High-Stakes Accountability in Urban Elementary Schools: Challenging or Reproducing Inequality?[1]

JOHN B. DIAMOND
University of Wisconsin–Milwaukee

JAMES P. SPILLANE
Northwestern University

In this article, the authors use data from interviews and observations in four urban elementary schools—two high-performing and two probation schools—to examine how schools respond to high-stakes accountability policies. The authors show that school responses to high-stakes accountability depend on the schools' accountability status. In probation schools, responses focus narrowly on complying with policy demands, focusing on improving the performance of certain students, within benchmark grades, and in certain subject areas. In contrast, higher performing schools emphasize enhancing the performance of all students regardless of grade level and across all subject areas. Given the concentration of poor students and students of color in the lowest performing schools, the authors conclude that issues of educational equity need to be given greater consideration in the implementation of high stakes accountability policies.

One of the most consistent findings in educational research is that family background is linked to children's educational outcomes, attainment, and adult occupational status (Blau & Duncan, 1967; Coleman et al., 1966; Jencks et al., 1972; Orfield, 1993). While education is viewed by many as an important mechanism for social mobility, many scholars argue that schools reproduce rather than challenge social inequality. Over the past decade, policy makers have mobilized multiple policy instruments in an effort to ensure that all children receive high quality education. One increasingly popular but controversial strategy relies on external accountability mechanisms, including high-stakes testing, to transform instructional practices and make teachers and students more accountable for their performance.

Critics argue that these policies will exacerbate inequalities by leading teachers to marginalize low-performing students (Clotfelter & Ladd, 1996;

Teachers College Record Volume 106, Number 6, June 2004, pp. 1145–1176

0161-4681

McDill, Natriello, & Pallas, 1986) and by causing teachers to teach these students only the material covered on standardized tests. Others contend that such policies are misguided because limited resources, unprepared teachers, and ineffective instructional practices rather than incentives are the problems that need to be addressed (Darling-Hammond, 1994).

Proponents argue that such assessments will reduce gate-keeping processes such as tracking and low teacher expectations that disadvantage certain students. External assessments, it is argued, provide objective information for school-based decision making and therefore work against more subjective judgments that contribute to stratification (Coleman, 1997; Muller & Schiller, 2000). For example, supporters argue that teachers' assessments of students' ability as well as decisions about course placement and grouping arrangements inside classrooms could be based on more objective information from standardized tests (Muller & Schiller, 2000).

Given the increasing emphasis on external assessment and accountability, and the strong arguments on both sides of the issue, it is interesting that with a few notable exceptions (Muller & Schiller, 2000; Roderick, Bryk, Jacob, Easton, & Allensworth, 1999) research on how these policies unfold in schools and the mechanisms through which they impact student learning has been limited. Moreover, few studies have explored how the implementation of these policies may be influenced by external pressures associated with high-stakes accountability.

The study reported on in this article examines the implementation of high-stakes testing and accountability in high- and low-performing schools. We seek answers to two related questions. First, how is high-stakes accountability policy perceived and implemented differently in high- and low-performing schools? Second, does the implementation of high-stakes testing in these schools suggest that it will reduce social stratification through the mechanisms outlined by testing proponents?

HIGH-STAKES ACCOUNTABILITY AND STRATIFICATION

One aim of accountability policies is to ensure that all students receive high-quality instruction and reach a certain level of competence in core subject areas (Muller & Schiller, 2000). Some districts, like Chicago, have adopted a high-stakes version of these policies that link student performance on examinations to consequences for schools and, in some instances, students themselves. Opponents of these policies argue that for these approaches to be fair, instructional changes should precede consequences for students (Heubert & Hauser, 1999) and that such policies create incentives for marginalizing low-performing students (Clotfelter and Ladd, 1996; McDill, Natriello, & Pallas, 1986). Neutral observers caution that those implement-

ing such policies must ensure the adequacy of educational resources for the tested students, attend to the reliability and validity of the exams for their intended purposes, and avoid basing decisions on one test (American Educational Research Association, 2000).

Proponents of these policies suggest that they can reduce inequality through increasing student motivation, creating incentives for teachers to seek improvements in student outcomes, providing more objective information about students' performance for school-based decision making, and increasing academic press in schools—particularly those serving low-income and minority students (Coleman et al., 1997; Muller & Schiller, 2000; Shouse, 1997).

Some research advocates the development of output-driven schools, a key component of which would be external assessment and accountability. Coleman et al. (1997) argues that external assessments such as student performance tests would create new incentives for school improvement, providing objective information for teachers to assess students and make course placement decisions, thereby reducing the gate-keeping functions of schools (Coleman et al., 1997; Muller & Schiller, 2000). Shouse (1997) argues that designing more output-driven schools would also increase academic press and have particularly beneficial consequences for students with lower socioeconomic status. Therefore, proponents of these policies suggest that three key mechanisms—the creation of new incentives, the provision of objective information for school decision making, and the increase of academic press—will combine to reduce schools' and teachers' gate-keeping practices and contribute to a reduction in stratification.

While arguments have been forwarded in support of and in opposition to these policies, surprisingly little research closely examines how they play out in schools. The data on outcomes that does exist presents a mixed picture. Data on the implementation of high-stakes testing in Chicago suggest that the percentage of students meeting minimum competency requirements has increased since the introduction of the policy (Roderick, Bryk, Jacob, et al., 1999). However, the policy has differential impacts on students based on their family background characteristics. For example, African American students were retained at a much higher rate than their White and Latino counterparts because they tend to score lower than Whites and because of the higher proportion of Latino students who are in bilingual programs and therefore exempt from the policy (Roderick et al., 1999). This likely has a stratifying effect for African American students because grade retention may be associated with negative long-term outcomes including reduction in self-esteem and increased likelihood of high school drop out (Roderick, 1994).

Muller and Schiller (2000) examine how state-level testing policy impacts high school students' graduation rates and mathematics course taking.

They show that these policies equalized students' academic attainment and reduced the impact of teachers' gate keeping through low-expectations, seeming to support the arguments of supporters of these policies (Muller & Schiller, 2000, p. 210). However, they also find that when state tests link students' performance to consequences for schools it leads to stratification based on SES, lending support to the arguments of opponents of testing policies. Therefore, their findings do not strongly support the arguments of proponents of these policies or their critics. They recommend more research using both qualitative and quantitative methods to explore the mechanisms through which these policies influence teachers' practices and students' outcomes. Research into the processes of accountability policy implementation will help inform this discussion.

In this article, we argue that to understand the implications of these policies it is important to examine how they are understood and implemented in particular school contexts. We examine schools' responses to high-stakes accountability policy, paying particular attention to the implications of these responses for issues of educational equity both within and across institutions. Specifically, we examine how teachers and administrators in high- and low-performing schools respond to high-stakes accountability policy focusing on their responses to incentive structures, their interpretation and use of test score data, and their subsequent instructional priorities. We argue that these responses to high-stakes accountability are situated in a school's status with regard to accountability policy—probation versus high performance—and argue that school status is correlated with students' race and social class. We conclude that differences in responses to accountability policy in different types of schools may increase rather than reduce educational stratification.

The article is organized as follows. First, we outline the theoretical tools used to frame our discussion. Second, we briefly discuss the Chicago context, describing the high-stakes accountability policy in that district. Third, we discuss the methodological approach that guided the research. We then examine differences in schools' responses to high-stakes accountability in four elementary schools—two high-performing schools[2] and two probation schools. In the final section, we discuss these cases, paying particular attention to the implications of school-level responses to high stakes testing for issues of stratification.

THEORETICAL TOOLS

Research on the role of family background and educational stratification demonstrates consistent links between socioeconomic status and students' outcomes. Some explanations for this pattern focus on direct effects of

family background such as class-based disparities in parents' beliefs and involvement patterns (Lareau, 1989; Sewell & Hauser, 1980; Sewell & Shah, 1968a, 1968b), family structure (i.e., number of parents in the home or the number of siblings in the family), and access to extrafamilial resources through parents' social networks and institutional affiliations (Carbonaro, 1998; Coleman, 1988; Coleman & Hoffer, 1987; Hao & Bonstead-Bruns, 1998; Hofferth et al., 1998; McNeal, 1999; Wilson, 1987; Wong, 1998). Other scholars focus on the interaction between these background characteristics and school practices. In these accounts, schools affect students though micropolitical processes such as low teacher expectations (Brophy & Good, 1974; Rist, 1970; Roscigno, 1998; Rosenthal & Jacobson, 1968), tracking (Oakes, 1995), and cultural reproduction processes (Bourdieu, 1979; Bourdieu & Passeron, 1990).

A third approach—institutional stratification—highlights the implications of interorganizational processes for social stratification (Roscigno, 2000). Because race and social class shape school attendance patterns and contribute to the creation of highly segregated school contexts (Orfield, Bachmeier, James, & Eitle, 1997), family background can contribute to stratification through the distinctly different characteristics of the schools students attend (Roscigno, 2000). Differences in schools' monetary resources (Elliott, 1998; Hedges & Greenwald, 1996), instructional quality (Smith, Lee, & Newman, 2001), presentation of valued knowledge (Anyon, 1981), course offerings (Ayalon, 1994), and social organization (Bowles & Gintis, 1976), all contribute to the maintenance of social stratification.

Roscigno's (2000) institutional stratification perspective captures these interorganizational dynamics, emphasizing the multilevel and interinstitutional nature of racial educational disadvantage. Roscigno writes:

> Arguably, the most important of these inter-institutional linkages in relation to race/class reproduction in education has to do with family background inequalities and their consequences for achievement *through* the character and resources of the schools one attends. What this means, more straightforwardly, is that family background shapes residential options. Where one resides, in turn, has a large impact on the school one attends and, consequently, achievement [emphasis in original]. (p. 271)

He concludes that residential segregation leads to indirect effects of family background through enrollment patterns in public versus private schools, the race and social class composition of schools, monetary expenditures, and school climate (Roscigno, 2000). Other work demonstrates that the concentration of low-income and African American students in certain schools may have detrimental implications for student outcomes apart from

the individual characteristics of students (Bankston & Caldas, 1996). Taken together, this work demonstrates that educational stratification based on race and social class is at least partially maintained through race and class-linked institutional processes.

Building on this work, we argue that the implications of policies like high-stakes accountability are also shaped by institutional stratification processes. In contemporary urban contexts there are several types of schools, including private schools (both religious and nonreligious), which are often thought to be the highest quality, magnet schools, which are often considered the "elite" public schools, and neighborhood schools, which can be further divided into high and low quality categories. Social class and race are important in patterning the schools that children attend, with the more highly valued settings being most accessible to middle-class and upper class children. The different types of public schools are likely to implement the policy differently. Therefore, if students are concentrated in different types of schools based on race and social class, they will be impacted by the policy in distinct ways.

THE CHICAGO CONTEXT[3]

The Chicago Public Schools (CPS) provide fertile ground for the examination of high-stakes testing and the distribution of students across high- and low-performing schools. Recent reforms press for schools to be held accountable for student achievement. Arguing for more intellectually rigorous content, reformers propose using specific student performance outcomes and rewards and sanctions for schools to hold schools accountable for students' mastery of this content (Clotfelter & Ladd, 1996). We identify each of these components below and use the Chicago Public School's accountability policy, an often referred to example of a successful accountability policy in both policy and academic circles, to illuminate each component.

To begin with, student performance outcomes as measured by tests, rather than inputs (e.g., number of certified staff), are the primary mechanism that state and local government agencies use to hold school's accountable. While the 1988 Chicago School Reform Act (P.A. 85-1418) decentralized decision making and led to the formation of Local School Councils (LSC), the 1995 Chicago School Reform Amendatory Act gave significant power to a mayoral appointee, the chief executive officer, who was empowered to place poorly performing schools (e.g., schools with poor performance on the Iowa Test of Basic Skills [ITBS]) in remediation or on probation. Specifically, students' scores on the ITBS at benchmark grades became the districts' most important accountability measure.

The second component of most accountability policies is the creation of a system of rewards, sanctions, and interventions as motivators of change. The key sanction in Chicago was the CEO's ability to place schools on probation as a result of low student test scores. The CEO did not hesitate to use this authority. In 1996 25% of the districts elementary schools (109 schools) in which 15% or fewer students performed at or above national norms in reading and mathematics on the ITBS were placed on probation (Hess 2000; Wong & Anagnostopoulos 1998). Schools on probation must develop supplemental school improvement plans that detail school strategies to improve student achievement and create mechanisms for tracking progress. Schools may also use their discretionary funds to purchase the services of district approved external partners who provide technical assistance to probation schools. If these interventions do not lead to adequate progress, probation schools can be reconstituted by the CEO. In such cases, the school receives a new principals, faculty, and Local School Council.

Some high-stakes accountability policies also include rewards or sanctions for students. This is important because, as some scholars note, teaching is coproduced by teachers and students (Cohen & Ball 1998). Hence, an accountability system that targets teachers and school administrators exclusively may place them in an impossible position as they depend on their students to improve school performance. In 1996, the CPS ended social promotion by forcing students who performed poorly on the ITBS to attend school during the summer. Moreover, if students failed to achieve the required level on the ITBS after summer school, they are forced to repeat the same grade. These developments are important because the incentive structure mobilized by the school district targets students in addition to teachers and administrators.

While the accountability policy is designed to impact the outcomes of all students, students are not evenly distributed across these schools. CPS data[4] shows that schools on academic probation have a higher percentage of African American and low-income students, on average, than the typical Chicago Public School. African Americans make up 52% of the district's student population but 83% of students attending probation schools. Likewise the district average of low-income students is 84%, while the average for probation schools has 92% low-income students. Perhaps the most interesting fact, however, is that while White students make up 10% of the district student population, they make up less than 1% of the students attending elementary schools on probation.

These figures are more striking when compared to data from Chicago magnet schools, which consistently rank among the district's highest performing schools. Elementary magnet schools contain 55% low-income students (compared to 92% for probation schools) and 53% Black students

Table 1. Mean percentage of students in CPS probation, magnet and high-performing schools by race and social class

	District	Probation	Magnet	High Performing
% African American	52%	83%	53%	27%
% White	10%	.12%	17%	34%
% Low Income	84%	92%	55%	56%

(compared to 83% for probation schools). White students make up 17% of elementary magnet school students (as compared to less than 1% in probation schools). These data demonstrate that structural processes related to family background shape students' access to schools of different quality. African American students from low-income families are more likely to be found in the lowest performing schools while white and middle income students are more likely to be found in higher performing magnet schools.[5] Table 1 presents the mean percentage of students by race (Black/White) and social class in the Chicago district as a whole as well as its magnet, high-performing, and probation elementary schools.

The four schools we studied do not conform to the district's race and class attendance patterns with regard to student composition and performance levels. However, we argue that our findings on school responses to accountability policy have theoretical significance given that low-income African American students are overrepresented in probation schools.

The multiple factors that contribute to this process are beyond the scope of this paper, however, as we shall see, the observed patterns likely have important consequences for students' educational experiences.[6] In the discussion that follows, we focus on school status in relation to high-stakes accountability (probation vs. high performance) as an important factor that shapes schools' responses to accountability policy and students' access to educational equity.[7]

RESEARCH METHODOLOGY

This article is based on data from the Distributed Leadership Project, a 4-year longitudinal study of elementary school leadership funded by the National Science Foundation and the Spencer Foundation. In this larger study we were chiefly interested in schools that had shown signs of improving mathematics, science, or literacy instruction, we also wanted to study some schools that had managed little change in student outcomes. We

Table 2. School demographics and test scores

School Name	Racial Composition	% Low Income	ITBS Mathematics Composite K–8	ITBS Reading Composite K–8
Blake(High Performing)	40% White 6% Black 26% Hispanic 26% Asian	69%	63.7	71.1
Kelly (High Performing)	100% African American	85%	65.3	70.7
Field (Probation)	100% African American	99%	16.4	20.0
Wexler (Probation)	100% African American	97%	24.7	22.0

therefore selected two schools that were on district-mandated probation. The study reported here is based on the examination of four schools, two were among the highest performing in our sample and the other two were probation schools and among the lowest performing. In Table 2 we outline schools by racial composition, percentage of low-income students, probation status, and test scores on the ITBS during the 2000–2001 academic year.

DATA COLLECTION

Members of the research team spent between 50 and 70 days of fieldwork in each of the four sites discussed in this paper. The research strategies include observations and structured and semistructured interviews. Researchers observed school leadership events, school meetings, and classroom instruction. Leadership events observed in these schools to date included grade level meetings, faculty meetings, school improvement planning meetings, professional development workshops, and supervisions of teaching practice. In addition, we observed a number of other events including homeroom conversations between teachers, lunchroom conversations, grade level meetings and subject specific workshops and meetings.

Interviews were conducted with teachers (primarily at the second and fifth grade levels) and formal and informal school leaders. Interview protocols focused on school leaders' agenda and goals, their responsibilities, and the key tasks they perform as part of promoting instructional change in mathematics, science and literacy associated with accountability policy. Using the protocols, researchers wrote detailed field notes following each observation. All interviews were tape-recorded and transcribed.

DATA ANALYSIS

We integrated data collection and data analysis to examine emergent patterns and hypotheses and refine data collection strategies (Miles & Huberman, 1984). Coding categories were developed based on initial analyses of our observation and interview data. A computer based qualitative data analysis program—NUDIST—was used to code and analyze all project data. NUDIST allowed us to code the emerging ideas and concepts from the data into free nodes, parent nodes, and child nodes that could then be compared and related to each other and stored in an index system that brings the different components of the project together. Coders worked together to code transcripts initially to develop a shared understanding of what each code meant. Once coders had developed a shared understanding of these codes, they worked independently.

FINDINGS[8]

SIMILARITIES ACROSS SCHOOLS IN RESPONSES TO ACCOUNTABILITY POLICY

One of the core arguments of proponents of high-stakes testing is that it provides incentives for school improvement. In simple terms, when schools, teachers, and students are made accountable for students' outcomes, it is argued, they will seek to improve them. Data from the four schools indicate that school leaders do pay attention to high-stakes accountability. There were general patterns that cut across these schools, some of which were in line with proponents arguments in support of these policies. Leaders at each school report paying some attention to the exams. For example Dr. Smith, the principal at Wexler School explained, "We still are aiming for national average being at 50%+ and we feel that we can get there . . . we still have a challenge ahead of us and we are working vigorously on it." Likewise, teachers and administrators at Field School were very focused on improving test scores.

However, even in the highest performing schools where 50% or more of the students score at or above national norms in core subject areas, school leaders reported paying some attention to exam results. In addition, leaders at both high- and low-performing schools reported seeking improvements in students' outcomes. As Mr. Preston, the principal at Blake school reported:

> When I look at the test results—and I happen to be one that believes that the test results do tell you something about curriculum—fifty

> percent are succeeding. I look at it the other way, fifty percent of our children are not succeeding. . . . Hopefully our scores will go a notch up.

The push for improvement was consistent among school leaders across all of our schools. None of the school officials we interviewed was completely satisfied with student performance on exams.

Each of the schools also engage in some form of explicit test preparation activities. There are differences, however, in the strategies used and the frequency of these activities. At one school, which engages in the most extensive test preparation, students are tested every Thursday. The stated goal at this school is to make the testing environment comfortable for the students so that when they take the actual exam they will not be overwhelmed or feel undue pressure. As Ms. Brown, the assistant principal at Kelly explained:

> We expose them to the testing situation at least once a week and we use the bubble forms except for the little ones because their test is not like that. In other words we try to simulate the testing even with the writing part of the test. I've shown the teachers how to do booklets that simulate the real test. So we try to do whatever we can so that our children are accustomed to taking tests so they're test smart kids and they're not nervous.

In contrast, these test preparation activities were more sporadic at other schools, occurring increasingly often just prior to the testing period. For example, the principal at Field School, Dr. Johnson, explained that there is an effort to refresh students basic skills just prior to testing in order to improve their performance.

> Right before the test [last year] we held a 'this is how important it is' for you to do well on the test meeting. And I'm going to hold that meeting again this year. And explain to them that you are receiving flash cards, each child will receive basic skills flash cards. Third grades will have, right now they'll be getting multiplication but we're going to do addition, subtraction. I mean really, basic skills.

We also found that testing structured schools' and teachers' priorities with regard to the content covered in classrooms and the attention paid to different subject areas within these schools and in the larger study. With regard to instructional content teachers reported that testing was outmatched only by other teachers and textbooks as an influence on the content they covered.[9] One specific manifestation of this influence on

content coverage was the lack of attention to science instruction when compared to reading and mathematics, a situation that has been documented elsewhere (Spillane, Diamond, et al., 2001).

Within the field of elementary education, mathematics, science, and language arts instruction are differentially valued. Although subject matter specialization and departmental structures do not directly lead to the development of subject matter identities among teachers, subject matter remains an important context for their work (Stodolsky, 1998). A general pattern at these four schools (and reflected in the larger sample of schools) was that science instruction was given lower priority than other subject areas. Across our schools we found that when compared to mathematics and language arts instruction, science had fewer formal and informal subject matter leaders, less attention from school administrators, and fewer instructional specialists (Spillane, Diamond, et al., 2001). Further, our data suggest that accountability policy may add to this undervaluing of science education. As the following quotations indicate, teachers in both high- and low-performing schools emphasized mathematics and language arts over science at least partially because of the standardized tests and accountability policy.

> Science and social studies are more flexible [than reading and mathematics] because the students are not tested on the IOWA's [Iowa Test of Basic Skills] in science and social studies so that's more, you know, on the teacher's personal decision.
>
> You know science isn't one of your guides for whether a child is promoted or graduates. So reading and math are what are stressed because those are what everybody looks at. . . . You've got to get this in math, you've got to get this . . . in reading. So those two always come first.
>
> We aren't able to teach science as much as I would like to. Mainly because on the 3rd grade level we aren't tested on [science and social studies] we're not tested on those subjects we are tested on reading and math. . . . I just can't fit it in. . . . I begin teaching science and social studies after the test. (Spillane, Diamond, et al., 2001)

Therefore, across all of our schools, and without regard to performance level, school staff paid attention to tests results and sought to improve students' outcomes on them. According to teachers' reports, testing and accountability policy also influenced the instructional content covered and the priority placed on different subject matter areas, with science receiving less attention than language arts and math. This is important because a key goal of accountability policy is to motivate teachers to focus on instruction and improved student outcomes. These common patterns across schools

suggest that these policies, at least partially, result in their intended purpose.

SCHOOL STATUS DIFFERENCES AND RESPONSES TO ACCOUNTABILITY POLICY

Despite these general patterns across all the schools, school personnel at high- and low-performing schools apprehended and responded to the policies in distinctly different ways. Hence, the impact of policy on school practice was not uniform across schools. Comparing responses to accountability policy at high- and low-performing schools in the following section, we explore differences in schools' responses to the incentives, their use of testing data, and their instructional responses. More specifically, we argue that incentives in probation schools were focused on sanctions rather than rewards and led to a superficial response based on external threats while high-performing schools responded in ways that were closer to the intentions of policy makers.

SCHOOL STATUS AND THE NATURE OF INCENTIVES

Proponents of high-stakes accountability argue that it will create incentives for instructional improvement. We argue that how schools respond to these incentives depends at least in part on their status in relationship to accountability policy. In Chicago, the incentives for probation schools are clear and direct; they are organized around the threat of reconstitution. To avoid this schools must get off of probation and this becomes the goal for school staff. In high-performing schools, the incentives are structured around rewards more than sanctions, with recognition for high performance being the primary motivator. The following section examines the ways in which school leaders (and schools as a whole to a certain extent) at high and low performing schools orient themselves toward accountability policy and the ways in which the incentive structures shape these orientations.

Incentives in Probation Schools

School leaders at the two probation schools—Wexler and Field—emphasize getting off of probation as a major goal for their schools. As the Wexler School's principal, Dr. Smith, stated when asked about the schools goals, "the obvious goal is to get off probation! Now that's it in a nutshell." This emphasis on getting off of probation in these two schools had an important impact on their responses to accountability policy. The need to respond to the immediate pressures of probation status (e.g., pressures from external

actors, such as probation managers) led to several superficial responses that were cosmetic with regard to classroom instruction.

Dr. Smith used the accountability policy to get teachers' attention and motivate them to change. Her first year at the school, she used accountability policy as a way to legitimize her push for teachers to improve their practice. She reminded teachers that she could and would replace those who were not pulling their weight. For example, in one staff meeting she informed her staff that the probation manager asked her to keep track of underperforming teachers. She explained to the teachers that she could not ignore these requests but would give each one of them "a chance" to improve. This was a tone she consistently sounded, seeking to motivate teachers. In this sense, accountability policy (and the set of actors associated with it) serves as a threat, a way for school leaders to get the attention of teachers and push them to change their practices.

The external partner at Field School believes that teachers at this school are open to change. As she has observed at this school and at others "teachers in probation schools are more likely to be open [to change]." Therefore, at both of the probation schools accountability policy is used to motivate teachers to do the bidding of school leaders.

Another motivation strategy used by school leaders at Wexler School was to encourage teachers' effort by expressing confidence in their abilities and promising that with hard work they could get off of probation. At one staff meeting with the school's new probation manager,

> there was clapping of hands as [the probation manager] stood up to speak. She began by saying, "It is possible to get off probation" (another loud applause) and "you are going to get off probation." (another loud applause) "Sometimes," [the probation manager continued], "it's only a question of knowing what to focus on, getting children to be ready for the tests and getting the right tools." (Spillane et al., 2002)

There was a revival meeting atmosphere in this meeting, the emphasis being liberation from the bondage of probation. This revival meeting approach was evident in other meetings we observed as well.

In a certain sense this strategy dovetails well with the use of probation as a threat to motivate teachers. Being placed on probation is likely depressing. It labels the school as a failure. In addition to this if school leaders work too hard to challenge teachers without some sense of emotional support, it is likely that it will be difficult to make change. This approach may be suited for creating social support in the push for academic improvement.

At the probation schools, accountability policy, and particularly probation status, enhanced school leaders' ability to get teachers' attention. It was a powerful tool in the hands of school leaders wanting to implement instructional change. Proponents of high-stakes accountability would not doubt respond, "I told you so."

Another set of incentives that existed in probation schools but not high-performing schools was the need to impress external observers. Probation schools must have an external partner and a probation manager. These actors play a role in determining whether or not a school gets off of probation. Some probation school efforts were designed to convince external observers that their schools were doing all they could to improve students outcomes. The pressure to do this was captured by the Field School principal, Dr. Johnson, who relayed a conversation she had had from an office of accountability representative.

> When Dr. Austin came out yesterday, he's . . . from the office of accountability. He said, "You know there are schools that are in worse condition, physical condition than your school. Those schools are open because they are at 40 percentile." And he said, "We don't care if they run naked through the hallways but they're not on probation. Whatever it is that you need to do then that's what you will have to do to get your school off of probation. So there's still something wrong that has to be fixed. . . . So whatever it takes, you have to fix it."

At Field School, some of these responses emphasized superficial changes designed to impress district officials and probation managers. For example, since being placed on probation, Field School has adopted several programs in a rather haphazard way in an attempt to demonstrate improvement efforts. This approach has been unsuccessful in the eyes of one of the schools external partners, a consultant from a local university, who argued that the problems of the school result from a lack of coherence in their improvement efforts. Likewise, the following field note illustrates a similar perceived lack of coherence of the school's program.

> After sitting through a number of . . . meetings, I'm still struggling to see coherence to the schools struggle to get off probation. While the components seem to be in place . . . the quality of those components seems to be lacking.

Schools adopting several disconnected programs have been characterized as "Christmas tree" schools, a pattern that surfaced at many unsuccessful elementary schools (Bryk, Easton, Kerbow, et al., 1993). In contrast, more recent work suggests that increasing instructional program coherence is an

important mechanism for instructional improvement (Newmann, Smith, Allensworth, & Bryk, 2001).

Another strategy used in probation schools was to impress external observers through emphasizing the appearance of student and teacher engagement in instructional activity and student discipline. For example, at a Field School staff meeting the assistant principal, Dr. Janet Harold, warned teachers that the school's probation manager would be visiting and advised teachers on what actions to take.

> I don't know if you saw Ms. Fox, our probation manager, here in the building today. . . . Make sure your university organizers are visible, there should be student work, your classroom should be attractive, they are going to be looking at the decorum of our students so make sure you continue with working on good behavior, most important they are looking for students on task and that teachers are effectively teaching their students . . . there is no telling when someone might be in to visit your class.

This reflects one of Field School's responses to external pressures. The emphasis here is on classroom management, décor, and the displaying of students work. Dr. Johnson does discuss the importance of demonstrating being "on task" but focuses less on instruction issues and more on classroom management and appearance. In other words, this strategy is cosmetic, emphasizing the trappings of instructional improvement while the actual instructional practices are not emphasized.

Responses to high stakes accountability in the probation schools naturally emphasized getting off of probation. The strategies used by school leaders included using the school's probationary status to motivate teachers—both through the threat of negative consequences for low-performing teachers and through the promise of getting to the probation free "promised land" and efforts to manage the impressions of external stakeholders through adopting multiple programs and increasing the appearance of instructional emphasis and innovation.

Incentives in High-Performing Schools

Incentives in the high-performing schools are less clear. One strategy used by school leaders in these schools revolved around praise for accomplishment. In Kelly School's professional development meetings the principal regularly praised teachers for working hard to produce high student outcomes. Teachers were encouraged to applaud their own and other accomplishments and test results were prominently posted inside the school

and shared with parents who visited the school. All of this reinforced pride in past accomplishments and encouraged continued improvement.

Given past success however, leaders were also forced to combat complacency on the part of faculty. In the case of Blake, Mr. Preston felt that teachers believed that the school was performing better than it actually was. He used trend test-score data to challenge teachers' assumptions. He demonstrated that while the schools absolute outcomes were excellent, when compared to other schools in their neighborhood their students were not gaining as much year to year. As he explained,

> the analysis made clear that out of the 12 schools [in the neighborhood], Blake was either at the bottom or really close to the bottom, in terms of the amount of actual growth that students were making. Forget about where the growth started, forget about the base. Forget about the end. Just, you know, how many months of progress, on an average were sixth graders achieving.

School leaders used this data in school meetings to help create incentives for continued instructional improvement. They took great pains to repackage standardized test data in ways that captured teachers' attention, transforming massive spreadsheets into relatively easy to read charts that were color-coded by grade level (Spillane et al., 2002).

Likewise, at Kelly School, a monitoring system with a template that connects teachers daily lesson plans to the material tested, district standards, and the level of mastery of specific skills for each student was used to maintain a focus on constant improvement. Ms. Brown explained how this skill chart works:

> It's just an organizational tool. You look at this chart and you see that child didn't master that skill. That is something you can do in a small group. You can assign your [teacher's] aide to work with that particular child on that skill and retest, cause we believe right away if the child didn't master it . . . most kids only miss it like a master is 80% maybe they got a 75 or a 72, so he just missed it by a little bit, quickly review, go over it again and retest. The child masters it then move him on.

At a professional development meeting the principal, Dr. Williams emphasized that teachers need to ensure that they are tracking the skill mastery of students:

> I noticed that the skill charts are not being filled out diligently enough. The [university program] got us on the path to improvement through

> charting our progress but we can't get lax on this. We have completed 2 fifth week assessments and if you have a lot of children not getting their skills you need to reteach. If a lot of your children are not getting the material it is not the children. It is something to do with the way you taught it. You can't teach the same way every year. It's always the children. People make excuses. But that does not hold up because we can take the same child in two different classes and they can do well in one and have trouble in the other. But if you see students are having trouble don't go on. It's going to be evident that the students are not getting it so think of another way to teach it.

Here, school leaders help create a school-based incentive structure tied to the maintenance of high performance. In this example, the teachers maintain primary responsibility for student performance and excuses based on students' limitations are discounted. In another meeting (the following week) she again emphasizes the need to maintain focus and work hard in order to keep test scores up:

> Dr. Williams said that someone inquired about Kelly and asked if the school has gifted students. She said that "our students are average . . . our instructional program is what makes the difference . . . the only way we continue to improve is through hard work. Just because we did well last year [on the ITBS] does not mean anything. We have to continue to work hard and align our lesson plans.

In high-performing schools leaders used praise for past performance and the need for constant improvement to heighten teachers' sense of accountability and motivate them.

In contrast to probation schools, the incentives in high-performing schools were based more on rewards than sanctions. In these schools, school leaders praised school accomplishments in professional development meetings and proudly displayed students' outcomes in the school and communicated these to parents. In addition, because the accountability policy did not create sanctions for these schools, school leaders sought to interject other forms of motivation through comparisons with other schools or through building on the school's past performance.

The nature of the incentives created by high-stakes accountability are different for high- and low-performing schools. High performing schools are arguably in a better position to focus on instructional improvement—the explicit intention of accountability policy—while probation schools must respond directly to policy requirements (e.g., getting a certain percentage of students to grade level). Therefore, because high-performing

schools are under less pressure, they can be more reflective and purposeful about instructional improvement.

SCHOOL STATUS AND THE USE OF DATA

Proponents of accountability policy also argue that test results provide objective information upon which to orient school decision making. The argument is that schools with clear, objective information about student performance will be able to make better informed decisions about instructional improvement (Coleman et al., 1997). Once again our data demonstrate distinct patterns in the interpretation and use of results in high-performing versus probation schools. We argue that the ways in which test results are interpreted depends, at least in part, on the school context.

Data Interpretation and Use in High-Performing Schools

At both high-performing schools, the full range of test score data were used to inform strategies of instructional improvement. At both schools, leaders use the outcomes for the entire school in addition to the item analysis to track overall school trends. For example, Blake school's principal and a collection of school leaders[10] are involved in data interpretation including longitudinal trend analysis and the analysis of movement between quartiles. During an initial meeting with the Mr. Preston, he shared a longitudinal data analysis for both mathematics and reading which highlighted the movement of students between quartiles for the entire school and for specific grade levels. In addition, he reported on the development of a particular grade level, the class of 2003, in terms of their performance in mathematics and reading.

At both Kelly and Blake school leaders use the item analysis to identify specific student needs within subject areas. For example, Dr. Williams explained how the item analysis allowed school leaders to identify overall student needs:

> We try to look at [test scores] in August if we have them back and we design our program a lot looking at those skills that are measured on the ITBS [Iowa Test of Basic Skills] and what the item analysis[11] indicate our weaknesses and our strengths and so forth.

Dr. Williams explained how the school uses test scores to track student performance. When asked about how this information informed the instructional program in mathematics she explained.[12]

> With the math I found that concepts, our children tend to do well in computation, pencil and paper, figuring out the problem, 2+2, whatever they do well but when it comes down to higher order thinking skills they tend to not do as well and we're working—we started last year we started focusing in on higher order thinking skills because the tests are moving more and more in that direction. The math problems they have to explain how they got the answers not just get the answer. So what we've been doing we have been working with this year a new initiative not all teachers are doing this but next year every teacher will be doing this. The math journal in which children they must explain—they must explain how it is that they arrived at the answer that they got and as I said the total math is not—it's pretty good cause we're at 61% but the math problem solving we tend to not do as well and especially in the area of math concepts. So we're working on that.

Her interpretation is enhanced through her interactions with others in the school who inform her interpretations and seek to implement instructional responses. The school's assistant principal, counselor, and technology coordinator all play active roles in data interpretation. The technology coordinator made a similar interpretation of the test score data.

> The biggest deficit in mathematics grade-wise, and we found it to be pretty much the same in every grade, was the word problems. Them interpreting word problems. Their computation skills are great. Just about in every grade. The computation scores were good, average or above average. Data interpretation was another area that we felt needed some work.

Blake made similar use of the test score data. Mr. Preston explained:

> One thing that helped us a lot in terms of being able to disaggregate our math data is we were able to disaggregate it into three sections called problem solving and data interpretation, concepts and estimation, and computation.

Therefore, both of the high performing schools used test results to identify macro-trends across the school and focus attention on areas of specific needs. In both schools, the "item analysis" was a diagnostic tool that helped them identify where they should focus their attention.

Data Interpretation and Use in Probation Schools

At the probation schools the interpretation and use of the data was less focused on its instructional implications. School leaders discussed the need to improve reading and mathematics and did speak in specific terms about subdimensions of these subject areas but did not speak about specific instructional approaches and strategies as they did in the high performing schools. In addition, the data was used at these schools in pretty much the form in which it came from the district. There was less repackaging of the information and limited analysis of specific trends (with the exception of the identification of specific students who were close to "passing" the exam).

The high-performing schools use the test score data in ways that seem consistent with the arguments of testing proponents. The test results are used to define students' specific instructional needs and provide a basis for school level instructional decision making. In contrast, the probation schools tend to focus more on the overall test results and have a less systematic strategy for turning test results into useful information for instructional change. This suggests that even the interpretation of "objective" data is situated. Within higher performing schools, there is a more substantive interpretation of the data. The resources to enable data interpretation are greater in higher performing schools. Therefore these schools are more likely to benefit from the information in ways that lead to instructional improvement.[13]

SCHOOL STATUS AND ACADEMIC PRESS

Academic press measures the normative emphasis on academic success and reaching certain standards of achievement among both teachers and students (Lee, Smith, Perry, & Smylie, 1999). We found that the probation schools increased instruction focus, but in ways that were designed to respond to the policy demands of the external environment—getting off of probation. Their efforts targeted certain students, certain grade levels, and certain academic subjects. In contrast, the high performing schools focused equally on mathematics and language arts instruction, emphasized improvement for all grade levels, and worked to enhance the learning opportunities of all students. We argue that academic press is likely to increase in these schools with high-stakes accountability. Schools will focus on academics more when the incentives structure shifts, with rewards and sanctions being tied to student performance. However, academic press as typically measured is content neutral. It suggests that high standards exist but tells us little about how these standards are operationalized inside schools. Therefore, what academic press means for different schools may vary. Moreover, when schools are positioned differently within the

accountability system their "press" might manifest in different ways. The following section discusses these patterns.

Probation Schools: Targeted Instructional Focus

One of the critical issues continually raised at the probation schools was getting off of probation. One way that the instructional focus manifested at these schools is in an effort to increase the number of students at or above cutoff points at benchmark grades. In these approaches, school leaders target certain students or certain grade levels for extra assistance in an attempt to reach minimum acceptable performance levels. For example, at Wexler School teachers and administrators focus on the benchmark grades in order to reduce their retention rates. At Wexler, half of the professional development meetings we observed were related to testing. In addition, professional development targeted certain grades (those that took the test) and subjects (e.g., language arts). The external partner focused its energy on teachers in the benchmark testing grades and provided exam preparation books only for teachers at those grade levels.

At Field School, one approach was to identify those students who were close to reaching national norms, and providing them with additional help. The school established an after-school tutoring program for these students. This tutoring process and the student selection process was discussed at a staff meeting by the assistant principal, Dr. Harold.

> The after-school program will start on Tuesday. All of you got the applications for your children and they need to be returned on Monday . . . we have one class for every grade level. The list of students may have seemed erratic. Ms. Lawrence chose those students according to their ITBS scores. She chose those students who she felt had the most potential to improve.

Later, another school administrator explained that the school targets those students who are closest to the threshold in its effort to get off of probation.

> The . . . program is . . . for students who are . . . very close to having the skills necessary to pass the test . . . .students in this program attend three times per week from 2:30 (dismissal) until 4:30 p.m. Other students are allowed to come but the students who are closest to passing the exam are targeted.

Ms Jackson later explained this further stating that the school will use this program to "work with selected students, taking a look at the IOWA scores

and just really focusing in on students that are at a median that we can work with and see if we can get some growth spurts on them."

The school's external partner, who attributed this focus to the external pressure being applied by the school district, also discussed this strategy.

> They [the school] leave behind [the lowest performing students] and focus on [the higher performing]. So many principals are under this pressure. It's the name of the game. When Vallas [the district CEO] comes and they have their region meetings, they are told they have to get off probation. Even if your school shows growth and doesn't get off probation, they realize they will be looked at as not doing the job.

Thus the lowest-performing students in this school received limited assistance in improving their scores. Instead, the instructional focus is on the students who were close to making the cut-off for probation requirements. In this case the external accountability mechanisms lead to a selective increase in instructional focus with limited implications for the lowest performing students. In addition to this selective focus, the program emphasizes content coverage (pacing) but not the teaching strategies that should be used. Thus the academic press in probation schools was targeted toward particular students rather than all students with the lowest performing students being least likely to get academic support.

Responses to testing in the probation schools were also structured by subject matter. As we noted previously all of the schools focused more on mathematics and reading instruction than science instruction. However, in the probation schools, the instructional focus emphasized one subject area—reading. For example, Wexler's administration was putting much of their efforts into language arts as opposed to other subject areas. Dr. Smith explained that the school's emphasis on language arts instruction was based on the fact that it had the potential to help the school get off of probation.

While instructional emphasis is one of the outcomes predicted by proponents of testing policy, the selective emphasis on certain students, grade levels, and subject matter areas may limit the impact of the policy for all of the students in the school. Moreover, the selective targeting of students seems to run counter to the intended impact of the policy. Probation schools adopt nonsystematic responses to accountability policy, which may marginalize students who face challenges.

High-Performing Schools: Global Instructional Focus

In both high-performing schools the exams are used to identify high and low-performing groups of students. However, in contrast to the probation

schools, the high-performing schools adopted interventions for all students, not just a subcategory. At Blake School, test score data were used by school leaders and teachers to diagnose the effectiveness of certain teaching approaches. For example, when fifth grade teachers at Blake met to discuss the prior years test results, they were pleased that the overall percentage of students at or above national norms had increased. However, upon closer examination, they discovered that the increases were among students moving from the second to the third quartile. In discussing this, they determined that as a group it was likely that they had been focusing instruction on students in the middle range, potentially not addressing the needs of the lowest and highest performing students. As Blake's principal, Mr. Preston explained, while the school advocates heterogeneous grouping the teachers had identified a potential pitfall.

> If you go and you look at the data that's out there about how student grouping effects achievement and you recognize that heterogeneous grouping with a competent teacher is gonna be your best shot at being able to give everybody an opportunity to succeed. What's the biggest challenge you then have as a teacher? Well the challenge is how do you deal with that? How do you manage that enormous diversity of talent that's in front of you? . . . How do you do that?
>
> That is to say you know the moment that you allow yourself as a teacher to either shoot at the middle you're gonna be, and underserve the youngsters who are really more ready. Or even worse, that you out of mostly good intentions to drive most of your teaching efforts by those youngsters who require the most remediation and are most needy of your time and who you're feeling most guilty because you're not serving.

Given their review of the test score data at the fifth grade level and similar analyses at other grade levels, the school developed an approach that sought to address the needs of all of the students while maintaining heterogeneous grouping.

> Well, our assumption is that in order to be able to [address the needs of all students] you need to teach high, if I can use that sort of piece of jargon, you need to teach high and re-teach to the middle and lower . . . .the strategy at the school is to continue to do the whole group instruction high. Teaching you know the mass majority of that whole group instruction is to teach high and then to make remedial provisions within the classroom as well as out of the classroom for

> additional tutorial and other kind of remedial work to get to those youngsters who are not getting it the first time. That's the strategy.

At Blake the analysis of movement within quartiles, and teachers' and school leaders' interpretation of that data, highlighted an instructional practice that needed to be revised. Test results therefore informed teaching strategies in ways that targeted the instructional focus toward all students rather than a subcategory.

The instructional focus of the high-performing schools also extended beyond the benchmark grades to include all grade levels. At Blake, school leaders tracked test score trends for all students and focused attention to all grade levels. For example, a team at Blake studied test data in math. They focused on longitudinal trend analysis of mathematics outcomes. They determined that high student performance in third grade was sustained through fifth grade but began to decline during sixth grade. They convened teachers across grade level to address this issue through aligning mathematics topic coverage across grade levels. Therefore, test data provides an important resource for planning at this school with the instructional focus extending to all grade levels rather than a subset. Kelly school used the monitoring system discussed above to target instruction to all students and all grade levels in ways similar to Blake School.

Testing resulted in very different patterns of instructional focus at the two sets of schools. In the probation schools, the instructional focus became targeted at certain grade levels, certain students, and certain subject areas. Low-performing students and students at nonbenchmark grades were unlikely to be impacted by these strategies. In contrast, the high-performing schools targeted instruction at all grade levels and all students. Though not reported in the data here, they also maintained a balance between mathematics and literacy instruction.[14]

DISCUSSION AND CONCLUSION

Having examined schools' responses to high-stakes accountability policy in high-performing schools and probation schools we argue that these contexts shape how the policy is enacted. Proponents of these policies argue that they will create new incentives for teachers, provide objective information for school decision making, increase academic press, and that the combination of these mechanisms will reduce gate-keeping practices. We argue that testing proponents are right that high-stakes accountability gets teachers attention, provides them with objective information, and increases their focus on issues of instruction. However, we argue that critics are right to caution us about the implementation of these policies because

the nature of incentives and teacher motivation, the use of testing data, and the targeted nature of instructional focus in probation schools marginalizes low-performing students, increases gate-keeping processes, and works against the spirit of the policy.

Our data show that probation schools and high-performing schools respond to accountability policy differently and that this has implications for issues of educational equity. In probation schools, responses to high-stakes accountability emphasize getting off of probation, partially through managing the impressions of external stakeholders. These efforts to convince outsiders that the school is engaged in change efforts perverted the intention of the policy in certain circumstances, prompting an emphasis on the trappings of instructional innovation rather than substantive change. In addition, leaders at these schools used test results to look at overall school and grade level outcomes but do not connect these results explicitly to instructional decisions as supporters of these policies suggest they would. Finally, with regard to increased instructional focus (academic press), while school leaders were able to demand teachers attention through a combination of threat and encouragement, they focused this attention in a process best described as "selective press" which targeted specific grade-levels, students, and subject matter areas all emphasizing reaching the goal of getting off of probation.

In contrast, the incentive structure at high-performing schools pushed school leaders to reward and encourage teachers for their accomplishments while creatively pushing for continued improvement. These schools use test data to track macro processes of student performance and set the schools' instructional agenda. This instructional agenda, unlike at the probation schools, focused on all students, all grade levels, and balanced attention to both of the core subject areas. Therefore, some of the key mechanisms through which high stakes accountability is supposed to impact students' educational opportunities are constructed very differently depending on the school's status in relation to the accountability system. This is very important because students' race and social class correlate with school accountability status.

Proponents of high stakes accountability suggest that these policies will reduce schools' gate-keeping processes and increase academic press. The data from these schools suggest that the extent to which this is the case may depend on the status of the schools in relationship to the high-stakes accountability policy. The data from Field and Wexler demonstrate that their response to accountability policy is situated in their probation status, which structures school-level responses to high stakes accountability in important ways. Because probation status adds pressure to school leaders and teachers, it can lead to practices that increase rather than reduce gatekeeping processes.

For example, Field School's practice of focusing tutoring programs to serve only the students who are close to national norms may have detrimental impacts on the lowest-performing students, those who the policy is designed to help. These students may be marginalized from interventions that could increase their educational outcomes. Moreover, the targeted responses of these schools, focusing on benchmark grades and subject areas with the greatest chance of passing the probation threshold, seems equally problematic. Focusing on the benchmark grades for intervention suggests that the increase of academic press may only impact a sub-set of students. Likewise, if the subject matter areas are selectively focused upon this may ultimately limit students' access to knowledge. The responses of Wexler and Field schools do not seem to represent an increase in academic press for all students, grade levels, and subjects as much as a calculated, strategic effort to respond to the policy demands of the external environment which may ultimately marginalize the lowest performing students. For instance, Wexler school focused on reading instruction because it was "one area that could impact probation," and Field school provided assistance for "students who are . . . very close to having the skills necessary to pass the test." This is not increased instructional focus for all students in any true sense.

Unlike the probation schools, the responses of school leaders at high-performing schools seem to more closely match the outcomes predicted by testing supporters. They contribute to the development of clearer instructional foci and seem to reduce gate-keeping processes within these schools by focusing on improving the learning opportunities for all students across all grade levels.

We believe that the different responses of these schools result, at least partially, from the external threat of the accountability policy and the lack of resources present in probation schools. That these schools are on probation because of dismal student performance on standardized tests suggests that teaching and learning in their classrooms needs improvement. However, given limited time to improve and very few additional resources, it is not surprising that probation schools scramble to make changes and often do so in ways that violate the stated intention and spirit of the policies. In contrast, officials at high-performing schools are able to build on past successes and accumulated expertise among teachers and administrators to more productively respond to high stakes accountability. Moreover, because they do not face external threat, leaders at these schools are able to focus on continuos improvement. While pervious work highlights additional factors relevant to understanding these processes, we argue that school accountability status is highly salient (Spillane et al., 2002).

The responses of these schools suggest that the implementation of accountability policy may work against increased educational equality. If

higher performing schools construct the policies in ways that increase their academic press and reduce gate-keeping while lower performing schools have the opposite effect, focusing primarily on responding to external threats, then the policies could exacerbate rather than challenge educational stratification. The situated nature of policy implementation should be an important consideration for school reformers. Policy implementation is very much a local affair and understanding the variation in context (even within districts) appears to be critical. Moreover, the fact that the likelihood of attending schools of different quality is associated with social class, race, and residential segregation suggests that broader structural factors impact school level processes in ways that should be attended to by policy makers.

This article is not meant to argue that high-stakes accountability policies have no positive impacts. In reality, these policies may have led to increases in students test results across the Chicago district and may increase schools' focus on instruction under certain conditions. However, we do question the ability of such policies to reduce race and class educational inequality. The findings reported here suggest that the highest-performing schools, those with higher percentages of middle-income and White students, may benefit more from accountability policy than the probation schools that are most in need of improvement. While this is not the stated intent of the policy, attention should be given to the potential implications of this process. Demanding accountability without providing schools with substantially more resources will not lead to sustainable improvement. In order to avoid detrimentally impacting low-performing schools, they must be provided with additional resources to enhance teacher quality and engage in meaningful professional development to improve instruction. Steps must also be taken to ensure that low-performing students are not further marginalized by a policy ostensibly designed to help them.

Notes

1 Preliminary versions of this paper were presented at the American Educational Research Association Annual Meeting, Seattle, Wa, 2001 and at the Center for Research on Educational Opportunity, The University of Notre Dame, November 10, 2002. Work on this paper was supported by the Distributed Leadership Project which was funded by research grants from the National Science Foundation (REC-9873583) and the Spencer Foundation (200000039). This research was also supported by the National Academy of Education/Spencer Postdoctoral Fellow Program. We are indebted to several project researchers for their help with data collection and analysis including Patricia Burch, Fred Brown, Amy Coldren, Loyiso Jita, Antonia Randolph, and Jennifer Zoltners. Northwestern University's School of Education and Social Policy and Institute for Policy Research also supported work on this paper. John Easton, Amy Stuart Wells, Gary Natriello, and the anonymous reviewers of Teachers College Record provided helpful comments on earlier drafts of this paper. All inquiries should be directed to John B. Diamond at University of Wisconsin–Milvaukee, Enderis Hall 545, P.O. Box 413-04143

or jdiamond@uwm.edu. All opinions and conclusions expressed in this paper are those of the authors and do not necessarily reflect the views of any funding agency or institution.

2 We define high-performing schools as schools where 50% or more of the students perform at or above national norms on the Iowa Test of Basic Skills. The Chicago Public Schools define probation schools as those in which fewer than 15% of students perform at or above national norms on the ITBS. This number was later increased to 20% and then 25%. While we focus on the data from these four schools, our data base consists of the 13 schools that are part of the Distributed Leadership Study.

3 This section borrows from a paper by Spillane et al. (2002).

4 The data reported here are taken from the Chicago Public Schools Office of Accountability public use data base. This information can be accessed through this site at http://acct.multi1.cps.k12.il.us/

5 We do not suggest that family background is an absolute determinant of the type of schools children attend. A substantial number of the students in magnet schools for example are African American and more than half are low-income. However, the data demonstrate a pattern of students being sorted into different types of schools in ways that correlate with race and class.

6 For those interest in school choice and parents' decision making around school attendance please refer to Fuller, Elmore and Orfield (1996) and Gewirtz, Ball, and Bowe (1995).

7 Our definition of high performance is specific to the Chicago context. Our definition of high performance in these schools means that more than 50% of students are scoring at or above national norms in reading and mathematics. This figure would not equate with high performance in many districts but is considered high performance in Chicago schools.

8 Some of the findings on school-level responses to high-stakes accountability policy reported here are also reported in Spillane et al. (2002) and Spillane, Diamond, et al. (2001). We indicate when we have borrowed from these previous reports where appropriate below.

9 The categories of influence included other teachers, principals, assistant principals, the local school council, parents, testing, standards, textbooks, and instructional specialists (i.e. mathematics teachers).

10 For example, in the case of a literacy committee report which drew extensively on test results, the Russian Bilingual teacher, a reading teacher, the librarian, the reading specialist, the drama teacher, and a local school council representative were all involved in the interpretation of test results and the development of the report.

11 The item analysis is a document that shows classroom and student level test scores by the items correct and incorrect.

12 The principal had a similar response for language arts. We limit the discussion to mathematics to save space.

13 While they may not gain as much in terms of absolute test score results, this likely occurs because of the fact that gains are more difficult at the top end because of ceiling effects.

14 It is important to note that in all of the schools more than 60% of students are working class (e.g., receive free and reduced lunch). We are not, therefore, talking about particularly privileged children.

References

American Educational Research Association. (2000). *Position statement concerning high stakes testing in pre-k–12 education*. Washington, DC: Author.

Anyon, J. (1981). Social class and school knowledge. *Curriculum Inquiry, 11*(1), 3–42.

Argyris, C., & Schon, D. A. (1974). *Theory in practice: Increasing personal effectiveness*. San Francisco: Jossey-Bass.

Ayalon (1994). Monopolizing knowledge? The ethnic composition and curriculum of Israeli high schools. *Sociology of Education*, *67*(4), 264–278.

Bankston, C. & Caldas, S. J. (1996). Majority African American schools and social injustice: The influence of de facto segregation on academic achievement. *Social Forces*, *75*(2), 535–555.

Blau, P. M., & Duncan, O. T. (1967). *The American occupational structure*. Chicago: University of Chicago Press.

Bourdieu, P. (1977). *Outline of a theory of practice*. Cambridge, UK: Cambridge University Press.

Bourdieu, P. (1986). The forms of capital. In J. G. Richardson (Ed.), *Handbook of theory and research of the sociology of education* (pp. 241–258). New York: Greenwood Press.

Bourdieu, P., & Passeron, J. C. (1990). *Reproduction in education and society.* Beverly Hills, CA, Sage.

Bowles, S., & Gintis, H. A. (1976). *Schooling in capitalist America*. New York: Basic Books.

Brophy, J., & Good, T. L. (1974). *Teacher-student relationships: Causes and consequences*. New York: Holt, Rinehart, and Winston.

Bryk, A. S., Easton, J. Q., Kerbow, D., Rollow, S. G., & Sebring, P. B. (1993). *A view from the elementary schools: The State of reform in Chicago*. Chicago: Consortium on Chicago School Research.

Carbonaro, W. J. (1998). A little help from my friends' parents: Intergenerational closure and educational outcomes. *Sociology of Education*, *71*, 295–313.

Clotfelter, C., & Ladd, H. (1996). Recognizing and rewarding success in public schools. In H. Ladd (Ed.), *Holding schools accountable: Performance-based reform in education* (pp. 23–64). Washinton, DC: The Brookings Institution.

Cohen, D. K., & Ball, D. L. (1998). *Instruction, capacity, and improvement* (CPRE Research Report Series, RR-42). Philadelphia: University of Pennsylvania.

Coleman, J., Scheider, B., Plank, S., Shiller, K. S., Shouse, R., Wong, H., & Lee, S. (Eds.). (1997). *Redesigning American education*. Boulder, CO: Westview Press.

Coleman, J. (1988). Social capital in the creation of human capital. *American Journal of Sociology, 94*, 94–120.

Coleman, J. S., & Hoffer, T. (1987). *Public and private high schools: The impact of communities*. New York: Basic Books.

Coleman, J., Sudha, E., Campbell, Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfield, F. D. & York, R. L. (Eds.). (1966). *Equality of educational opportunity.* Washington, DC: U.S. Government Printing Office.

Cremin (1951). *The American common school: An historical conception*. New York: Teacher's College Press.

Darling-Hammond, L. (1994). National standards and assessments: Will they improve education? *American Journal of Education*, *102*(4), 478–510.

Elloitt, M. (1998). School finance and opportunities to learn: Does money well spent enhance students' achievement? *Sociology of Education*, *71*, 223–245.

Elmore, R. F., Peterson, P. L., & McCarthey, S. J. (1996). *Restructuring in the classroom: Teaching, learning, and school organization*. San Francisco: Jossey-Bass.

Gewirtz, S., Ball, S. J., & Bowe, R. (1995). *Markets, choice, and equity in education*. Philadelphia: Open University Press.

Hedges, L. V., & Greenwald, R. (1996). Have times changed? The relation between school resources and student performance. In G. Burtless (Ed.), *Does money matter? The effect of school resources on student achievement and adult success* (pp. 74–92). Washington, DC: Brookings Institution.

Hess, F. (2000). *Changes in students' achievement in Illinois and Chicago, 1999-2000*. Washington, DC: The Brookings Institution.

Heubert, J. P., & Hauser, R. M. (Eds.). (1999). *High stakes testing for tracking, promotion, and graduation*. Washington DC: National Academy Press.

Hoa, L., & Bonstead-Bruns (1998). Parent-child differences in educational expectations and the academic achievement of immigrant and native students. *Sociology of Education*, *69*, 126–141.

Hofferth, S. L., & Boisjoly, J., et al, (1998). Parent extra familial resources and children's school attainment. *Sociology of Education*, *71*, 246–268.

Jencks, C.S, Smith, S. M., Ackland, H., Bane, M. J., Cohen, D., & Gintis, H., et al, (1972). *Inequality: A reassessment of family and schooling in America*. New York: Basic Books.

Knapp, M. S., & Associates. (1995). *Teaching for meaning in high-poverty classrooms*. New York: Teachers College Press.

Lareau, A. (1989). *Home advantage: Social class and parent intervention in elementary education*. New York: Falmer Press.

Lee, V. E., Smith, J. B., Perry, T. E., & Smylie, M. A. (1999). *Social support, academic press, and student achievement: A view from the middle grades in Chicago*. Chicago: Consortium on Chicago School Research.

McDill, E. L., Natriello, G., & Pallas, A. M. (1986). A population at risk: Potential consequences of tougher school standards for student dropouts. *American Journal of Education*, *94*(2), 135–181.

McNeal, R. B. (1999). Parent involvement as social capital: Differential effectiveness on science achievement. *Social Forces*, *78*, 117–144.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Thousand Oaks, CA: Sage.

Muller, C., & Schiller, K. (2000). Leveling the playing field? Students' educational attainment and states' performance testing. *Sociology of Education*, *73*, 196–218.

Newmann, F. M., Smith, B., Allensworth, E., & Byrk, A. (2001). *School instructional program coherence: Benefits and challenges*. Chicago: Consortium on Chicago School Research.

Oakes, J. (1985). *Keeping track: How schools structure inequality*. New Haven: Yale University Press.

Orfield, G. (1993). *The growth of segregation in American schools: Changing patterns of separation and poverty since 1968*. Cambridge, MA: Harvard Project on School Desegregation, National School Boards Association Council on Urban Boards of Education.

Orfield, G., Bachmeier, M. D., James, D. R., & Eitle, T. (1997). *Deepening segregation in American public schools*. Cambridge, MA: Harvard Project on School Desegregation.

Rist, R. (1970). Student social class and teacher expectation: The self-fulfilling prophesy of ghetto education. *Harvard Education Review*, *40*(3), 411–454.

Roderick, M. (1994). Grade retension and school dropout. *American Educational Research Journal*, *31*(4), 729–761.

Roderick, M., Bryk, T., Jacob, B., Easton, J. Q., & Allensworth, E. (1999). *Ending social promotion: Results from the first two years*. Chicago: Consortium on Chicago School Research.

Roscigno, V. J. (1998). Race and the reproduction of educational disadvantage. *Social Forces*, *76*(3), 1033–1060.

Roscigno, V. J. (2000). Family/school inequality and African American/Hispanic achievement. *Social Problems*, *47*(2), 266–299.

Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom*. New York: Holt, Rinehart, and Winston.

Sewell, W. H., & Hauser, R. M. (1980). The Wisconsin Longitudinal Study of Social and Psychological Factors in Aspirations and Achievement. In A. C. Kerckhoff (Ed.), *Research in sociology of education and socialization* (pp. 59–100). Greenwhich, CT: JAI Press.

Sewell, W., & Shah, V. P. (1968a). Parents' education and childrens' educational aspirations and achievements. *American Sociological Review*, *33*, 191–209.

Sewell, W., & Shah, V. P. (1968b). Social class, parent encouragement, and educational aspirations. *American Journal of Sociology, 73*(5), 559–572.
Shouse, R. (1997). Academic press, sense of community, and student achievement. In J. Coleman et al (Eds.), *Redesigning American education*. Boulder, CO: Westview Press.
Smith, J. B., Lee, V. E., & Newman, F. M. (2001). *Instruction and achievement in Chicago elementary schools*. Chicago: Consortium on Chicago School Research, University of Chicago.
Spillane, J. P., Diamond, J. B., Walker, L., Halverson, R., & Jita, L. (2001). Urban school leadership and elementary science instruction: Identifying, mobilizing, and activating resources in a devalued subject area. *Journal of Research in Science Teaching, 38*(8), 918–940.
Spillane, J. P., Diamond, J. B., Burch, P., Hallett, T., Jita, L., & Zoltners, J. (2002). Managing in the middle: School leaders and the enactment of accountability policy. *Educational Policy, 16*(5), 731–763.
Spillane, J. P., Halverson, R., & Diamond, J. B. (2001). Investigating school leadership practice: A distributed perspective. *Educational Researcher, 30*(3), 23–28.
Stodolsky, S. (1998). *The subject matters*. Chicago: University of Chicago Press.
Wilson, W. J. (1987). *The truly disadvantaged: The inner city, the underclass and social policy*. Chicago: University of Chicago Press.
Wong, R. S. (1998). Multidimensional influences of family environment in education: The case of socialist Czechoslovakia. *Sociology of Education, 71*, 1–22.
Wong, K., & Anagnostopoulos, D. (1998). Can integrated governance reconstruct teaching? Lessons learned from two low-performing Chicago high schools. *Educational Policy, 12*, 31–47.

JOHN B. DIAMOND is assistant professor of education at the University of Wisconsin–Milwaukee where he teachers in the Department of Educational Policy and Community Studies. His fields of interest include the sociology of education; stratification; race and class stratification; parent involvement; and leadership and organizational change. He served as research director of the Distributed Leadership Project (http://www.letus.org/dls), a program of research investigating the practice of school leadership in urban elementary schools, from 1999–2002. Recent publications can be found in *Sociology of Education, Education Researcher, The Journal of Curriculum Studies, Educational Policy, Anthropology and Education Quarterly*, and *The Berkeley Journal of Sociology*.

JAMES P. SPILLANE is associate professor of education and social policy, and a Faculty Fellow at the Institute for Policy Research, Northwestern University, where he teaches in both the Learning Sciences, and Human Development and Social Policy graduate programs. His research explores the policy implementation process at the state, school district, school, and classroom levels, focusing on intergovernmental relations and policy-practice relations. Spillane is principal investigator of the Distributed Leadership Project (http://www.letus.org/dls), a program of research investigating the practice of school leadership in urban elementary schools. He is associate editor of *Educational Evaluation and Policy Analysis*. Recent publications can be found in *Cognition and Instruction, Education Researcher, Journal of Curriculum Studies, Teacher's College Record, Educational Policy*, and *Sociology of Education*.