

$$1) \beta_j = \Phi_{1j}$$

$$2) \beta_j = \Phi_{2j}$$

$$\Phi_{3j} = 1 - \Phi_{1j} - \Phi_{2j}$$

to specify all three possible outcomes, it is ties, with Category 3 serving as a reference tural models were estimated:

$$\beta_{0(k)} + \beta_{1(k)} X_{1j} + \dots + \beta_{p(k)} X_{pj}$$

$$\beta_{0(k2)} + \beta_{1(k2)} X_{1j} + \dots + \beta_{p(k2)} X_{pj}$$

models:

$$\dots + \gamma_{0(k)} W_{aj} + \gamma_{0(k)} X_{1j}$$

$$\dots + \gamma_{0(k2)} W_{aj} + \gamma_{0(k2)} X_{1j}$$

the same set of predictor variables and con- vel achievement growth model included transfer, so the predicted learning rates were

Promise, Progress, Problems, and Paradoxes of Three Phases of Accountability: A Longitudinal Case Study of the Baltimore City Public Schools

Samuel C. Stringfield

University of Louisville

Mary E. Yakimowski-Srebniak

Council of Chief State School Officers

Issues associated with attempting to achieve academic equity while raising achievement gains for all Americans are being negotiated at a large scale and with particular urgency in our urban school districts. Building from a decade of archival data, 6 years of student-level quantitative data, semi-structured interviews, document analyses, and observations of key informants, the authors examine the long-term interplay of shifting state and federal policies related to accountability requirements, organizational responses, and student outcome measures in the Baltimore City Public School System, a large, high-poverty, majority-minority urban school system. Analyses conducted from 1992 through the spring of 2003 are presented in light of both increasing accountability requirements and national and state calls for urban school reform. The authors conclude by examining possible implications for districts and states serving large concentrations of students at risk.

KEYWORDS: assessment, at-risk students, educational equity, systemic reform, urban school reform.

In this article, we present data from a series of interrelated efforts to cast useful light on the challenges associated with attempting to increase both equity and achievement among urban children. We examine a decade of efforts to improve student outcomes at the elementary through high school levels in the Baltimore City Public School System (BCPSS), a large, high-poverty central-city school system. After an overview of the context of and need for reform in the BCPSS, we review research on school reform in the United States, particularly as it has affected urban school districts. As will be seen, accountability has been the watchword in urban school reform; this has resulted in both a flood of student-, school-, and district-level data and growing criticism of the increasing emphasis on

test preparation activities. Several recent studies of reform efforts in urban districts are highlighted.

We then shift our focus to one historically challenged urban setting, the BCPSS, and describe three outcome-focused phases of efforts to use state and federally mandated accountability measures to increase the academic achievement of Baltimore's children: (a) an initial phase of testing and measuring during which practical support was not provided to the district; (b) a phase involving increased changes in governance and leadership; an infusion of new funds, and the forming of an improved city-state partnership; and (c) a still-ongoing phase driven by the requirements of the federal No Child Left Behind (NCLB) legislation. These phases are detailed over the period from fall 1992 to spring 2003. We discuss outcome data associated with each phase, focusing in particular on the accountability shifts experienced by the state and the BCPSS.

The discussion section focuses on the promises and paradoxes of accountability-based reform efforts, in the case of both Baltimore and other large, diverse urban districts. What was found in the case of BCPSS was that the state, at least in Accountability Phase 1, did not achieve the results it desired. The threat of state takeover (consistently underperforming schools could be deemed "reconstitution eligible" [RE]) produced high levels of stress among teachers and administrators in low-performing schools, but the state's stepped-up testing measures, while shedding valuable light on the high levels of need in the BCPSS, did not produce marked improvements in student learning. Furthermore, the state's focus on school-level improvement failed to take into account the systemic nature of many of the problems plaguing the BCPSS.

If Accountability Phase 1 of the BCPSS reforms was marked largely by problems and paradoxes, Accountability Phase 2, with its emphasis on a shared accountability system called the City-State Partnership and an infusion of financial and technical support from the state, represented a new era of steady, continuous progress in which student test scores and graduation rates began to rise. Working closely with the state department of education,

SAMUEL C. STRINGFIELD is the Academic Director of the Nystrand Center of Excellence in Education and a professor in the Departments of Leadership, Foundations, and Human Resource Education and Teaching and Learning at the University of Louisville, College of Education and Human Development, Room 332, Louisville, KY 40292; e-mail sam.stringfield@louisville.edu. From 1999 through 2004, he was an appointed member of the Board of School Commissioners of the Baltimore City Public School System. His research interests include teacher and school effects, systemic effects, school boards, and international issues in school effectiveness.

MARY E. YAKIMOWSKI-SREBNICK is Director of Assessments, Council of Chief State School Officers, One Massachusetts Avenue, NW, Suite 700, Washington, DC 20001-1431; e-mail Yakimowski@aol.com. Her research focus is on assessment, urban education, program evaluation, and school improvement.

the BCPSS was making progress, but it still faced significant obstacles as it moved into Accountability Phase 3. The NCLB legislation, which emphasizes accountability-driven reform, increased pressure on the state to comply with federal mandates related to assessment, accountability, content standards, and data collection and reporting procedures. Throughout 2003, the Maryland State Department of Education (MSDE) made a concerted effort to meet these demands while continuing to provide support at multiple levels.

In Baltimore's case, however, NCLB's requirements have not been without paradox. NCLB's recommendation to turn over management of low-achieving schools to for-profit corporations has already been proved a failure in Baltimore, where three RE elementary schools, run by a for-profit corporation since 2000, continued making little progress in RE status in 2003. The BCPSS, under NCLB guidelines, is itself technically in a phase of "corrective action" and in need of reconstitution; yet, the system has been in such a state for more than 6 years, and it ranks as one of the fastest improving districts in Maryland.

Baltimore as Context

The past is never dead. It's not even past. (Faulkner, 1959)

For most of America's history, Baltimore was one of the largest cities in the country. After an economic boom built on the nation's World War II need for steel, shipbuilding, and Atlantic shipping, Baltimore's total population peaked in 1950 at 949,000, of whom 225,000 (24%) were African American. The subsequent half century saw a continuous population decline that has only recently abated. As of 2003, the U.S. Bureau of the Census estimated the city's population at just over 651,000, of whom approximately 424,000 (65.2%) were African American.

The total population of students served in the BCPSS peaked at 199,000 in the early 1970s and stood at 96,304 during the 2002-2003 school year (see Table 1). During that school year, the student body was 86.2% African American, and more than 80% of BCPSS students were eligible for the free/reduced-price lunch program. As can be seen in Table 1, the city schools are serving a population that is increasingly minority and poor. The system comprises 184 schools, ranging from small primaries to very large high schools and including variations such as high schools with multiple "small learning communities" as well as schools designed to serve small numbers of students with specific medical needs.

It would be almost impossible to talk about Baltimore or Baltimore education without discussing race/ethnicity (Orr, 1999). Prior to the Civil War, Baltimore was home to both legal slavery and the largest free Black population of any city in the United States.¹ Baltimore began desegregating its schools in 1952 and did so more aggressively after the Supreme Court's 1954 *Brown v. Board of Education* decision. Patricia Welch, current chair of the BCPSS Board of Commissioners, vividly remembers attending an overcrowded, two-room,

Table 1
BCPSS Enrollment by Race/Ethnicity and Free/Reduced-Price Meal Program (FARM) Eligibility

Year	American Indian	Asian	Black	White	Hispanic	Total	Black (%)	White (%)	Hispanic (%)	FARM (%)
2003	288	583	82,775	9,252	1,133	96,034	86.2	9.6	1.2	83.0
2002	323	339	83,719	9,967	907	97,257	86.1	10.2	0.9	81.5
2001	350	594	85,778	10,768	726	100,217	85.6	10.7	0.7	76.2
2000	381	607	89,793	11,619	600	105,000	85.5	11.1	0.6	75.0
1999	417	617	92,092	12,953	524	108,602	84.8	11.9	0.5	68.4
1998	453	636	92,055	13,789	483	109,414	84.1	12.6	0.4	67.9
1997	493	614	92,570	14,626	456	110,756	83.6	13.2	0.4	66.8
1996	515	646	92,665	15,730	424	111,976	82.8	14.0	0.4	70.1
1995	500	587	94,866	17,075	400	115,423	82.2	14.8	0.3	68.7
1994	524	525	92,973	17,703	369	114,088	81.5	15.5	0.3	69.4
1993	452	519	91,029	18,297	365	112,655	80.8	16.2	0.3	67.7

Note. BCPSS = Baltimore City Public School System.

Case Study of the Baltimore City Public Schools

wooden, wood-store-heated all-Black school just two blocks away from a new, brick, centrally heated all-white school. She recalls being one of the first dozen young Black women at the historically all-white, college-preparatory Eastern High School. "The administration didn't want us there, the teachers didn't want us there, and the students didn't want us there . . . but we worked hard, and graduated" (Welch, personal communication, March 2002).

By the mid-1960s, Baltimore's economic position had slipped to a sufficient degree that the BCPSS was spending 18% less per student per year than the surrounding Baltimore County area. By the mid-1980s, given a rapidly declining tax base caused by industrial decline, "White flight," and, increasingly, middle-class Black flight to the suburbs, the BCPSS ranked 22nd among 24 Maryland school districts in per-pupil spending (Bowler, 1997). Schools were increasingly in disrepair, and teacher pay had fallen several thousand dollars per year below that of surrounding jurisdictions. A series of reports from both White business groups and Black community leaders between the late 1980s and mid-1990s—each calling for the combination of increased funding, multiple reforms, and increased accountability—failed to bring significant change to the system on any of these dimensions (Bowler, 1997; Chulka, 2003a, 2003b; Orr, 1999). By a range of measures, Baltimore presents a challenging context in which to attempt educational reform.

Urban Schools: Accountability and School Reform

For more than 20 years, the United States has been in a period of continuous calls for dramatic school reform, especially in our urban districts. The calls have been for improved processes and outcomes, the most significant outcomes being the raising of student achievement scores and high school graduation rates. Virtually all efforts have combined a strong moral call for social justice with demands for implementation of various "research-based" technical fixes. Yet, few have pressed for data on whether these reforms have been working to boost urban students' achievement levels (Snipes, Doolittle, & Heitly, 2002).

At various points, these calls have focused on the use of research on teacher effectiveness (e.g., Brophy & Good, 1986; Rosenshine & Stevens, 1986) and school effects (e.g., Edmonds, 1979; Teddlie & Reynolds, 2000), implementation of promising programs (Berends, Chun, Schuyler, Stockly, & Briggs, 2002; Borman, Hewes, Overman, & Brown, 2003; Darnow, Borman, Stringfield, Radaba, & Castellano, 2003), and institution of systemic reform (O'Day, 2002; O'Day & Smith, 1993). Calls for reform have been so continuous that, currently, "reform is the status quo" (F. Hess, 1999, p. 5).

Increasing Public School Accountability

Perhaps the most unifying theme of the national- and state-level calls for public school reform has been the demand for increased accountability. This typically takes the form of specific measures of outcomes at the student, school, and district levels (Jain, 2003; Rothman, 2003). Olson (2003) concluded that,

to date, NCLB has produced "one unambiguous result: an avalanche of data on the performance of public schools in the United States" (p. 1).

Beyond the production of more data, responses to demands for increased levels of academic accountability have been mixed. While most policy-makers and practitioners would agree that accountability is "everyone's concern, everyone's job" (Wolf, 2003, p. 4), many are far from sanguine about the narrow focus of most of the current generation of measures (Herman, 2003; Linn, 2003). Several have expressed concern that such narrow measures are resulting in increased time spent studying test preparation materials (McNeil & Valenzuela, 2001). Some scholars caution that high-stakes accountability systems may not improve actual learning and may lead to unethical situations (Amrein & Berliner, 2002; Elmore, 2002; O'Day, 2002). Fine, Bloom, and Chajet (2003) have seen America's current testing regimen, in which at-risk urban students are held accountable for meeting high academic standards even though their schools, texts, and libraries and the qualifications of their teachers often do not meet minimum established standards, as a deep betrayal of our nation's most vulnerable young people. The reservations of these and other scholars notwithstanding, under the current generation of systemic reforms generally, and under the federal NCLB in particular, aggregated scores on various states' designated achievement measures have become key components of accountability for America's public schools.

The Need for Urban School Reform

A primary focus of these reform and accountability efforts has been America's urban schools and school systems (Cuban & Usdan, 2003). Urban districts have disproportionately large percentages of students who are in danger of not thriving in school (Stringfield & Land, 2002). They are more likely to live in poverty; to suffer health-related problems; to live in single-parent, single-grandparent, or foster-parent homes; and more likely to have been exposed to violence, lead from peeling paint, and a wide range of other risks that are negative predictors of academic achievement (Land & Legters, 2002).

The enormity of the academic challenge currently facing students, their families, and schools in America's urban districts was highlighted in a recent set of reanalyses of scores from the National Assessment of Educational Progress (NAEP). Flanagan and Grissmer (2002) disaggregated a decade of NAEP data by four regions (Northeast, Midwest, West, and Southeast) and, further, by three types of localities (rural, suburban, and central city). Statistically averaging scores from several rounds of NAEP allowed these authors to conduct a detailed examination of the resulting 12 cells. They found stark contrasts. The four cells with the lowest scores were the four central-city groupings, regardless of region. When the data were further broken down according to race/ethnicity, four of the six lowest-scoring groups were African Americans living in the central cities of the four regions. These four groups had average scores that were between 0.79 and 0.84 standard deviations below the national average.

Similarly, Balfanz and Legters (2003) analyzed national data from the Common Core of Data to determine areas with high dropout rates. They found that the substantial majority of high schools in which ninth-grade classes were at least 70% larger than the graduating classes 4 years later were located in urban districts of the "rust belt" (stretching along the east coast from Boston to Washington, D.C., and west to Indianapolis and Chicago) or in Atlanta, Birmingham, Nashville, or one of several cities in Texas. It is reasonable to assume that much of the success or failure of the current generation of U.S. school reform will be judged on the extent of improvement in the measured achievements of our most vulnerable children in urban schools.

Research on Reform in High-Poverty Urban Areas

Over the past decade, there has been an increase in the number of studies of reform efforts in high-poverty urban districts. Chicago's reform efforts have been particularly well chronicled (Bryk, Sebring, Kerbow, Rollo, & Easton, 1998; A. G. Hess, 1995, 2003). Hess (2003) presented a detailed analysis of Chicago students' outcomes on various measures over a decade. Two measures on which national comparisons were most available were the percentages of sixth-grade students scoring at or above national norms on the Iowa Test of Basic Skills (ITBS) and graduation rates. The percentage of students scoring at or above the national average on the ITBS reading section rose from 26% in 1993 to 35% in 1998, a 9-point gain. On the ITBS mathematics test, the percentage increased from 35% to 41%, a rise of 6 percentage points. Over the 9 years from 1990 to 1998, Hess calculated that Chicago's actual dropout rate decreased from 45.9% to 41.7%, an 8-year improvement of 4.2% that was significant in this very hard-to-move measure of school reform effects.

The Council of Great City Schools recently completed a set of case studies of urban school systems that had demonstrated substantial improvements in student achievement (Snipes et al., 2002). To be included, the systems had to have demonstrated a multiyear trend of improved-overall student achievement that exceeded that of their state. The study identified and focused on four large, urban districts in which at least 35% of students received free or reduced-price lunches and in which the percentage of White non-Hispanic students was below 50%. These districts were Houston, Charlotte-Mecklenburg, Sacramento, and New York City's (former) Community District #2. All four districts had made impressive gains on state tests in the elementary grades. However, while these and several other urban systems have begun to produce significant gains at the elementary and middle school levels, demonstrated gains on high school measures, the most significant of which is high school graduation, have been less impressive. Most of the districts studied are only beginning to turn their attention to secondary school reform and gains in secondary education, and specifically to efforts to improve high school graduation rates.

Method

The present investigation was based on a longitudinal mixed-methodology (Tashakkori & Teddlie, 1998, 2003) case study (Yin, 1989) of a large, high-poverty district's experiences during a decade of accountability-focused reform efforts. A mixed-method design was chosen in a pragmatic effort to capture the widest range of effects of accountability efforts (the *what* and *so what* of reform efforts) together with a range of participants' perspectives of *how* and *why* various reforms were attempted. Following this pragmatic methodological approach (Howe, 1988), we made use of seven types of data, as described in the following sections.

Time Frame

This case study covered the period between fall 1992 and spring 2003 and involved a particular emphasis on the effects over time of changes in accountability: both who is accountable and for what. Most interviews were conducted between the spring of 2002 and the summer of 2003; the focus was on the period of the most active reforms, beginning in 1997, but interviews with advocates often covered events occurring throughout the 1990s and through the summer of 2003. As can be seen in Table 2, the various strands of quantitative data were available for different periods. The city-state comparative achievement data involving the Maryland School Performance Assessment Program (MSPAP) were available for the years 1993–2002, after

Table 2
Data Sources for Baltimore Case Study

Data type	Time frame for which data were available
Qualitative	
Document analyses (prior scholarship, reports, newspapers, board minutes, etc.)	1982–2003
Interviews (board members, community activists, others)	2000–2003
Author notes and periodic journals	1998–2003
Quantitative	
MSPAP (performance assessment)	1992–2002
TerraNova (NRT)	1998–2003
Graduation Rates (MSDE Analyses)	1996–2002
Graduation Rates (Johns Hopkins Analyses)	1997–2002
Maryland School Assessment	2003

Note. MSPAP = Maryland School Performance Assessment Program; NRT = norm-referenced testing; MSDE = Maryland State Department of Education.

which the state discontinued that test. The BCPSS had discontinued gathering of norm-referenced testing data prior to the 1997 establishment of the new board, and hence CTB/TerraNova information was available for each spring assessment from 1998–2003. Maryland began systemically reporting high school dropout and graduation data for the class of 1997 and then altered the definitions of these data in 2003; hence, 1997–2002 data are used. As a check against these state analyses, Mac Iver, Farley, and Wayman (2003) conducted supplemental analyses of high school graduation rates using student-level longitudinal files covering the graduating classes of 1998–2002.

Qualitative Data

Interview Data

Detailed, semistructured interviews were conducted with seven of the original nine members of Baltimore's new Board of School Commissioners appointed under Senate Bill 795. Initial interviews were conducted during the spring of 2001, with follow-up interviews conducted through 2003.² These interviews were transcribed and analyzed according to the guidelines of Yin (1989) and Miles and Huberman (1994), with particular attention to setting of goals and system-level processes such as allocation of funds for teacher recruitment and professional development and issues related to student assessments. Between 2000 and 2003, additional interviews related to specific assessment topics were conducted with members of the local media, the deputy state superintendent of education for accountability, key members of advocacy groups such as the American Civil Liberties Union, and staff of the Council of Great City Schools.

Archival Data/Document Analyses

Four studies examining Baltimore educational reform efforts prior to the 1997 reforms or the politics of current reforms provided valuable contextual information for the present study (Bowler, 1997; Orr, 1999; Cibulka, 2003a, 2003b; Westat, 2001). Archival reports from the MSDE³ and the BCPSS,⁴ as well as newspaper accounts and results from an independent evaluation⁵ of various aspects of BCPSS operations, were gathered and analyzed.

Participant Observations

The first author was appointed to the BCPSS school board in 1999, 2 years after its re-formation; he resigned in 2004. During his tenure on the board, he gathered detailed notes on the formal and informal proceedings of the system. The second author served as chief of educational accountability for the BCPSS Division of Research, Evaluation, Assessment, and Accountability from 2000 to 2004. These circumstances allowed for rich data gathering and syntheses.

Quantitative Outcome Measures

Four types of quantitative student outcome measures are reported here. The first, the MSPAP tests, were designed to be "performance measures."⁶ The MSPAP involved no multiple-choice items; students wrote all responses, regardless of content area. Primacy was placed on display of academic skills, especially higher order skills deemed to be important in a world in which no citizen could know all of the facts of a content area but might be able to develop skills in accessing specific facts. Parts of most of the MSPAP subtests involved cooperative learning among the students, with each student then being responsible for writing up one aspect of the group's work.⁷ A matrix sampling methodology was used for the administration of the MSPAP to students in the third, fifth, and eighth grades. MSPAP results are presented in terms of the percentages of students at the school, district, and state levels at excellent and satisfactory levels in terms of each of the six content areas assessed. Also presented are overall composite and school performance indices.⁸

MSPAP test booklets were blind-scored on a statewide basis. While substantial efforts were made to standardize scoring across years, in fact, state-average scores varied as former scorers departed and new scorers were added. Equally problematic, teachers' and schools' motivations to excel on the test varied over time, with an increase in the first few years, as it became obvious that the MSPAP was potentially a "high-stakes" test for schools and districts, and a decrease in the final year. The most plausible reason for the statewide decline in motivation in the final year was that 1 week before the MSPAP administration in the spring of 2002, the MSDE announced that the test would be discontinued and that a new test would be developed and administered the following year to ensure alignment with newly implemented federal accountability legislation.

Given these conditions, it is probable that the most meaningful MSPAP analyses are concerned not with absolute scores but, rather, with the gap between any given school or local education authority (LEA) and the state. While the MSPAP was not without its critics, it was very well received within the academic community (e.g., Herman, 2003), and it earned Maryland an "A" in *Education Week's* first "Quality Counts" (1997) report. We report MSPAP data longitudinally, with an emphasis on changes in the Maryland-BCPSS gap.

In addition to school-level test scores from the MSPAP, the annual district and school report card produced by the MSDE, called the Maryland School Performance Program (MSPP), included reports on student attendance and, at the high school level, dropout rates and percentages of students passing the state's minimum-competency functional tests. MSPAP test scores, attendance rates, functional test scores, and graduation rates were combined to form a school performance index (SPI) used by the state to indicate both a school's academic status and progress over time.

The second student outcome data source we used was the norm-referenced TerraNova test battery from CTB/McGraw-Hill (2001). Reliability

coefficients for this battery have been found to be within acceptable ranges.⁹ The third outcome variable examined was high school graduation rate. Greene and Caire (2001) found that graduation from high school is a particularly important measure, in part because the economic disparity between individuals who earn a high school diploma and those who fail to do so is approaching 100%.¹⁰

The longest term graduation data available for this study were those from MSDE reports (MSDE, 2003). The MSDE graduation rate is an estimate of the percentage of students who entered Grade 9 and received a Maryland diploma 4 years later. This estimated cohort graduation rate is calculated by dividing the number of high school graduates by the sum of students in that class who dropped out in each of the previous 4 years and the number of high school graduates.

Because the MSDE measure is inexact, we also used a second metric for measuring BCPSS dropout rates. Mac Iver et al. (2003) developed longitudinal, individual student-level files for entering BCPSS ninth graders from 1994 through 2002. These authors followed individual students and computed the percentage of students who graduated from the same or any other BCPSS school 4 years later. The Mac Iver et al. dropout rate analyses were compared with the state's longitudinal analyses.

The final student outcome data were derived from the Maryland School Assessment (MSA) program. MSA was initiated in the spring of 2003 as part of Maryland's compliance with the recently enacted NCLB. Measures gathered in the spring of 2003 were from students in Grades 3, 5, and 8 in the content areas of reading and mathematics. Results are presented in terms of the percentages of students in the basic, proficient, and advanced performance categories, in alignment with the assessment requirements of NCLB.¹¹

Three Accountability Phases of Maryland Educational Reform as They Have Related to Baltimore City

Accountability Phase 1: Introduction of the Standards Movement

The first phase of accountability-based reform in Maryland education came as part of the state's response to the federal *A Nation At Risk* (National Commission on Excellence in Education, 1983) report. In 1987, Maryland's governor asked the MSDE to appoint a commission to conduct a detailed study of the preparedness of Maryland's students to perform in the modern economy. The MSDE released the commission's report in August 1989. The authors concluded that Maryland's students were not prepared to succeed in the new economy and that new, academically demanding state tests were necessary to accurately measure schools' performance in preparing students (MSDE, 1989). The stated purpose and promise of the state's reforms were to increase academic achievement among all students through increased accountability on the part of all schools and districts. The implicit theory of action was that meaningful improvement was possible at low cost to the state, simply by setting

standards, publishing school- and district-level results of tests of those standards, pointing out individual schools that were not scoring well or producing measures of progress, and encouraging districts to improve the limited numbers of low-scoring schools. The MSDE responded to the report by funding the development of the MSPAP, as described earlier.

As has been the case in many other states, the release of the first annual MSPAP state report card produced a great deal of media coverage and discussion. For Baltimore City, the data were distressing: BCPSS students' MSPAP scores were at the bottom of the state's 24 districts in every subject area and every grade. In terms of student achievement scores, the BCPSS was, by a substantial margin, the state's furthest negative outlier. These data were a shock to the state, the city, and the BCPSS.

Over the next 4 years, BCPSS's MSPAP news only worsened. As can be seen in the top half of Table 3, the overall average of Maryland students' scores on the MSPAP subtests (as reported in the cumulative index, or CI) rose by 10.1 points. In contrast, the BCPSS produced a modest 3.5-point gain in CI scores over these 4 years. In 1996, not only was the BCPSS still producing by far the lowest student achievement scores in the state, but each year it was falling further behind the state average.

The state's implicit theory of action in Accountability Phase 1 did not produce the intended result in Baltimore. One component of the theory had been that external threats and efforts to induce crisis would result in improvements among the lowest-performing schools. However, as the data in Table 3

Table 3
MSPAP Composite Index (CI) Scores for BCPSS and Maryland
and CI Gap Analysis

Year	BCPSS	Maryland	State-city gap	Annual change in gap
Phase 1				
1993	10.4	31.7	21.3	
1994	11.7	35.3	23.6	2.3
1995	13.8	39.6	25.8	2.2
1996	13.5	40.7	27.2	1.4
1997	13.9	41.8	27.9	0.7
Total change				+6.6
Phase 2				
1998	16.1	44.1	28.0	0.1
1999	17.0	43.8	26.8	-1.2
2000	20.5	45.3	24.8	-2.0
2001	22.5	43.7	21.2	-3.6
2002	20.4	39.1	18.7	-2.5
Total change				-9.2

Note. MSPAP = Maryland School Performance Assessment Program; BCPSS = Baltimore City Public School System.

demonstrate at a system level, and Mintrop (2003) demonstrated at the school level; external threats may induce anxiety and stress among teachers and administrators, but these emotional states do not create an automatic trigger of learning among students.

The MSPAP program had imbedded in it the potential to declare schools as RE, and if an RE school's low scores did not improve over time, it could be declared "state reconstituted." But the first schools declared RE—all in Baltimore City—did not significantly improve. Clearly, the schools lacked the necessary capacity for constructive action to go along with the pressure, and the result was that the majority of the RE schools made little or no progress toward the state average.

A second problem was related to the first. The system's efforts to target improvement were at best chaotic, and they were clearly inconsequential. The MSDE response had been to send in monitors with checklists, critiques, and suggestions, largely related to teaching skills related to test taking. To the extent that the criterion was measured improvement, both of the strategies failed. Not only did the individual schools lack the capacity to improve, but the BCPSS, and in some regards the MSDE, lacked the capacity to successfully guide improvement in historically underperforming schools.

A third problem that had not been theorized at all in the MSPP eventually came into focus. By definition, the assessment portion of the MSPP assumed that the core assessment needed to take place at the school level. However, during the first 5 years of the MSPP, 50 of the 52 Maryland schools declared RE were in Baltimore City. Clearly, a school-level assessment had identified deep, system-level problems.

Politically, other forces, such as pressure from Baltimore's mayor and a school funding adequacy case working its way through the state courts, had some effect (Cibulka, 2003a, 2003b). However, it was the annual, statewide publication of the MSPP report card, which included MSPAP data clearly demonstrating the dire needs of BCPSS students and the lack of progress under the then-current arrangements, that persuaded the city's and state's political powers to take action (M. Bowler, personal communication, December 2003; B. Verdery, personal communication, December 2003).

At the district level, Accountability Phase 1—principally the introduction of statewide testing and identification of the BCPSS as a deeply challenged district—had produced minimal evidence of positive effects. The one clear effect of Phase 1 was that the accountability measures left no doubt that BCPSS students were in need of substantially improved schooling and that the then-current reform efforts were not succeeding. While a range of other legislative and court-based change efforts had been attempted, it seems clear that the depressingly low accountability scores on state-mandated tests contributed directly to the passage of Maryland's Senate Bill 795 and initiation of Accountability Phase 2 in Baltimore.

A striking paradox of Accountability Phase 1 was that Maryland's early reform efforts had been developed in the hope that a focus on measures of student outcomes and a public shaming of low-performing schools would

provide a low-cost solution to Maryland's and Baltimore's educational needs. Over time, the testing regimen had the opposite effect. Testing alone produced no significant effect on Baltimore's educational processes or its students' measured outcomes; however, as described subsequently, the intended-to-be-low-cost testing program eventually led to an unprecedented increase in the state's fiscal support for the BCPSS.

Accountability Phase 2: Senate Bill 795 and District Reform

In large part on the basis of the MSPAP-measured failure of Accountability Phase 1 reforms to produce measurable increases in student outcome targets for BCPSS students and schools, Baltimore's mayor, the state superintendent of instruction, the governor, and two key members of the Baltimore delegation of the state legislature negotiated a series of changes that took effect in the spring of 1997. This second round of reforms was formalized in Senate Bill 795, a carefully crafted educational improvement plan that represented a compromise.

Senate Bill 795 mandated the following: (a) a new, nine-member board of school commissioners¹² drawn from a list of nominees provided by the MSDE, with each new member jointly recommended by the mayor and the governor (Baltimore's school board had historically been appointed by the mayor); (b) increased state funding for the BCPSS (the exact levels of these additional funds were negotiated up to the final hours but eventually included a \$30 million increase in Fiscal Year 1998, followed by an additional \$50–\$55 million in each of the subsequent 4 years); (c) creation of a transition plan and a 5-year "master plan" to provide direction and oversight for the system; (d) establishment of the positions of chief executive officer (CEO), chief academic officer, director of research and evaluation, and chief financial officer and a parent and community advisory board; and (e) actions designed to address the specific recommendations of (previously ignored or minimally implemented) various court cases, citizen commissions, and so forth. In addition, these mandates included establishing better business systems, building better district-level databases, increasing parent involvement, and, ultimately, improving student outcomes.

The new school board began work in the spring of 1997. It replaced 14 of the most senior executives (C. Daniel, personal communication, May 2002) and worked with the new leadership team to develop a strategic master plan. The board and its newly appointed senior executives studied a range of options for new citywide reading/language arts curricula, held a series of public hearings, modified choices on the basis of community feedback, and purchased new reading/English texts and supporting materials for all students. The system then focused all 1998–1999 professional development activities on building teachers' skills in the use of the new reading curricula. The next year, the board repeated the process with K–12 mathematics texts and professional development, followed by the sciences, history/social studies, and the arts.

The board also created a process for allowing Comprehensive School Reform Design schools to pursue their reform models independently from many BCPSS requirements¹³; reduced student/teacher ratios in Grades 1–5 to 18:1; established prekindergarten and all-day kindergarten programs in most elementary schools; expanded before-school, after-school, and summer-school programs for the most academically at-risk students; targeted assistance to schools in need of improvement; and developed a "CEO's district" to provide particularly intensive services to a subset of the system's lowest-performing schools. Working with BCPSS and MSDE staff, the board developed more uniform promotion and retention policies, expanded professional development opportunities, and moved toward compliance with the nation's special education consent decree.

In terms of infrastructure, teacher and principal salaries were raised to levels competitive with surrounding jurisdictions.¹⁴ Mentors were trained and assigned to support new teachers in historically low-achieving schools. The system purchased and implemented a new student database and upgraded the court-mandated special education tracking system. In addition, the system accelerated the rate at which schools were provided with modern computer wiring and Internet access, greatly increased the numbers of computers in the schools, and quadrupled its school maintenance and repair budget.

Significantly, most of the efforts just described involved a system of shared accountability. The legislature and governor promised and provided extra funding, and the state department provided more coordinated technical assistance. Senate Bill 795 mandated that the first CEO of the new BCPSS not be a permanent hire. The board hired and worked with the initial, acting CEO and three subsequent CEOs to create a systemwide sense of accountability. When several senior executives failed to work within this new accountability system, they were replaced. Through retirement, demotion, termination, and other steps, BCPSS replaced more than half of its principals between 1997 and 2003.

The results of these shared-accountability/multiparty-supported reform efforts can be seen in three system-level trends, an independent evaluation of the system's progress, and two school-level trends. The three systemic trends concern the state's performance-based MSPAP examinations in Grades 3, 5, and 8, BCPSS norm-referenced TerraNova results, and data on high school graduation rates.

MSPAP Results

As noted earlier (see Table 3), during Accountability Phase 1, the BCPSS not only had the lowest MSPAP scores in the state but was annually falling further behind the state average, with the city-state gap widening by 6.6 CI points between 1993 and 1997. After passage of Senate Bill 795, the gap ceased expanding. The BCPSS produced essentially the same CI gain as the state in 1998, and it then narrowed the gap for 4 consecutive years, culminating in an overall Phase 2 gap reduction of 9.2 CI points in 2002.

Although the gap was still substantial, Baltimore City had broken a multi-year trend of expansion and was closing the MSPAP gap as measured against state figures.¹⁵

TerraNova Results

In the years leading up to the 1997 reforms, the BCPSS was in difficult financial straits and had discontinued administration of all standardized testing other than the MSPAP. At the insistence of the new board, the system re-initiated testing in the spring of 1998 and chose the CTB/McGraw-Hill TerraNova assessments for its norm-referenced testing program. Data from this program are presented in Table 4, and two features of these data are particularly notable. The first is the level to which achievement in the BCPSS had been allowed to fall prior to the 1997 reforms. In the spring of 1998, the system's median national percentile rankings in regard to norm-referenced testing were in the teens and twenties across the content areas and grades. By any reasonable measure, the BCPSS faced the task of recovering from a badly failed reform process.

Table 4
CTBS/TerraNova Median Percentile Scores and Gains:
BCPSS, 1998-2003

Subject area	Grade							
	1	2	3	4	5	6	7	8
Total reading								
1998	25	23	25	22	16	12		
1999	37	27	26	22	16	12		21
2000	46	38	34	29	35	18	25	22
2001	55	44	41	35	41	22	28	25
2002	59	44	41	37	40	30	31	30
2003	59	50	41	40	41	31	33	33
5-year gain	34	27	16	18	25	19	8 ^a	12 ^b
Total mathematics								
1997-1998	24	19	21	15	15	17		
1998-1999	23	21	18	15	15	17		20
1999-2000	37	32	32	26	28	21	22	20
2000-2001	51	41	41	33	34	24	24	21
2001-2002	54	43	43	35	36	29	28	32
2002-2003	58	52	45	41	41	31	32	37
5-year gain	34	33	24	26	26	14	10 ^a	17 ^b

Note. BCPSS = Baltimore City Public School System.

^a3-year gain.

^b4-year gain.

Case Study of the Baltimore City Public Schools

Second, the gains were dramatic in the primary grades, ranging from 16 to 34 national percentile points, and the gains were large in the upper elementary grades as well, ranging from 18 to 26 percentile points. Although the system did not test in eighth grade until 1999 or seventh grade until 2000, reading and mathematics scores rose in those grades as well, with median percentile gains ranging from 8 to 19 points. Although often based on measures and metrics different from those reported for the districts in the previously noted Council of Great City Schools analyses of test scores in rapidly improving districts, these gains would appear to compare favorably with the gains reported for the four districts the council identified as making exemplary progress as well as those from the A. G. Hess (2003) Chicago analysis.

High School Graduation Results

At the high school level, the most plausible measure of effect is graduation rate. Students who fail to graduate from high school face extremely difficult subsequent life challenges. Data comparing Maryland statewide graduation rates and those of the BCPSS over the years 1996-2002 are presented in Figure 1. The MSDE's calculation of the statewide graduation rate in the spring of 1996, just before the start of Phase 2 Baltimore reforms, was 80.2%. This rate rose over the 5 subsequent years by 3.6%, to 83.8%.¹⁶ The MSDE calculated

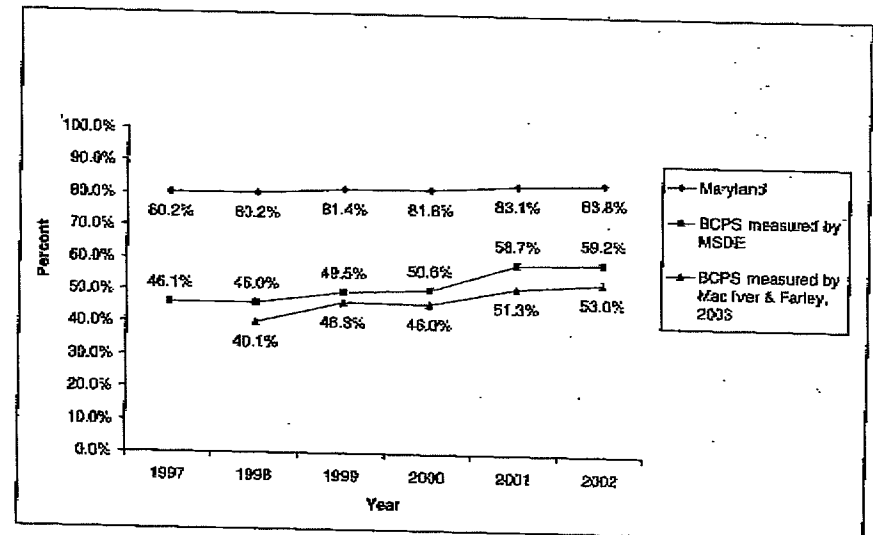


Figure 1. Overall comparison of graduation rates between the Baltimore City public schools (BCPS) and Maryland as a whole. MSDE = Maryland State Department of Education.

Baltimore City's graduation rate as 46.1% for the 1996–1997 school year, the most plausible baseline year for judging the effects of the 1997 reforms. The MSDE reported that, over the subsequent 5 years, the BCPSS graduation rate rose by a substantial 13.1%, to 59.2%.

Given various previously noted scholars' cautions regarding analyzing aggregated data when calculating graduation rates, Mac Iver et al. (2003) examined the records of individual incoming ninth graders over much of the decade and calculated BCPSS graduation rates by following first-time ninth graders in the system over subsequent years. The 4-year graduation rates of these students are also reported in Figure 1. Although Mac Iver et al. found 4-year graduation rates that were consistently 3–6 percentile points below the MSDE-reported figures, a more important finding was that their more rigorous analyses produced an upward trend line that was essentially identical to that of the MSDE, with a 12.9% gain in graduation rate over a 5-year period.¹⁷

By either calculation, the BCPSS rise in graduation rates during Accountability Phase 2 was significant. The 28% reduction in the Maryland-BCPSS graduation rate gap over the 6 years in question must be regarded as substantial progress. In practical terms, the gain meant that nearly a thousand additional young people per year were leaving school more nearly ready to either begin college or compete for relatively well-paying employment. According to Sharon Lewis (personal communication, December 2003), director of research for the Council of Great City Schools, the BCPSS rise in graduation rates during those years had few if any parallels in the urban United States. Data from Maryland's performance measures, Baltimore's use of a norm-referenced test, and two calculations of high school graduation rates all pointed to significant progress during Accountability Phase 2's period of substantially increased systemic supports.

Findings From an Independent Evaluation

Senate Bill 795 had mandated that an independent evaluation of Baltimore's reforms be conducted toward the end of the initial 5-year cycle. Westat won the competitively bid contract to conduct the evaluation. The Westat (2001) report can be summarized as revealing that the system had made a great deal of progress but still had a great distance to travel to achieve state-average progress. The authors of the report found that academic achievement gains had been substantial and that, in general, management of the system had improved. They included two overarching suggestions for moving forward: (a) involve stakeholders outside the central office more fully in decision making and (b) improve and enhance the information available for guiding reform and assessing its results.¹⁸

The Westat report was widely perceived to be supportive of the 1997 systemic reforms and of improvements in student outcomes, yet it provided valuable recommendations for additional improvements. In part on the basis of that report, and in part as a result of the positive trends in student outcome

data just described, Maryland's legislature voted to continue the Senate Bill 795 reforms for an additional 5 years.

Reforms Under Senate Bill 795: One School's Contrasting Case

Along with the district's overall rise, the majority of individual schools were making academic progress in the years between 1997 and 2002. However, the picture was not universally positive. The state's MSPAP program was intended to focus on the school as the unit of accountability. At the school level, Accountability Phase 2 brought three additional problems to light.

The first of these problems concerned the limits of the MSPAP tests themselves. The MSPAP tests were designed to be performance assessments and to be worthy of being "taught to." As an example of one of the test measures, teams of students might be asked to place beans in varying conditions of water and soil and make observations over a week as to the progress of each toward sprouting and healthy growth. At the end of the experiment, each student would receive a writing task, with one focusing on the scientific method, another on the actual condition of the beans at week's end, and so on. While the MSPAP had a certain elegance and many admirers (e.g., Herman, 2003), scoring proved problematic on several fronts. One difficulty particularly relevant in the current context was that the scores themselves had to be aggregated to the school level to become moderately reliable. Even at the school level, relatively stable schools often saw their scores "bounce" up and down by several percentage points.

This bouncing was compounded by the fact that individual cohorts coming through schools are very rarely uniform. With the combination of test-generated error variance and between-cohort variance, it was inevitable that schools would produce a certain amount of "noise" along with their MSPAP "signal." As a practical matter, a school was declared RE if two conditions were met. First, the school's MSPAP scores had to be well below the state average. Second, the school's SPI score had to drop below the average of the 2 previous years for 2 consecutive years.¹⁹ While these criteria were rational on their surface, the practical implication was that, over time, the majority of historically low-scoring schools would migrate to the RE list. The reason was that no allowance was made for possible measurement error or differences between cohorts.

As an example case, SPI actual scores and annual changes between 1993 and 1998 for Baltimore City's School #50 are presented in the top half of Table 5. School #50's SPI may well have reflected a flat academic trajectory, with an average SPI of approximately 18.8. However, because 1994 produced an unusually high score, the average of the SPIs for 1994 and 1995 was 21.1. Although the school's SPI for 1996 was up (however marginally) from 1995, it was down from the average of the 2 previous years, with the result that 1996 became the first of 2 consecutive "down" years. School #50's SIP for 1997 was down nearly 3 points from 1996 and was also below the average of the 2 previous years. Hence, the school was declared RE. The

Table 5
School Performance Index (SPI) Values for BCPSS School #50:
1993–1998

Year	SPI	1-year change	Change from mean of 2 previous years
1993	17.8		
1994	23.8	6.0	
1995	18.3	-5.5	-2.5
1996	18.7	0.4	-2.4
1997	15.8	-2.9	-2.7
1998	18.7	2.9	1.5
Results with 1994 and 1997 scores reversed			
1993	17.8		
1994	15.8	-2.0	
1995	18.3	2.5	1.5
1996	18.7	0.4	1.7
1997	23.8	5.1	5.3
1998	18.7	-5.1	-2.6

Note. BCPSS = Baltimore City Public School System.

irony in regard to School #50 was that all of these years' SIP scores were within one standard error of the mean for the years 1993–1998, so simple error variance was a possible explanation for the differences over time. Had the scores for 1994 and 1997 been reversed (as demonstrated in the bottom half of Table 5), the school would not have been declared RE in 1997. Other BCPSS schools had lower average scores; however, because their scores had not reflected the particular pattern of 2 consecutive declining years, these other schools were never declared RE.

Just as there had been schools in Baltimore with lower average SIP scores but without a 2-year downward trend, there were schools in other parts of Maryland that served much more affluent communities and had as many as 5 consecutive years of declining SPIs but scores above the state cut-off. These schools were not considered for RE status.²⁰ In ways that foreshadowed NCLB adequate yearly progress (AYP) issues, several schools were targeted under the MSPP for special attention on the basis of what may have been error variance or other arbitrary components of the rules.

The second school-level problem with Accountability Phase 2 was that for several years there were no published RE exit criteria, and once criteria were published they were so far above the scores of all RE schools that only two schools statewide (one in the BCPSS) were able to reach those levels during the few years of implementation of the criteria. During the no-specified-criteria years, schools were not motivated; during later years, the majority of local educators saw the exit criteria as unattainable. RE came to be perceived by some BCPSS educators as essentially an arbitrary, permanent designation.

Third, RE schools that did not make progress for any 2 consecutive years, as measured by the SIP, were to be "reconstituted." This produced a range of situations that hovered between unfortunate and seemingly farcical. For example, Maryland's first two schools to be declared RE had been BCPSS neighborhood high schools. Because these schools had made negligible progress over the subsequent several years without a threatened state takeover, it became the opinion of the BCPSS school board and the majority of local educators that the state was unable to identify any plausible alternative organizational structure or credible, research-based, high school management firm and would never "reconstitute" the schools. Whether or not these opinions were completely accurate, no Maryland RE high school has been reconstituted.

Several elementary schools produced a different scenario. Three BCPSS elementary schools were declared to have not made adequate progress under MSPAP and RE status, and in 2000 the state awarded a contract to manage all three to a nationally recognized for-profit corporation. The following year, two of the schools reported essentially no progress, and the third reported gains that were so large as to strain credulity. Three years later, in the summer of 2003, none of the three schools had made sufficient progress to exit the need for improvement designation. Under the new NCLB rules, one of the remedies available to the state to improve student outcomes in such circumstances would be to turn the three managed-for-profit schools over to a management-for-profit corporation to achieve better results. Accountability-based school reform is not without paradox.

Summarizing the effects of Accountability Phase 2, the governor and legislature made good on their pledge of additional funding; the state department worked to find useful support mechanisms; the BCPSS board worked with a series of three CEOs to build workable academic supports for students ranging from new texts and materials to summer school, teachers, and schools; and students were increasingly held accountable for their progress in specified content areas. The BCPSS also made substantial, educationally significant gains. These gains were spread from first-grade test scores to high school graduation rates and stood in clear contrast to the lack of gains associated with the Phase 1 testing-only reform efforts. Phase 2 was not without paradoxes, but, as both Linn (2003) and Herman (2003) would predict, its emphasis on providing support at multiple levels produced educationally important, measured gains among Baltimore's often highly at-risk students.

Accountability Phase 3: Initiation of Federally Legislated NCLB Reform

Implementation of the third wave of accountability-driven reform began with initiatives implemented by the MSDE to meet the requirements of the NCLB federal legislation. Following NCLB mandates, the MSDE reviewed content standards, funded the development of new assessment and accountability programs, and modified its data collection and reporting processes. Compliance efforts made by the MSDE have included adoption of grade-specific

content standards by the Maryland Board of Education and a voluntary Maryland curriculum aligned with these standards. The department has eliminated the MSPAP, which had been aligned with the then-current Maryland Learner Outcomes.

With the spring 2003 release of results from the MSA, the state's accountability focus turned to examining the percentages of students scoring at three performance levels: *basic*, *proficient*, and *advanced*. Results are used as the basis for determining the NCLB's AYP requirements. More specifically, these criterion-referenced scores are used in the newly implemented Maryland accountability program that incorporates "annual measurable objectives" targets to determine whether a school meets the AYP requirements. Maryland's newest accountability program, designed to be in alignment with this federal legislation, requires schools and school systems to meet targets or be within the 95% confidence band of meeting the criteria for aggregate grade-span achievement performance and for subgroups having five or more students. This was an improvement on the MSPAP, which had made no allowance for measurement error. Finally, whereas the majority of states have chosen to raise accountability standards in a set of steps, thereby having the bulk of improvement requirements coming in the subsequent several years, the MSDE chose a continuous equidistance, stepwise improvement pattern in which every school must show measured improvement every year, beginning immediately.

As can be seen in Table 6, summer 2003 results from the first-time administration of the MSA at the elementary grade levels (i.e., Grades 3 and 5) revealed that, in both of the content areas tested (reading and mathematics), more than half of the BCPSS students at each grade level scored in the basic category. In comparison with their peers statewide, nearly 20% more

Table 6
Percentages of Baltimore City and Maryland Students Scoring at the Proficient or Advanced Level on the 2003 MSA

Grade and content area	BCPSS (%)	Maryland (%)
Grade 3		
Reading	39.1	58.1
Mathematics	41.9	65.0
Grade 5		
Reading	44.4	65.7
Mathematics	31.3	55.0
Grade 8		
Reading	32.8	59.9
Mathematics	11.5	39.7

Note. MSA = Maryland School Assessment; BCPSS = Baltimore City Public School System.
*Sum of percentages of students scoring at the proficient and advanced levels.

BCPSS students were at reading performance levels in the basic category in Grades 3 and 5, and this was true for Grade 3 mathematics as well.

More than two thirds of BCPSS middle-grade-level students demonstrated performance at the basic level in reading and mathematics (67.2% and 88.5%, respectively). Statewide, fewer youngsters scored in the basic category (40.1% and 60.4%, respectively). Analyses by school type revealed that more students in the traditional Grades 6–8 middle schools in the BCPSS scored in the basic category than students in BCPSS schools configured as prekindergarten through eighth grade. This is particularly pertinent given the decision of the board to implement more schools with the latter configuration. While these results reflect all students tested, an analysis of the subgroup of students assessed for federal/state accountability purposes (and combined with data from the alternate versions of the MSA that were developed for special education students) in terms of AYP requirements revealed that the BCPSS did not meet the designated targets for all students. However, all 23 of the other jurisdictions in Maryland also failed to meet these requirements in the first year of this new accountability program in the case of one or more of the subgroups of students on which data had to be disaggregated.

Furthermore, the MSA results confirm that BCPSS students and schools, though still ranking lower than the state overall, are no longer extreme outliers as they were in the early years of the MSPAP. First-year MSA data verified the need for continuing improvement in the BCPSS, but these data also showed the progress made by the system since the beginning of the 1997 reforms.

School-level analyses revealed wide variations. In some schools, more than 90% of students scored at the proficient level in at least one content or grade level. On the basis of the new tests, eight BCPSS schools were removed from what was previously labeled RE. These schools were no longer deemed RE because, overall and for each student subgroup, they met AYP requirements and had a positive CI on the former state measure, the MSPAP, in the previous year. This is particularly noteworthy given that, during the decade of Phases 1 and 2 of the reforms, only two BCPSS schools had exited RE status.

Hence, one positive note for the BCPSS in this phase is that it is now clear how a school can exit RE status, as the targets for each year through 2013–2014 have been established. At least in the early years, the targets are more imaginably reachable for many of the RE schools in the BCPSS than were the Phase 2 accountability criteria.

It is too early to judge the full impact of NCLB, which for Baltimore represents a third phase of accountability-based reforms. It is obvious that in Phase 3, as with previous accountability-led reform efforts, NCLB accountability efforts are already replete with paradox. The first of two ready examples was previously noted: NCLB suggests the introduction of a for-profit management company as a solution to low achievement. However, what should a responsible district or state do when, for 3 years, a for-profit management company managed three schools to their current (low) condition?

Second, while the BCPSS is clearly one of the fastest improving LEAs in the state, one section of Maryland's accountability plan for NCLB states that when 25% of the schools in an LEA are deemed in need of corrective action, the district itself is declared to be in "corrective action." That status would allow the state department to completely disband the school board and central administration and begin anew. However, this is precisely what the state did 6 years earlier, and through the spring of 2004 it was the MSDE's position that the efforts to date have been largely successful. At either the school or district level, when the new remedy is the old remedy, it is difficult to imagine that it will result in a greatly increased rate of improvement. NCLB has no stated classification for "stay the course as systemic improvement continues producing positive effects."

Discussion

Baltimore has experienced three phases of accountability-driven reform in the past decade. The implicit theory of action in Accountability Phase 1 was that the simple and relatively modestly priced introduction of a demanding new testing regimen would drive schools and districts toward meaningful improvements. This effort had virtually no positive effect on the very needy district or the individual schools therein. The system lacked both the human and monetary capital necessary to mount student-achievement-bearing educational reforms (Orr, 1999). However, statewide testing did throw credible new light on the levels of need in the BCPSS. Abject test scores and sustained lack of progress in improving these scores were repeatedly cited in both court cases and advocacy before the state legislature as reasons to increase funding for the BCPSS.

In 1997, the political compromise resulting in Senate Bill 795 set the foundation for Accountability Phase 2, in which increased funding was paired with increased state oversight of and support for targeted changes in the governance structure, business practices, and academic focus of the BCPSS. Senate Bill 795 initiated a broad-based effort to increase the human capital of the system and to engage parents, teachers, principals, central administrators, the school board, and the state department of education in a concerted effort to improve student achievement.

By a variety of measures, the shared-engagement/focused-accountability effort of Phase 2 was successful. The BCPSS reversed more than a decade of educational frustration and lack of success (Bowler, 1997; Orr, 1999). Longitudinal analyses conducted by multiple authors and research teams, including data presented in this article, indicate substantial gains in students' academic achievement across grades as well as significant gains in high school graduation rates. Analyses by Cibulka (2003a, 2003b) and Westat (2001; see also Frechtling, 2003a, 2003b) indicate that these gains are the direct result of focused efforts supported by both the state department of education and the local school district, especially the new board of school commissioners. Data from Baltimore's Accountability Phase 2, contrasted with the results of Phase 1, add

proof of the value of the types of multiparty accountability efforts advocated by Linn (2003) and Herman (2003).

The accountability changes produced by NCLB form a third, still unfolding phase of accountability-based reforms in the BCPSS. Were Baltimore and Maryland not building on Phase 2 reforms, NCLB, by introducing another accountability program but providing no inherent addition to human capital and only limited new funds, would probably meet the same failing fate of Phase 1. However, the majority of Baltimore's Phase 2 supports have remained in place, and the ultimate success of Phase 3 awaits future data.

Cuban and Usdan (2003) concluded that many of the urban school reform efforts of the past decade have been potentially powerful but have had "shallow roots." It is possible that, through the marriage of accountability and supported reform, the BCPSS is sending down deeper roots. The system has yet to reach the end of its accountability-driven reform journey. However, at least at the system level, after years of stagnation and worse, the BCPSS has documented 6 years of clear academic progress. Baltimore's decade of reform efforts, as discussed in this article, lead us to the several conclusions.

1. First, increasingly, educational assessment is not without paradox. One of the BCPSS examples came when Maryland's more affluent, politically conservative suburban jurisdictions found themselves virtually forced to direct some of their taxes to support a predominantly high-poverty urban system's reform efforts. A second example has come as the MSDE faces the potential dilemma of having to reconstitute schools that it had previously reconstituted and assigned to a for-profit management agent.
2. Second, over time, the combination of mandated assessments, additional funding, and mandated improvements in governance that result in a shared focus on and support for academics can have substantial positive effects on measures of students' achievement. The shared accountability of Phase 2, beginning with new legislation, new governance structures, new partnerships, and a substantial infusion of new funds, produced positive, sometimes dramatic results from first grade through high school graduation. The state-city partnership has produced measured gains worthy of being replicated and studied elsewhere.
3. Finally, assessment programs can have equally documentable negative effects, such as punishing schools for small variations in year-to-year test scores. Furthermore, a mandated assessment/accountability program lacking provision of substantial, additional ongoing investment in human capital is unlikely to produce documented positive effects on student achievement. In Maryland, this was particularly true when annual changes were clearly within the confidence intervals of the measures, and no allowance was made for the possibility of differences being nothing more than error variance. When faculties perceive measures to be unfair or unreliable, a plausible school-level

product of the accountability effort is bitter, unproductive cynicism. At the school level, there were examples of the Phase 1 accountability efforts producing passivity (Mintrop, 2003) and a learned helplessness response when test scores were not paired with capacity building.

4. One of the cautionary lessons from Baltimore that can be applied to other cities and states would be to avoid overreactions to what may well be hyper-rationalized analyses of small differences in moderately reliable measures over time. Phase 1 testing alone produced little evidence of improvement in any student outcomes. However, publication of test results did have one longer term outcome in Maryland. The reforms of Accountability Phase 2 would have not been funded, and later progress associated with that increased funding would not have occurred, without dismal MSPAP scores and the increasing state-BCPSS test-score gap.

Educational reform in a context of high concentrations of poverty and historically disadvantaged groups is a very long process. The post-1997 progress of BCPSS is among the more promising urban reform stories in the country. For example, in a nation where high school graduation rates are stable or dropping (Greene & Forster, 2003), a 13% gain in graduation rates over a 5-year period compares very favorably with the gains in deemed-to-be-exemplary urban LEAs and is laudable. Yet, if the laudable BCPSS graduation trend were to continue in a linear fashion, the high school graduation rate would match the rising state rate in the summer of 2015. It is not clear that Baltimore's 5-year trend can continue for a decade and a half.²¹ But even if such a trend is possible, it is equally unclear that the necessarily long series of elected and other government officials will choose to stay the course for that long.

One of the greatest challenges along the path to dramatic academic improvement in the BCPSS has been attempting to respond to a long series of reform for social capital. To the extent that there has been success in Baltimore's reform efforts of a decade and counting, it has built on the multi-faceted professional development of current staff and the hiring of increasingly highly qualified administrators and teachers. The potential of the BCPSS to meet the rising accountability challenges of NCLB and any other challenges it sets for itself will be directly related to its ability to continue recruiting excellent new staff and upgrading current administrators' and teachers' skills. As a nation, the United States needs to invest in the long-term development of a new generation of urban educational leaders with an unprecedented level of skills.

Another great challenge has involved the system's ability—and often inability—to convert the large quantities of data it houses to useful, timely information. In the modern information society, the ability to access data and convert diverse data sets into usable information becomes paramount in regard to school improvement and data-driven reform. The BCPSS has invested millions of dollars in computer systems to facilitate modern financial and aca-

demical accounting. In many ways, these systems have yielded positive results. Yet, far too often, important budgeting and programmatic decisions have been made in environments that are virtually awash in figures but in which decision makers perceive themselves to be "data rich and information poor." In some areas, this has spoken to the need for ongoing professional development both in learning to "ask the right questions" of data and in interpreting data. In others, it has demonstrated the system's inability to date to build adequate linkages among diverse data sets and to provide more user-friendly interfaces between these data sets and central administrators and teachers.

Throughout the phases of reform, there have been examples of principals and teachers who are generally competent but who lack adequate sophistication in interpreting the increasing numbers of often psychometrically sophisticated measures for which they were being held accountable. One of the clearest areas of future capacity building in the BCPSS—and, we suspect, the nation—involves the development and procurement of school-level analytic instruments combined with professional development in making sophisticated interpretations and practical use of analyses.²²

A great deal of research remains to be done on Baltimore's reform efforts, and exponentially larger amounts are needed on urban educational reform nationally. In the case of the BCPSS, since the Phase 2 reforms began 7 years ago, the state has elected a new governor who is ideologically quite different from his predecessor, the city has elected a new mayor, and seven of the nine original members of the new school board have rotated off the board. The system has had one acting and three permanent CEOs; four chief financial officers, chief academic officers, and directors of human resources and information technology; three chief operating officers; and three directors of research and evaluation. In addition, there have been many other senior management changes, as well as a nearly 50% reduction in the overall size of the central administration. Through all of this, the stabilizing force had been the school board itself (Westat, 2001); in the past year, however, several of the more senior members of the board have completed their legal maximum of two full terms and have rotated off the board. This would appear to require a new source of stability in the system. Can the BCPSS—or any system—sustain a vision, direction, and capacity in the face of so many changes? To what extent will the skills that brought the system up from near chaos need to be modified to move the system to the full goals of NCLB and the restored dreams of Baltimore's citizens? What reforms will prove most effective in moving forward, and how can they be best separated and measured? Only future, long-term research can address these questions.

Nationally, what is the range of reforms that can be productive in urban districts? Reform in the BCPSS has been much more pragmatic than theoretical, but one would presume that other reasonably successful districts would have co-constructed (Datnow & Stringfield, 2000) other choices that were different yet equally pragmatic within their contexts. To what extent can one generalize from these situations? Do accountability issues play as large an array of positive