevertheless, we felt that the strategy of including imperfect controls for the characteristics of the population of the local community area was preferable to omitting controls for the characteristics of the population served by the schools.

<sup>38</sup>Capell (1981).

<sup>39</sup>For instance, if, on average, reform raised the self-esteem of students and thus increased the average student's likelihood of succeeding in employment, then the reform would be a success according to the criteria established in the Chicago School Reform Act, even if the standard measures of student performance were unchanged.

<sup>40</sup>While educators continue to express concern about the link between scores on the ACT tests, the IGAP tests, or any other achievement tests and the success of the classroom, use of performance on these tests as an indicator of classroom and school success is becoming widespread. This has led the authors of these tests to link their tests more closely to the classroom experience. Further, according to Ballantine (1993), for better or for worse, the tests now shape what goes on in the classroom. For further discussion of achievement tests and their strengths and weaknesses, see Ballantine (ibid.).

<sup>41</sup>Meyer (1993) and Bryk et al. (1994).

<sup>42</sup>Bryk et al. (1994).

<sup>43</sup>Hanushek (1986).

<sup>44</sup>Beginning in 1990, the American College Testing Program switched to a new test instrument. To adjust for this change, we have added 2 to pre-1990 ACT scores.

<sup>45</sup>We were, however, concerned that certain characteristics of a school's program were not exogenous. For example, average class size and minutes per day spent in academic subjects, variation in which might reflect the extent of flexibility accorded to schools after reform, might be correlated with unobserved attributes of schools that affect the success of reform efforts. Because of this potential endogeneity, we present results that account directly for average class size and minutes per day spent in the academic subject being tested along with reduced form specifications that omit these potentially endogenous factors.

<sup>46</sup>We also included interactions with community-area characteristics. However, there was no indication that any of these interactions were systematically related to post-reform changes, so we omitted them from the final specifications.

<sup>47</sup>Kirp (1992).

<sup>48</sup>The question of how to account for the administration of the Chicago public schools remained. In the regression results that follow, we assumed that prior to reform, differences could exist in the within- and outside-Chicago intercepts but that changes in student characteristics translated into changes in performance in the same way within and outside the city. The empirical results that followed from these assumptions are valid only if the single-school districts are a reasonable control group. However, since districts in this group were nonrandomly chosen, trends in these districts may have been nonrepresentative of those in the state or nation in ways that were not independent of the unique governance status of schools in these districts. For this reason, we used all of the districts in Illinois as another control group. The main results proved to be insensitive to the choice of control group.

<sup>49</sup>Prior to the 1993-94 academic year, students were assigned an IGAP score of 1 if their actual test score was below the score achievable by guessing. As a result, some fraction of the 1993-94 increase in mean test scores was attributable to the change in scoring practices. We did not have the data needed to adjust for this change, nor did we have the data needed to adjust for changes in the calculation of scores at the top of the score distribution. However, analysis of adjusted data by the Department of Research, Evaluation, and Planning of the Chicago Public Schools (1994) indicated that reported scores in 1993-94 substantially overstated actual improvement. Any changes in scores between 1992-93 and 1993-94 must be interpreted with caution.

<sup>50</sup>We performed the same type of analysis of pre- and post-reform means of IGAP math scores for grades 3, 6, and 8. Math scores increased substantially over the past three years, with the across-school mean in 1993-94 above the pre-reform level in all three grades. Nevertheless, the patterns of change in these scores relative to scores elsewhere in the state paralleled the patterns of change for reading scores. In all grades in the single-school districts, mean IGAP mathematics scores increased more post-reform than they did in Chicago schools.

<sup>51</sup>Increases in the mean ACT composite score could also be attributable to increasing selectness in the students being tested; however, there is no evidence that such a trend occurred. Both within and outside Chicago, the fraction of eligible students taking the ACT increased after reform.

<sup>52</sup>Meyer (1993) shows that changes in the ability of cohorts can attenuate the effects of reforms; Downes (1992) provides an example of such an effect.

<sup>53</sup>Concern over the possible biases generated by the changes in scoring methodology between 1992-93 and 1993-94 led us to drop the 1993-94 school year and to include a dummy variable for the 1993-94 school year. The central conclusions discussed in this and the next subsection were insensitive to these modifications to the specification and to the data analyzed.

<sup>54</sup>An alternative interpretation of the estimates in table 4 is that the absence of any estimated effect indicates the schools failed to take advantage of the increased flexibility by reducing average class sizes and increasing the minutes per day in academic subjects. The data do not support this interpretation. For example, in the school year prior to the reform (1988-89), the average class size in grade 3 was 26.75 students, the average number of minutes per day in English classes in grade 3 was 141.57, and the average number of minutes per day in mathematics classes in grade 3 was 46.52. In the 1993-94 school year, the average class size in grade 3 was 25.21 students, the average number of minutes per day in English classes in grade 3 was 142.38, and the average number of minutes per day in mathematics classes in grade 3 was 48.19. Similar downward trends in average class size and upward trends in minutes per day in academic subjects occurred in grades 6 and 8.

<sup>55</sup>Ferguson (1991), Sander (1993), and Downes and Pogue (1994).

<sup>56</sup>Ferguson (1991); Sander (1993).

<sup>57</sup>The relationship between school size and educational success, which is documented in Ballantine (1993), is not unique to Chicago. Harrington and Cookson (1992) argue that size was an important factor in the success of East Harlem's choice program.

<sup>58</sup>Many of the high schools with long traditions of success, such as Whitney Young Magnet High School, are both large and among the schools with the highest ACT scores. However, the relationship between ACT scores and school size cannot be explained by tradition alone; school size continues to be positively related to mean ACT score in fixed-effects variants of the specification presented in table 4.

<sup>59</sup>Ferguson (1991).

<sup>60</sup>Sander (1993).

<sup>61</sup>Because they entered the schools after reform was implemented, third graders are the most natural subjects for analysts who want to gauge the success of reform. Our basic conclusions would be unchanged if we considered either sixth-grade or eighth-grade IGAP reading scores.

<sup>62</sup>Unavailability of school-district-level Census data for 1990 prevented us from controlling for the characteristics of the surrounding communities in these regressions. Omitting such controls did not affect the basic conclusions concerning reform effects in the within-Chicago regressions. Thus, we feel that the implications of table 5 are robust. Nevertheless, we plan to explore the sensitivity to exclusion of community characteristics once Census data are available.

<sup>63</sup>Rosenkranz (1994).

<sup>64</sup>Similar conclusions would be reached for the "mean" school and for the other example schools if instead we considered either sixth-grade or eighth-grade IGAP reading scores.

<sup>65</sup>Bryk, et al. (1994).

<sup>66</sup>For analyses of the East Harlem experience, see Kirp (1992) and Harrington and Cookson (1992).

<sup>67</sup>For example, Easton and Storey (1994).

---

## REFERENCES

- Ballantine, Jeanne H.**, *The Sociology of Education: A Systematic Analysis*, Englewood Cliffs, NJ: Prentice Hall, 1993.
- Brown, Daniel J.**, *Decentralization and School-Based Management*, New York: The Falmer Press, 1990.
- Bryk, Anthony S., Paul E. Deabster, John Q. Easton, Stuart Luppescu, and Yeow Meng Thum**, "Measuring achievement gains in the Chicago public schools," *Education and Urban Society*, Vol. 26, May 1994, pp. 306-319.
- Capell, Frank J.**, *A Study of Alternatives in American Education, Volume VI: Student Outcomes at Alum Rock 1974-1976*, (R-2170/6-NIE), Santa Monica, CA: The Rand Corporation, July 1981.
- Chicago Panel on Public School Policy and Finance**, *Chicago Public Schools Databook, School Year 1988-1989*, Chicago: Chicago Panel on Public School Policy and Finance, 1990.
- Chicago Public Schools, Department of Research, Evaluation and Planning**, "Report of Chicago Public Schools achievement trends," Chicago: Chicago Public Schools, October 26, 1994.
- City of Chicago, Department of Planning**, *Chicago Statistical Abstract: Community Area Profile*, 1994.
- Downes, Thomas**, "Evaluating the impact of school finance reform on the provision of public education: The California case," *National Tax Journal*, Vol. 45, December 1992, pp. 405-419.
- Downes, Thomas A., and Thomas F. Pogue**, "Adjusting school aid formulas for the higher cost of educating disadvantaged students," *National Tax Journal*, Vol. 47, March 1994, pp. 89-110.
- Easton, John Q., and Sandra L. Storey**, "The development of local school councils," *Education and Urban Society*, Vol. 26, May 1994, pp. 220-237.
- Education and Urban Society*, Vol. 26, May 1994, pp. 206-319.
- Ferguson, Ronald F.**, "Paying for public education: New evidence on how and why money matters," *Harvard Journal on Legislation*, Vol. 28, Summer 1991, pp. 465-498.
- Flinspach, Susan Leigh, and Susan P. Ryan**, "Diversity of outcomes: Local schools under school reform," *Education and Urban Society*, Vol. 26, May 1994, pp. 292-305.
- Grady, William, and Karen M. Thomas**, "City school reform plan is ruled illegal," *Chicago Tribune*, December 1, 1990, pp. 1 and 6.
- Hannaway, Jane, and Martin Carnoy (eds.)**, *Decentralization and School Improvement*, San Francisco: Jossey-Bass Publishers, 1993.
- Hanson, Marjorie K., Don R. Morris, and Robert A. Collins**, "Empowering teachers in Dade County's school-based management pilot," in *Empowering Teachers and Parents: School Restructuring Through the Eyes of Anthropologists*, G. Alfred Hess, Jr. (ed.), Westport, CT: Bergin and Garvey, 1992, pp. 71-87.
- Hanushek, Eric A.**, "The economics of schooling: Production and efficiency in public schools," *Journal of Economic Literature*, Vol. 24, September 1986, pp. 1141-1177.
- Harrington, Diane, and Peter W. Cookson, Jr.**, "School reform in East Harlem: Alternative schools vs. 'schools of choice,'" in *Empowering Teachers and Parents: School Restructuring Through the Eyes of Anthropologists*, G. Alfred Hess, Jr. (ed.), Westport, CT: Bergin and Garvey, 1992, pp. 177-186.
- Hatry, Harry P., et al.**, *Implementing School-Based Management: Insights into Decentralization from Science and Mathematics Departments*, Washington, DC: The Urban Institute, 1993.
- Hess, G. Alfred, Jr.**, *Chicago School Reform: What It Is and How It Came to Be*, Chicago: Chicago Panel on Public School Policy and Finance, November 1990.

\_\_\_\_\_, *School Restructuring, Chicago Style: A Midway Report*, Chicago: Chicago Panel on Public School Policy and Finance, February 1992a.

\_\_\_\_\_, "Through the eyes of anthropologists," in *Empowering Teachers and Parents: School Restructuring through the Eyes of Anthropologists*, G. Alfred Hess, Jr. (ed.), Westport, CT: Bergin and Garvey, 1992b, pp. 229-243.

**Katz, Michael B.**, "Chicago school reform as history," *Teachers College Record*, Vol. 94, Fall 1992, pp. 58-71.

**Kirp, David L.**, "What school choice really means," *The Atlantic Monthly*, No. 270, November 1992, pp. 119-132.

**Lane, John J.**, "Instructional leadership and community: A perspective on school-based management," *Theory Into Practice*, Vol. 30, Spring 1991, pp. 119-123.

**Meyer, Robert H.**, "Can schools be held accountable for performance? A critique of common educational performance indicators," University of Chicago, Harris Graduate School of Public Policy, working paper, 1993-94.

**Mirel, Jeffrey**, "School reform, Chicago style," *Urban Education*, Vol. 28, July 1993, pp. 116-149.

**Oates, Wallace E.**, *Fiscal Federalism*, New York: Harcourt Brace Jovanovich, 1972.

\_\_\_\_\_, "Searching for Leviathan: An empirical study," *American Economic Review*, Vol. 75, September 1985, pp. 748-757.

**Pink, William T.**, "The politics of reforming urban schools," *Education and Urban Society*, Vol. 25, November 1992, pp. 96-113.

**Rogers, David, and Norman H. Chung.** *110 Livingston Street Revisited*, New York: New York University Press, 1983.

**Rosenkranz, Todd**, "Reallocating resources: Discretionary funds provide engine for change," *Education and Urban Society*, Vol. 26, May 1994, pp. 264-284.

**Sander, William**, "Expenditures and student achievement in Illinois," *Journal of Public Economics*, Vol. 52, October 1993, pp. 403-416.

**State of Illinois**, Department of Education, "School report cards," data file, November 1988 - November 1994 (annual).

**U. S. Department of Commerce**, Bureau of the Census, *Census of Population*, 1994.

**Walberg, Herbert J., M.J. Bakalis, J.L. Bast, and S. Baer**, "Reconstructing the nation's worst schools," *Phi Delta Kappan*, Vol. 70, June 1989, pp. 802-805.

**Wohlstetter, Priscilla, and Karen McCurdy**, "The link between school decentralization and school politics," *Urban Education*, Vol. 25, January 1991, pp. 391-414.

**Wong, Kenneth K., and Gail L. Sunderman**, "Redesigning accountability at the systemwide level: The politics of school reform in Chicago," in *Midwest Approaches to School Reform*, Thomas A. Downes and William A. Testa (eds.), Chicago: Federal Reserve Bank of Chicago, 1995, pp. 162-187.

**Zax, Jeffrey S.**, "The effects of jurisdiction types and numbers on local public finance," in *Fiscal Federalism: Quantitative Studies*, Harvey S. Rosen (ed.), Chicago: University of Chicago Press, 1988, pp. 79-103.

## ECONOMIC PERSPECTIVES

*Public Information Center*  
Federal Reserve Bank of Chicago  
P.O. Box 834  
Chicago, Illinois 60690-0834

**Do not forward**  
**Address correction requested**  
**Return postage guaranteed**

### **Mailing label corrections or deletions**

Correct or mark **Delete from mailing list** on the label and fax it to 312-322-2341, or send to:

Mail Services  
Federal Reserve Bank of Chicago  
P.O. Box 834  
Chicago, Illinois 60690-0834

BULK RATE  
U.S. POSTAGE  
**PAID**  
CHICAGO, ILLINOIS  
PERMIT NO. 1942

FEDERAL RESERVE BANK  
OF CHICAGO