Note: This paper has been published as: Betts, J., "The Promise and Challenge of Accountability in Public Schooling," in Frederick M. Hess (Ed.) URBAN SCHOOL REFORM: LESSONS FROM SAN DIEGO, Cambridge, MA: Harvard Education Press, pp. 157-176. 2005.

The Promise and Challenge of Accountability in Public Schooling: The Case of San Diego

Julian R. Betts
Professor of Economics, UCSD
and
Senior Fellow, Public Policy Institute of California

October 29, 2004

This paper was written for the San Diego Review Conference, September 27-28, 2004. The author thanks Andrew Zau for assistance and many district staff for sharing insights, in particular Karen Bachofer and Superintendent Alan Bersin.

INTRODUCTION

One of the major recent reforms to American public education is the spread of accountability systems, consisting of content standards, student testing, and interventions and rewards for students and schools based on performance. The potential benefits from school accountability are many. It can raise public awareness of inequalities in outcomes across schools, reallocate funding toward the schools and students most in need, and reduce achievement gaps. Accountability systems can also create incentives for all participants in public education, from superintendents to the students themselves, to work toward the common goal of improving student achievement.

However, accountability systems may demand improvement in student achievement without providing the financial means to get there.

A second risk of accountability reforms is that they sometimes feature a "one-size-fits-all" approach that can create opposition from unexpected quarters. Betts and Costrell (2001) note that in some states the most vociferous complaints about standards and testing have come from well-to-do suburbs, apparently on the grounds that accountability systems interfere with schools that were already meeting affluent parents' goals for their children. [1]

A third risk is that unless information on student achievement flows smoothly to parents, administrators and teachers, the accountability system cannot effectively funnel remediation and intervention to the students most in need.

San Diego City Schools provides a fascinating case study of these potential benefits and pitfalls of accountability systems, because the district has taken steps to integrate the federal and state accountability systems, and at the same time has developed its own ambitious and controversial system of interventions for underperforming schools and students. Specifically, in the summer of 2000, the San Diego City Schools embarked on a set of major reforms collectively known as the Blueprint for Student Success. The Blueprint created a district-level system of accountability that uses achievement tests to identify students who are lagging behind. The district then intervenes with resources and methods designed to boost these students' performance. Initially, the Blueprint focused on promoting literacy, building on the model Chancellor of Instruction Anthony Alvarado

had designed and implemented while serving as superintendent of Community School District #2 in New York City. In recent years the Blueprint has expanded to include mathematics.

While San Diego was designing and implementing a district-level system of accountability, California was developing what is arguably one of the more rigorous statewide accountability systems in the nation. The state adopted content standards in English/language arts and mathematics in late 1997, and in science and history/social science in late 1998. In the spring of 1998, the state instituted a new testing system, the California Standardized Testing and Reporting program (STAR), that initially used a nationally normed test, the Stanford 9. The state used the scores on this test to calculate the Academic Performance Index (API), a widely publicized summary measure of school performance. The state has since added to the API criterion-referenced tests, including the California Standards Test (CST) and the California High School Exit Examination (CAHSEE), the latter of which was first given to grade 9 students in 2000. In calculating the API the state also replaced the Stanford 9 with another nationally normed test, the CAT/6, in spring 2003.

Once the standards and tests were in place, the state introduced rewards and sanctions for schools based on their students' scores on the Academic Performance Index. The Public Schools Accountability Act of 1999 and subsequent legislation established financial awards for schools, staff, and students. High-performing schools would be eligible for financial rewards under the High-Achieving/Improving Schools Program, which includes the Governor's Performance Award Program. In addition, the Certificated Staff Performance Incentive Awards and the Schoolsite Employee Performance Bonus, both created in 2000, awarded bonuses to certified teachers and all employees, respectively, at top-performing schools. In addition, high school students with high scores on the state test were given university scholarships. However, the state's troubled financial condition has effectively gutted the system of financial rewards for top-performing schools.[2] Another key element of this system is the Immediate Intervention/Underperforming Schools Program (II/USP), a voluntary program which

provides additional funding to struggling schools but which also metes out sanctions should the schools fail to improve sufficiently.

California uses schools' API scores to measure schools' progress. By 2020, all schools are expected to score 800 or above on the API scale. A school is judged to be making satisfactory progress if it has already reached that level or if its API score grows by at least 5 percent of the gap between 800 and the previous year's score (a school exactly meeting this progress each year would reach an API score of 800 in about 20 years). Schools must meet targets for all students as well as similar targets for specific "numerically significant" subgroups of students.

San Diego has devoted considerable resources to complying with California's accountability system while also complying with the more recent accountability and school choice provisions of the federal No Child Left Behind Act of 2001. The federal law requires states to administer annual tests in core academic subjects to students in grades 3 through 8 and at least once in high school. States are also required to establish proficiency standards for students, with the goal of all students meeting these standards by 2014. Schools that fail to make "adequate yearly progress" toward meeting these standards are required to offer school choice and free tutoring (known in the law as supplementary services) to their students.

The state and federal accountability systems differ substantially in the long-term goals they set for schools, and in the means they use to identify failing schools. As we will discuss, in June 2004, the San Diego school board approved a District Accountability Framework that attempts to integrate the federal and state accountability systems. This Framework partially reduces the various tensions between the state and federal systems. The district's own Blueprint in some senses is a separate and parallel accountability system, which focuses on instructional delivery, and interventions at the student rather than the school level.

OVERVIEW OF THE BLUEPRINT FOR STUDENT SUCCESS

The main difference between San Diego's accountability system and the state's is that the Blueprint focuses on teaching methods, especially in the area of literacy. The Blueprint emphasizes the concept of "Balanced Literacy," which calls for teachers to promote reading "by, with, and to children." In this approach, teachers become more actively involved as they introduce more difficult text to their students, by reading *with* children or, in the case of advanced text, reading *to* children. Similar approaches are taken for other elements of literacy including writing. [3]

The Blueprint includes three main strategies. The first strategy, *prevention*, applies to all students and schools and focuses on extensive training for teachers, enhancing teaching methods, and additional high-quality classroom materials. The second strategy is *intervention*. Teachers identify students performing below grade level, who then receive extra instruction through various programs. The final strategy is *retention*, the practice of having students repeat a grade with increased support. For simplicity, we will treat this below as one of the interventions.

During the first year of the Blueprint, 2000–01, the prevention strategies in place for all students included:

a) use of a new literacy framework in all grades,
b) "enhanced classes" for all kindergarten and 1st grade students. These classes receive additional classroom materials and the teachers receive additional professional development.
c) "Genre Studies," also known as "Enhanced Literacy," consisting of a two-period English class and related professional development, for all students in the entering grade of middle or junior high school who are near to, at, or above grade level,
d) One or two peer coaches for all schools, to help teachers learn proven teaching methods. Peer coaches are overseen by Instructional Leaders, seven administrators with responsibility for overseeing all aspects of the Blueprint at their assigned schools.

In addition, Focus Schools (elementary schools ranking in the bottom tenth of API scores statewide) received an extended school year, a second peer coach, and other funds and staff. The elementary schools that ranked in the second lowest decile of the state ranks, known as "API 2" schools, received a second peer coach and additional funds, but not an extended school year.

The second category of Blueprint reforms consists of various interventions. Unlike the preventive measures, the interventions are targeted at students identified as lagging behind in reading, based on test scores.

The key intervention strategies are as follows:

a) Literacy block. This variant of genre studies is given to students who lag "below" or "significantly below" grade level. Literacy block is a double-length English class offered in grades 6 through 10.

b) Literacy core. For students significantly below grade level in 9th grade, the literacy class is extended to three periods. In 2001-2002 students in grades 6 and 7 also began to participate in literacy core.

c) Extended Day Reading Program. In all schools with grades 1-9, students below grade level in grades 3, 6–8, and in 9th grade beginning in the winter of 2001 receive three hours each week of supervised reading before or after school.

d) Summer School. The Blueprint calls for two types of summer school. The first and more novel type is aimed at students in most grades from kindergarten through grade 9 who are below grade level. Students are asked to attend for six weeks, for four hours per day. In addition, all secondary school students with D/F grades attend a more traditional type of summer school consisting of six weeks of courses in core subjects. Some schools in the district, mostly elementary schools, are year-round schools, which means that their schedules do not permit the implementation of Blueprint summer school. At these schools, students in affected grades who lagged behind in reading participated in special Intersession studies.

e) Grade retention. In extreme cases, students may be asked to repeat a grade, and are given additional supports in the year that they repeat the grade.

Retention is limited to entry-level grades, that is, 1st grade and either 6th grade in middle schools or 7th grade in junior high schools.

The district decided not to use any of the state tests in making decisions about assigning students to the interventions. When the Blueprint was being developed, the state used only a norm-referenced test, the Stanford 9. District officials believed it was important to instead use a criterion-referenced test—that is, one that measures a student's level of performance compared with an absolute standard rather than relative to a nationally representative sample of students. For the lowest grades it has used the Developmental Reading Assessment, in which teachers work one-on-one with students and then evaluate where students' performance places them among several dozen stages of reading development. In grades 4–10 the district has used the Stanford Diagnostic Reading Test. Teachers may assign students to interventions if their test scores indicate that they are "below grade level" or "significantly below grade level." Students are often given the chance to take another, different test to confirm that they are below grade level. Students learning English are automatically eligible for many of the interventions, solely on the basis of not being sufficiently fluent in English.

The Blueprint relies on certain aspects of the state's accountability system while substituting for others. For instance, the reading curriculum focuses on the state's standards in reading. However, as discussed above, the district has decided to use reading assessments apart from those required by the state in order to identify students in need of assistance. The district continues to administer all required state tests, but for its Blueprint intervention programs instead uses criterion-referenced tests that tell teachers and administrators the extent to which each student has mastered the reading elements expected for a given grade.

Based on numerous interviews, my sense is that teachers and principals do make substantial use of the results of the district-administered tests. This is particularly so in the early grades, where the teacher individually administers the Developmental Reading Assessment to his or her students. This direct interaction then enables the teacher to tailor the subsequent "reading by, with and to students" accordingly.

The district's accountability interventions operate quite independently from the state's performance-based rewards and sanctions and there are no obvious synergies or

conflicts between them. For instance, the state's Immediate Intervention/Underperforming Schools Program is voluntary and gives schools additional financial resources in return for developing a suitable plan for improvement with the help of an external consultant. This program is distinct from the Blueprint's schoolwide measures, such as the additional resources devoted to Focus and API 2 elementary schools, neither of which is voluntary. In addition, the Blueprint elements are prescribed centrally, whereas the Immediate Intervention reforms are designed largely by the school staff with the help of an external consultant. Notably, though, the district retained one consultant for all of its II/USP schools, to create some uniformity.

It is also worth mentioning the novel ways in which the district has funded the Blueprint. First, the district obtained a federal waiver to use Title I money for its Blueprint reforms in Title I schools, while providing the same services in other schools using separate funds. This was essential because under the Blueprint any student regardless of poverty levels at her school can participate in the interventions if her reading scores are sufficiently low. Second the district has done a great deal to obtain funding from private foundations to support its reform efforts. Cumulatively, this funding amounts to over $35 million dollars and continues to grow.

The Blueprint has proven controversial locally. Business groups such as the Business Roundtable for Education have supported the introduction of the Blueprint, but teachers have expressed strong concerns about many of the Blueprint elements, such as the peer coaching program. [4]

A Latino lobby group criticized the Blueprint from the outset. The coalition's main concern was that the double- and triple-length classes in English for some students would exacerbate tracking, and divert students' attention from other subjects. In addition, the lobby group worried that these potential negative side effects would fall mainly upon the shoulders of Hispanic students. [5] Conversely, in late 2002 Hugh Price and Cecil Steppe, the presidents of the National Urban League and the San Diego Urban League respectively, wrote an op-ed praising the Blueprint for the contributions it has

made to narrowing the achievement gap. [6] Thus, local reception to the Blueprint has been complex.

Also, consistent with the observations of Betts and Costrell (2001) in other states, some of the most vocal opposition came from the more prosperous suburbs. Teachers and parents at La Jolla High School, with some of the highest test scores in the county, threatened to convert the school into a charter school, to free it from the strictures of the Blueprint. Ultimately, the district reached a compromise in which the school converted to a contract school, which is not subject to the Blueprint, but which remains a district-controlled school.

The Blueprint is both less than and more than a full-blown accountability system. It is more than simple accountability because it is a resource allocation mechanism, based on testing, that provides significant additional resources to students who are struggling. But unlike a full-blown accountability system it does not create its own content standards, nor does it clearly assign incentives or sanctions. The closest it comes to a sanction for students is grade retention in grades 1, 6, or 7 for a subsample of students who are more than three grade equivalents behind in reading. As the penultimate section will discuss, the Blueprint does not provide clear incentives or sanctions for teachers or principals.

THE IMPACT OF NO CHILD LEFT BEHIND

San Diego has faced the challenge of aligning its own accountability system with not only California's system but also with the extensive requirements of the federal No Child Left Behind Act of 2001. A core concept in the federal accountability system is Adequate Yearly Progress (AYP). Each state is required to test students in certain grades annually using one or more tests of its choosing. States must also define various achievement levels: Below Basic, Basic, Proficient, and Advanced. For each year, California has decided to set a minimum percentage of students at each school that should meet state standards, which requires students to be at Proficient or Advanced levels. To meet AYP requirements, the school must have at least this percentage of students at Proficient or Advanced each year. This minimum increases over time so that by the year 2014, 100 percent of students will be expected to be at Proficient or better.

Between 2002 and 2004, California elementary and middle schools had to have 16.0 percent of students at Proficient or better in math and 13.6 percent in English Language Arts, based on the relevant California Standards Tests. For high school, the initial target was for 11.2 and 9.6 percent of 10th graders to reach Proficient or higher levels on the English Language Arts and math sections on CAHSEE (the high school exit exam), respectively. California also requires that a school's Academic Performance Index score be above a certain point (560 during the 2002–03 through 2004–05 school years), rising to the long-term target of 800 by 2014. Alternatively, if a school's API is below the required level, showing at least a one-point gain will satisfy this criterion. It is not clear what the state's motivation is in allowing schools with such low scores to meet the criterion by gaining only one API point. Starting in 2003, high school graduation rates are also being evaluated, with an absolute target of 82.8 percent, and a target increase of 0.1 percent per year or 0.2 percent over two years.

Arguably the most important difference between the federal and state accountability requirements is how the two systems judge whether a school is making adequate progress. Under the state system, a school can pass muster either by scoring at or above the target level of 800 on the API or by improving its API score by 5 percent of the gap between its API in the previous year and 800. Most California schools (including San Diego's) have API scores below 800, but many of these still increase their API far enough to pass the state requirements. By contrast, under NCLB, to make Adequate Yearly Progress a school must have a certain proportion of students at Proficient or above, with this proportion increasing over time. Thus, in any given year, NCLB sets an absolute level of performance. The state system allows a school to "pass" either by meeting an absolute level of performance or by showing *improvement.* Thus, California schools face two subtly different evaluation systems. It is quite possible for a school to succeed in one system and fail in the other.

These differences highlight varying conceptions of pathways to reform. The state system allows for the realistic case in which schools are initially quite heterogeneous, but requires schools that are performing far below the long-run target of 800 to improve by a larger absolute amount than schools with higher initial test scores. Thus, all schools are expected to converge ultimately. The federal system instead sets the bar at an absolute

level, raising it over time. With the federal approach, schools that already performed well above the low initial requirements set for 2002 through 2004 have little incentive to improve. Perhaps it is only when the AYP "bar" rises close to these schools' own performance that the fear of sanctions will encourage these schools to improve.

Two other important differences between the federal and state systems involve the calculation of the overall measures of school success. Under NCLB, AYP uses a single cutpoint on test scores (Proficient or above), which creates a very real possibility that a school could boost scores of very low-achieving (or high-achieving) students and yet receive no credit. California's API better reflects of the entire distribution of test scores. A second difference is that the API weights improvements among very low-scoring students more heavily than improvements among very high-scoring students.

A key element of NCLB is that it creates up to 44 hoops through which a school must jump in order to meet AYP. Not only must test scores meet the stipulated level, but 95 percent of students must take the test (although the federal Department of Education recently loosened this stipulation). Schools with substantial numbers of students in any of 11 groups (8 racial/ethnic groups, disadvantaged students, English Learners, and special education students), must meet both the "percent proficient" and participation requirements for each subgroup. A school that is both large and diverse could face up to what district officials half jokingly refer to as "44 opportunities for success" (these are composed of two subjects (math and English language arts), times two hurdles (test score and participation requirements), times eleven student groups).

In a trial run, the state released 2002 Base AYP reports using 2002 test results. These reports used the four performance criteria related to math and English Language Arts test scores and participation rates, but did not use the API or graduation rate criteria. As a district, SDCS made AYP in 38 of 44 categories. Among the eight largest urban districts in the state, SDCS fared relatively well, scoring better than all the other districts save for Long Beach.

Notably, all of the largest districts fared much better in the 2003 AYP calculations than they had in the trial run in 2002. SDCS and Long Beach both met AYP in *all* 44 categories.

Schools that fail to meet any of the AYP requirements for two consecutive years in the same subject area are deemed to be in need of improvement. California calls such schools Program Improvement (PI) schools. In the first year of PI status, a school must offer busing to students who would like to attend a non-PI school. In later years, the school must also offer supplementary services (tutoring). Students may choose from a list of providers.

For students in the district, SDCS has become the main provider of the supplementary services required by NCLB. In 2002-2003 the district had a 99.9 percent market share, which fell to 74 percent in 2003-2004, as the number of non-district providers rose from one to five. In 2002-2003, students who accepted supplementary services from the district received a version of the Extended Day Reading Program (EDRP) that featured reduced class sizes. In 2003-2004, the district cut back EDRP because of financial pressures. Students electing for the program through NCLB's supplementary services continued to receive EDRP, but the length of the program was cut back. Other students who were not at PI schools in year 2 or higher were no longer able to participate in EDRP at all.

In interviews, SDCS staff displayed overall support for the accountability provisions of NCLB but chafed at some of the internal inconsistencies. Karen Bachofer explained that it was highly unlikely that by 2014 100% of EL students could score Proficient or higher because this group gets refreshed each year as new EL's enter the system. She continued:

> So overall, I don't think NCLB is a bad thing. It has got people's attention and has focused us on standards. But statewide in 2-3 years 65% of our schools could be in PI. There is something wrong with that picture. NCLB helps us target support and allocate resources. It does a lot for us, but I don't think that it is workable in its present form. In a few years where will we bus kids to? We won't have enough space in non-PI schools for PI students who would like to transfer.

Bachofer believes that the Blueprint complements the state and federal accountability programs because the Blueprint is targeted at individual students and

instruction while the other systems focus more on the big picture, helping to "prevent complacency."

In an interview, Alan Bersin, Superintendent of SDCS, agreed that the Blueprint itself was more fundamentally "a resource allocation strategy that moved resources in support of instructional strategies in places that we thought needed them most." However, he strongly supports the need for an accountability system, saying:

> [We need an accountability system] such that judgment day comes, either because parents march with their feet or because the district together with the Program Improvement regime has the wherewithal and the political courage to actually shut down schools that are not serving kids. I'm amazed when we talk about some of those middle schools – they have been at the bottom for 35 years now. And yet they continue, as students and families circulate through them, never staying long enough to actually identify them as being institutions of social dysfunction and the perpetuation of inequities. So you've got to have an accountability system that works. None of this functions if you don't have an accountability system.

Overall, though, the federal and state systems are somewhat redundant, and the different emphases in the state and federal accountability systems, with their separate focuses on growth versus absolute levels of achievement, are likely to continue to leave many parents and teachers somewhat confused.

GAPS IN COMMUNICATION AND TECHNOLOGY

There is little point in creating an accountability system unless information from the system is well understood and easily conveyed. This requires good communication of test score results. In turn, good communication of results depends critically on the quality of both the state's and the district's computer systems. It is quite clear from interviews with a number of participants in the system that much work is needed here.

A key example is that the introduction of the federal accountability system on top of the state's pre-existing system creates the potential for confusion. In a paper assessing the first two years of NCLB in San Diego, Betts and Danenberg (2004) interviewed nine representatives of community-based organizations representing parents and students.[7] These representatives as well as district administrators told the authors that the public as well as teachers had much to learn about NCLB, and that the sheer complexity of the federal and state systems would likely lead to public confusion.

What information do the state and district give to parents and teachers, and in what form? Does the computerization of test score results, or the lack thereof, limit the diagnostic powers of the testing regime?

Consider first the Blueprint. Parents receive information about how their children have fared on the Blueprint reading assessments. For those students deemed Below Grade Level or Significantly Below Grade Level, the district informs parents but goes beyond this, by asking parents to sign a Learning Plan. This Plan indicates recommended interventions. Parents can opt their children out of these interventions, but in signing the document the parents attest that they understand that if their child remains below grade level it can hurt their child's long-term educational outcomes. In this way the Learning Plan serves the dual goals of informing parents and obtaining their "buy-in".

The results of the district-provided reading assessments used for the Blueprint are readily available at the school level, and school staff uses these in making decisions about which students to assign to interventions. Based on interviews, it appears that teachers do use the results of the district-administered Blueprint tests in their teaching.

As for the state tests, parents receive information on their children's scores, and they also have access to annual school report cards that depict the overall characteristics of the school's students and teachers as well as performance on the state tests. The district also sends brochures to parents that indicate where to find out about the state content standards on the web. The district is also working on a revised report card that would tell parents the extent to which their children were mastering specific aspects of the state content standards.

Information flows related to the federal accountability program have been relatively weaker. Betts and Danenberg (2004) report that in the first two years of NCLB the district has had difficulty dealing with the timetable for providing school choice to students in newly designated Program Improvement schools. The state announces the list of Program Improvement schools in late August, leaving the district with only about one to two weeks to design busing options for students and to inform the parents. Administrators mentioned this as one of the most important informational bottlenecks in the overall three-tier accountability system.

One activity that does improve information flows concerning all three tiers of accountability is the Parent Congress, which provides updates to parents on district initiatives. Several parents from each school attend this quarterly meeting.

One sore point among both administrators is that the detailed student-level results of the state-mandated tests are initially provided to schools on paper only. Without a computerized format, school administrators have to do considerable analysis by hand to discover patterns of strengths and weaknesses in, for instance, individual classes. Later in the school year the district does receive computerized results from the state, but these are not user friendly. The district has prepared several pre-specified queries, but the district is not yet at a level where a principal or counselor can use a computer to study specific strands of a test either in a grade or in a specific classroom.

It appears that both the state and district could do more to provide teachers with computerized queries on a more detailed basis. This would help tremendously in diagnosing problems in specific classrooms or with specific students.

As for NCLB, it seems that information at the student level is being disseminated well, but few beyond administrators understand AYP or Program Improvement Status well. One district administrator told us that she had found from site visits to struggling schools that beyond the principal, few knew of all of the sanctions that could be levied against PI schools.

SAN DIEGO'S ATTEMPT TO RECONCILE THE STATE AND FEDERAL ACCOUNTABILITY SYSTEMS, WHILE IMPROVING INFORMATION FLOWS

SDCS has recently created a District Accountability Framework that achieves three important goals. First, the Framework attempts to bring together the state and federal accountability systems, to provide a single overall measure of each school's performance. Second, it provides for additional rewards and sanctions for schools. Third, it formalizes mechanisms for informing numerous parties at PI schools of the potential sanctions and interventions that such schools will face if they fail to improve. Variants of this system could prove quite useful in other districts nationwide, because, like San Diego, they almost always have to contend with both state and federal tiers of accountability.

The SDCS School Board approved the Framework in June 2004. As of this writing it has not been formally implemented but will be during the 2004-05 school year.

The School Classification Matrix

The first element in the Framework is a School Classification Matrix that places each school into one of 5 categories: Reward, Recognition, Monitor, Support and Intervention. The matrix, reproduced in Table 1, was adapted from a template developed by the California Department of Education. It compromises by allowing high API scores and/or meeting API growth targets to partly counterbalance having failed to meet AYP, and vice versa. For instance, as shown in the table, a school could fail to meet AYP under federal accountability and still score as high as "Recognition" status. Similarly, a school could fail to meet API growth targets under state accountability and still score at this same level. In both cases the trick is for the school to do well in the other accountability system.

District staff did a revealing simulation based on 2003 data. Figure 1 shows the percentage of schools in each of the five categories, while distinguishing between elementary, middle, high and charter schools. Clearly, elementary schools are faring far better than the other categories of schools. In part this has to do with genuinely better performance by the average student in the lower grades. But a second explanation is that elementary schools have fewer students, and therefore on average have fewer numerically significant subgroups that must meet the criteria of AYP and API growth. Notably, no middle school was in the Recognition group, although some were in the top Rewards

group, and no high school reached the Rewards group. Charter schools are quite heterogeneous, populating each of the five categories.

The accountability framework will go into effect officially during the 2004-2005 school year. It seems quite likely that over time there will be a shift of schools into the lower-performing categories because the AYP performance levels are set to increase markedly in 2005.

Rewards and Sanctions

The Framework lists all of the interventions mandated by NCLB. The Framework goes beyond this though, and stipulates additional rewards, interventions and sanctions. At the top end, Reward schools receive public acknowledgment and official school board recognition. Perhaps more meaningful, Reward schools obtain the right to negotiate increased flexibility. As of summer 2004 the exact form of this flexibility had yet to be determined, but it seemed quite likely that it would entail greater school site ability to spend portions of Blueprint funding. (Charter schools will instead receive a one-year charter extension.) [8]

At the next level, Recognition schools receive public recognition but do not receive increased flexibility or, in the case of charters, an automatic charter extension.

At the third level, schools in the Monitoring group receive monitoring from the Instructional Leader (IL) assigned to that school. The IL will help identify problem areas as revealed by test scores and recommend changes. Because charter schools are exempt from the Blueprint, and are thus not under the supervision of an IL, charter schools in the Monitoring level instead are subject to an annual programmatic audit.

At the fourth level, Support schools receive help from the IL who facilitates both revision of the school's plan and a variety of forms of technical assistance, including data analysis and professional development.

At the fifth and bottom level, Intervention schools receive similar assistance to that received by Support schools, but received direct oversight by the Superintendent and Chief Academic Officer of the district.

Informing Relevant Parties about PI Status

One of the main goals of the Framework is to improve dissemination of information about each school's strengths and areas that need improvement. To this end, the district requires the principal and Instructional Leader responsible for a given school to sign an accountability agreement annually. This agreement sets out performance expectations, and lists in detail potential rewards and interventions. In addition, the schools School Site Council chairperson and Site Governance Team chairperson signs the form.

These documents are particularly clear about the implications of falling into years 3 or 4 of PI status. The document states that in year 3, a school must work with the district to develop a plan for District Corrective Action that the district board must approve. On June 30 of year 4, the school will be restructured.

The accountability agreement for charter schools clearly states that charter schools that reach year 4 of PI status will not have their charter renewed by the district. As one official told us, charter schools in this important sense have much more direct accountability than do regular public schools, because charters explicitly face the threat of outright closure. [9]

The need for these accountability documents was made clear by Karen Bachofer, who told me that "We have visited all of the year 3 PI schools and were surprised at how many parents, community members, teachers and advisory members had no idea that they were in danger of entering into year 4 PI and therefore subject to reconstitution." The requirement for signatures on these documents may help to inform all stakeholders.

MISSING AND INFORMAL ELEMENTS OF ACCOUNTABILITY

Accountability at the district, state and federal levels is clearly a work in progress, and some important elements of both sanctions and rewards have yet to be fully articulated.

Perhaps nowhere is this problem clearer than in the sanctions facing schools at the end of their fourth year in PI status. NCLB allows for many possible sanctions, including re-opening the school as a charter, replacing all or most of the staff, contracting with a third party to run the school, or a state takeover. Adding to the uncertainty is the late

August release of AYP results by the state of California, which will give the district under a year to decide on appropriate restructuring. As of July 2004 the federal government and the state had yet to issue guidance on how to proceed.

Both the Blueprint and NCLB's choice and supplementary service provisions deliver additional resources and choices to students who lag behind or who attend underperforming schools. More nebulous is the set of rewards and sanctions for teachers and principals. Teacher and administrator pay is not related to individual student performance or to school-wide results related to AYP or API, with two statewide salary bonus programs enacted in 2000 having been canceled for lack of funding after the 1999-2000 school year. Neither is the assignment of teachers and administrators to specific schools or, indeed, the job security of teachers or administrators, specifically related to any of the three tiers of accountability. The clearest sanction in place is that one of the four potential sanctions for schools in year 5 of PI status is to replace all or most of the staff including the principal. Teachers and principals would presumably want to avoid re-assignment to other schools, but it is far from clear how often the district will opt for this choice. Furthermore, according to district officials, it is extremely difficult to fire teachers or principals. So, in practice the most that will happen is that principals and or teachers may be re-assigned: they will *not* be fired.

The district has reassigned principals in some cases and interviews suggest that student test scores played a part, but not a decisive part, in these decisions. For example, in an interview Superintendent Bersin suggested that test scores had played some role in re-assigning principals, but that the oversight role of Instructional Leaders under the Blueprint has been the more decisive factor in identifying principals who should be re-assigned.

CONCLUDING THOUGHTS

SDCS mirrors a broader community of districts around the nation that face similar hopes and challenges regarding their states' accountability systems and the federal system. It is likely that district administrators around the country have dealt with similar problems related to the degree of consistency between the two sets of accountability requirements.

The introduction mentioned several potential benefits from accountability. One of the most important is its impact on student achievement. Success in terms of student achievement has been difficult to assess because studies to date have examined overall test-score trends without observation of which reforms were implemented for specific students. Raymond (this volume) provides an overview of test score trends and the AIR's two published evaluations of the Blueprint. To date, no researchers have produced a detailed longitudinal evaluation at the student level of the impact of specific Blueprint interventions. Betts, Zau and King (in review) are undertaking such an analysis and final results should be released soon. As for the impact of the federal accountability system it is far too soon to measure its impact, and the impact of the state system on San Diego has yet to be studied in detail.

Despite the limited evidence to date on achievement, the success of the three accountability systems can be meaningfully assessed in terms of implementation and changes in culture that have resulted. For instance, the introduction mentioned that accountability systems can raise public awareness of achievement inequalities. Both the federal and state systems have succeeded to some degree here. However the co-existence of two not entirely consistent accountability systems definitely poses challenges to public understanding.

In this regard, the SDCS District Accountability Framework could potentially be of use to districts around the nation. Not only does it combine the state and federal accountability systems so as to provide a single overall rating of each school, but it also provides for various rewards and sanctions, only some of which are required by state and federal law. By going beyond these mandated requirements, the district has signaled that it takes accountability seriously, and is willing to devote at least some additional resources to struggling schools. The Framework also requires that each school discuss its specific accountability agreement with administrators, teachers, parents and the School Site Council.

Another potential benefit of accountability mentioned in the introduction is that it can cause all actors from the superintendent to parents, administrators and teachers to "pull together" on behalf of improving student achievement. Accountability has certainly become a central element in the work of both administrators and teachers. The Blueprint

reforms in particular have affected how and what teachers teach in each grade. In an interview, district administrator Karen Bachofer praised the overall accountability system because it provides "focus and consistency in application,"—that is, it forces schools to focus attention on student achievement, while holding regular district-managed, charter and contract schools to the same standards.

At the same time, she said, accountability also has the potential to distract educators. She said:

> … if we are not careful it [accountability] will pull our thinking away from instruction. … Sure I said that AYP and API can focus us, but on the other hand it could move us away from focusing on effective and appropriate teaching methods and practice, which is the only way to improve our standing in the accountability system.

A quite separate lesson from the San Diego experience is that a district need not feel hemmed in by either state or federal accountability. The district developed its Blueprint for Student Success fairly independently, and it complements the federal and state systems.

What about the potential pitfalls to accountability mentioned in the introduction, the possibility of poor flows of information to teachers and parents, a lack of funding to achieve the mandate, and the potential reaction to one-size-fits-all reforms? More broadly, what has been the political fallout?

Timely dissemination of results from the state test continues to be a key weakness of the accountability system. Because newly designated PI schools are announced in late August, the district is forced to design new bus routes, and notify parents of their choice options in the space of one to two weeks.

Perhaps an even bigger informational problem is that schools initially receive results of the California Standards Test on paper only. This makes it exceedingly difficult for teaches and counselors to pinpoint weaknesses in curriculum in a timely manner.

The inaccessibility of the data on the state-mandated tests stands in fairly stark contrast to the way in which schools and teachers make use of the district-mandated reading assessments under the Blueprint for Student Success. This is especially so in the

lower grades, where teachers administer one-one-one reading assessments with their students. The results from such assessments directly feed into how the teacher works with his or her class.

On the funding question, the state accountability system has largely been an unfunded mandate apart from limited pots of money such as the II/USP program for underperforming schools. Nationally, many have criticized the federal government for underfunding NCLB. In many interviews, for this paper and Betts and Danenberg (2004) we did not hear district officials echo this complaint. However administrators are clearly concerned that when the test-score bar is raised in future years, the federal funding for choice and supplementary services will quickly become exhausted.

The district's own reforms under the Blueprint for Student Success are less susceptible to criticisms of underfunding. Indeed, the district has perhaps set an example for other large districts to emulate. First, the district obtained a federal waiver to use Title I money for its Blueprint reforms in Title I schools, even though the same programs were provided to other schools with separate funding. Second, it has done a great deal to obtain tens of millions of dollars of funding from private foundations to support its reform efforts. At the same time, cutbacks related to California's state budget crisis have reduced the scope of the Blueprint in recent years. The district could do much more with more state funding than is currently possible.

The third potential landmine confronting accountability systems is the political fallout. Clearly, politics has been the Achilles heel of the Blueprint for Student Success. Stein, Hubbard and Mehan (2004), Cuban and Usdan (2003), and others have characterized the Blueprint implementation as "top-down". [10] This approach has led to considerable teacher unhappiness, as shown by these interviews conducted by Stein et al. and by teacher surveys performed by AIR. Similarly, the declaration of one of the top high schools in the district as a 'contract' school exempt from the Blueprint was typical of the negative response to accountability by high achieving schools observed elsewhere.

A reasonable conclusion might be that other districts contemplating similar reforms should implement more slowly, in consultation with teachers, their union, school administrators and parents. This conclusion may be right, but it is impossible to know

what would have happened if an alternative, less contentious approach, had been taken. The reforms could have been watered down substantially, and surely would have been implemented later.

While strongly supporting accountability, Superintendent Bersin is concerned that resistance to using test scores, and the likely difficulty of shutting down PI schools in year 5, threaten the long-term viability of accountability. He said:

> Now what I care about is that because so many actors inside the education world consider those data points to be illegitimate whether they be AYP points or API growth points, that is one of the reasons the sector has been able to escape year in and year out from accountability. Because parents move on with their kids, educationists don't accept the legitimacy of these data points, and therefore they capitalize on the difficulty of shutting down schools. As Admiral Rickover once said so eloquently, there is only one thing more difficult than closing an American school, and that is moving an American cemetery.

Only time will tell what happens with regards to school restructuring in San Diego. The Superintendent expects that some schools will indeed be closed and re-opened in coming years as schools enter their fourth year of Program Improvement status.

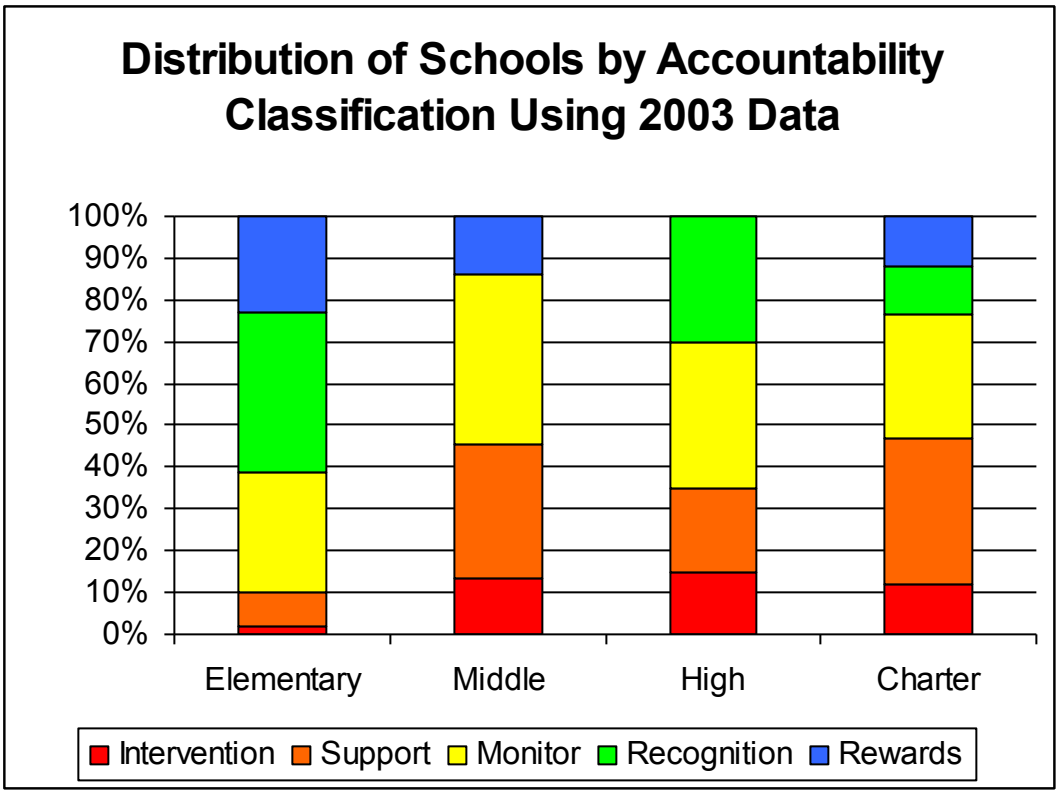
Apart from political challenges, perhaps the biggest challenges to accountability is that as the requirements to meet AYP ramp up sharply in coming years, the specter feared by district staff – that the majority of schools will fall into PI status – is quite likely to become a reality. At that time, either the standards or the timetable will have to be altered. Failing that, the entire accountability system may collapse, which would be doing a disservice to the accomplishments already achieved in San Diego and so many other districts nationwide.

Table 1 The SDCS School Classification Matrix

	API Score		API Growth		AYP Requirements
Rewards	800-1000	and	Met all targets	and	Made AYP
Recognition	800-1000	and	Met all targets	or	Made AYP
	600-799	and	Met all targets	and	Made AYP
Monitor	800-1000	and	Did not meet all targets	and	Did not make AYP
	600-799	and	Met all targets	or	Made AYP
	200-599	and	Met all targets	and	Made AYP
Support	600-799	and	Did not meet all targets	and	Did not make AYP
	200-599	and	Met all targets	or	Made AYP
Intervention	200-599	and	Did not meet all targets	and	Did not make AYP

To meet API growth targets, a school must meet its school-wide target as well as targets for all numerically significant subgroups. To meet AYP requirements, it must meet objectives in English language arts and math, and the additional API requirement and for high schools, graduate rate requirements.

Figure 1


Distribution of Schools by Accountability Classification Using 2003 Data
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Elementary
Middle
High
Charter
Intervention
Support
Monitor
Recognition
Rewards

[1] Julian R Betts and Robert M. Costrell, "Incentives and Equity under Standards Based Reform," in *Brookings Papers on Education Policy 2001*, ed. Diane Ravitch, (Washington, DC: The Brookings Institution, 2001).

[2] The Governor's Performance Awards were made for school years 1999-2000 and 2000-2001 only, and the other two programs were funded in 1999-2000 only.

[3] Mary Kay Stein, Lea Hubbard, and Hugh Mehan, "Reform Ideas that Travel Far Afield: The Two Cultures of Reform in New York City's District #2 and San Diego", *Journal of Educational Change* 5, no. 2 (June 2004), 161-197.

[4] See American Institutes for Research, *Evaluation of the Blueprint for Student Success in a Standards-Based System*, Palo Alto, California, American Institutes for Research, (2002) and Quick, Heather E., Beatrice F. Birman, Lawrence P. Gallagher, Jean Wolman, Kassandra Chaney, and Hiroyuki Hikawa, *Evaluation of the Blueprint for Student Success in a Standards-Based System, Year 2 Interim Report*, Palo Alto, California, American Institutes for Research, (2003).
[5] Alberto Ochoa, Op-ed, *San Diego Union Tribune*, October 26, 2001.

[6] Hugh B. Price and Cecil H. Steppe, "The Achievement Gap is Closing," op-ed, *San Diego Union Tribune*, October 25, 2002.

[7] Julian Betts and Anne Danenberg, "San Diego: Do Too Many Cooks Spoil the Broth?" in *Leaving No Child Behind,* eds. Frederick Hess and Chester Finn (London, Palgrave Macmillan, 2004).

[8] In addition to regular public schools and charter schools, the district has a third category of schools known as contract schools. One of these is La Jolla High School, a high-scoring school that signed a contract exempting it from the Blueprint for Student Success. Similarly, three schools in City Heights have signed contracts that grant them considerable autonomy. Rewards and sanctions are similar to those for charter schools.

[9] Similarly, the contract schools face non-renewal of their contracts if they enter Intervention status, or are in Support status unless the AYP target is met.

[10] Stein et al., op. cit., and Larry Cuban and Michael Usdan, "Fast and Top-Down: Systemic Reform and Student Achievement in San Diego City Schools," in *Powerful Reforms with Shallow Roots: Improving America's Urban Schools,* eds. Larry Cuban and Michael Usdan (New York: Teachers College Press, 2003).