

Pilot Study Report: The Local District Performance Measures

Steven M. Cantrell, Ph.D.

Nancy Lyon

Rosa Valdes, Ph.D.

Jeffrey White

Abelardo Recio, Ph.D.

Program Evaluation and Research Branch

6/25/01

Table of Contents

Table of Contents	ii
List of Tables	iv
EXECUTIVE SUMMARY	1
Pilot Testing.....	1
Indicator 1: API Growth Target Findings.....	2
Indicator 2: Matched Student Gains in Reading Findings.....	3
Indicator 3: School Organization Index Findings.....	3
Indicator 4: Classroom Assignments	7
Implications and Future Directions.....	8
1 Why the Local District Performance Measures?	9
2 A Balanced System for Accountability: Performance and Process Indicators.....	10
2.1 API Growth Target.....	10
2.2 Matched Student Gains in Reading.....	11
2.3 School Organization Index	11
2.4 Classroom Assignments.....	13
2.5 Further Technical and Operational Considerations.....	14
3 Local District P— Demographic Comparisons with LAUSD	15
4 Data Collection • Performance Indicators.....	16
4.1 API Growth Targets	17
4.2 Matched NCE Gains	18
5 Process Indicators • School Organization Index & Classroom Assignments	19
6 Data Collection • The School Organization Index.....	20
6.1 The Instrument.....	20
6.2 The Sample	20
6.3 The Protocols.....	21
7 Measures and Results • School Organization Index.....	24
7.1 Domain 1: Leadership	24
7.2 Domain 2: Quality of Teaching	28
7.3 Domain 3: School Climate and Culture	32
7.4 Domain 4: Appropriate Use of Student Assessment	36
7.5 Domain 5: Site-Level Commitment to School Improvement and Professional Development.....	38
8 Data Collection • The Classroom Assignments	40
8.1 The Instruments	40
8.2 The Sample	41
9 Measures and Results • Classroom Assignments	44
9.1 Classroom assignment scales	44
9.2 Sample Fourth-Grade Assignments	52
10 Scaling-up the Local District Performance Measures – 2001-2002	57
10.1 Scoring Considerations	57
11 Modifications to LDPM Design for 2001-02 School Year.....	60

Appendix A: Ranking Schools Using SAT/9: Six Approaches.....	A-1
School-level Residuals or Adjusted Gain Scores?	A-4
Appendix B: Indicators of Effective School Organization	B-1
Correlate One: Leadership	B-2
Correlate Two: Quality of Teaching	B-3
Correlate Three: School Climate and Culture	B-5
Correlate Four: Appropriate Use of Student Assessment	B-6
Correlate Five: Site-Level Commitment to School Improvement and Professional Development	B-7
References	B-9
Appendix C: Indicators of Classroom Practice.....	C-1
Appendix D: Changes to the SOI Protocol for 2001-2002.....	D-1

List of Tables

Table 1. Demographic Comparison of Local District P and LAUSD 16

Table 2. Number and Percentage of Schools meeting API Targets by School Level.... 18

Table 3. Number and Percent of Schools Meeting NCE Gain Expectations by School Level 19

Table 4. Harris Interactive Data Collection November -December 2000..... 21

Table 5. Overall Data Collection November 2000 – April 2001 21

Table 6. Number and Percent of Schools Meeting Performance Goals by Leadership Level 25

Table 7. Number and Percent of Schools Attaining an API Award by Level of Importance of Standards-based Grading 31

Table 8. Number and Percent of Schools Meeting Performance Goals by OCB 36

Table 9. Number and Percent of Schools Meeting Performance Goals by Presence of an Instructional Monitoring System..... 39

Table 10. Fourth-Grade Writing Assignment Scores by Dimension 48

Table 11. Fourth-Grade Reading Comprehension Assignment (#1) Scores by Dimension 49

Table 12. Fourth-Grade Reading Comprehension Assignment (#2) Scores by Dimension 49

Table 13. Seventh-Grade Writing Assignment Scores by Dimension 50

Table 14. Seventh-Grade Reading Comprehension Assignment (#1) Scores by Dimension 50

Table 15. Seventh-Grade Reading Comprehension Assignment (#2) Scores by Dimension 51

Table 16. Tenth-Grade Writing Assignment Scores by Dimension 51

Table 17. Tenth-Grade Reading Comprehension Assignment (#1) Scores by Dimension 52

Table 18. Tenth-Grade Reading Comprehension Assignment (#2) Scores by Dimension 52

Table 19. Scorecard for Local District P 57

Table 20. Quality Ratings of Classroom Assignments by Dimension and School Level 59

Table A-1: Distribution of Rankings by Initial NCE scores (Thirds).....A-4

Table A-3 Spring 1998 to Spring 1999 Grade-Level Gains by Initial NCE Score (Quintiles).....A-6

Table A-4 Spring 1998 to Spring 1999 Grade-Level Gains by Initial NCE Score (Thirds)A-7

EXECUTIVE SUMMARY

The Local District Performance Measures (LDPM) were commissioned by the General Superintendent in May 2000 to track the performance of Local District Superintendents. The system is designed to report on and to influence those elements of the school system most important to student performance. As originally proposed, the system was designed for formative use during the 2000-01 school year, and summative use during the 2001-02 school year.

The Program Evaluation and Research Branch has undertaken a pilot study to test the use of four indicators to assess the performance of Local Districts. Indicators 1 and 2 reflect direct outcomes of student performance, while indicators 3 and 4 reflect school processes expected to increase student performance.

LDPM indicators:

1. Percent of schools meeting their API Growth Target;
2. Percent of schools meeting their expected matched student reading gain on the SAT/9 (Matched NCE Gains);
3. Percent of schools reaching a satisfactory rating on the School Organization Index; and
4. Percent of schools reaching a satisfactory rating on the Classroom Assignments.

The first year (2000-01) was intended to provide formative feedback to the system's design and implementation. The pilot test data also helped establish standards for the four indicators. During the 2001-02 school year, the system will yield data that will inform local districts about their performance on these indicators.

Pilot Testing

For Indicators 1--API growth target and 2--matched NCE gains, a fictitious, local "District P(ilot)" was constructed, using student performance data across all local district schools (i.e., geographic stratification). District P performance on Indicators 1 and 2 was then compared to the entire District.

For Indicator 3--School Organization Index, 35 of the 55 District P schools were selected for the pilot. Principals, teachers, parents and students took part in a

20-minute telephone interview. Reasonable response rates were met for most of the respondent groups except for students (principals—86%, teachers—70%, parents—50%, and students—29%).

For Indicator 4--Classroom Assignments, fourth, seventh, and tenth grade language arts teachers from 35 of the 55 District P schools were asked to submit samples of their classroom assignments. Scales developed at the National Center for Research on Evaluation, Standards, and Student Testing (CRESST) were used to rate the quality of assignments. Response rates for this sizeable request averaged to 29% across the local districts.

Indicator 1: API Growth Target Findings

In line with the actual statewide API award rate, the API growth target for the district is 67% of schools meeting a 5% growth rate. (All under performing schools in California are expected to meet a 5% growth target for 2000-01, based on the computational approach embedded in the API.)

In order for a school to receive an award, two different growth targets must be achieved, a schoolwide target and a comparable improvement target. The schoolwide target states that a total school's API score must have grown by 5% of the distance between their baseline API score and the statewide goal of 800. The comparable improvement target specifies that each numerically significant subgroup achieve 80% of the schoolwide growth target.

Overall, the schools within District P were less likely to meet state targets and win awards than were other schools across LAUSD, which we attribute in large part to our sample design. (In order for us to test the full range of instrumentation, we oversampled secondary schools; compared to elementary schools, secondary schools are less likely to meet their targets.) The 66% of total district schools winning an API Award fell just short of the 67% API Growth standard.

Indicator 2: Matched Student Gains in Reading Findings

For each school, the expected gain target was calculated using the historic averages for their level of schooling (elementary, middle school, senior high) and 1999 performance grouping (quintile). Performance on this indicator was based upon the percent of schools that had met or exceeded their expected gain targets.

In this comparison, District P did not meet its goal of having 60% of schools meeting their expected gains. Elementary and senior high schools roughly matched the performance of the total district whereas middle schools exceeded that of the total district.

Indicator 3: School Organization Index Findings

We have preliminarily identified five aspects of school organization that we believe are correlated with school effectiveness: 1) leadership; 2) quality of teaching; 3) school climate and culture; 4) appropriate use of student assessment; and 5) commitment to school improvement and professional development. We believe that the use of these elements in an LDPM system will encourage schools to attend to these critical organizational features. Findings for the key elements of school organization are described next.

Leadership. We conceptualized leadership as consisting of instructional leadership and trust. We found a strong relationship between leadership and API performance (see Table 6). Of those District P schools receiving an API award, 80% (16 of 20) were led by principals who were scored by their teachers at or above 4.00 on our leadership measure which ranged from 1 (low) to 5 (high). Conversely, of those schools not receiving an API award, 66.7% (8 of 12) were led by principals who scored below 4.00 on our leadership measure.

Quality of Teaching. Several components make up Quality of Teaching. One component concerns *task and skill emphasis*, which was examined by asking teachers and their students to report on the instructional strategies used within the classroom during the past week. In general, teachers reported modes of

instruction that require students to be active learners. Students at the middle school generally confirmed the teacher self-reports. At the senior high school, however, students reported more passive forms of instruction.

Regarding *classroom effectiveness*, or the extent that teaching practices are viewed as effective by various stakeholders, parents reported that they are generally confident about how well teachers are prepared to help their children.

Moreover, principals reported that their primary emphasis is facilitating each student's progress in language and number skills (\bar{x} =3.94 of 4.00, s.d. = 0.25). This is followed closely by raising average performance on standardized tests (\bar{x} =3.72, s.d.=0.46). Not coincidentally, these are also the two areas where principals rated teachers as being the least successful (respectively, \bar{x} =2.94, s.d.=0.76; \bar{x} =2.97, s.d.=0.93).

For *grading practices*, we looked at the relationship between standards-based grading and student performance. Results show that schools that practiced standards-based grading tended to meet their NCE matched reading targets and API-related performance goals. Using 3.25 (out of 4) as our cutoff point, we looked at the performance of schools whose teachers reported using absolute level of achievement to determine student grades. Only 9 of 35 schools met our 3.25 criterion for standards-based grading. Of these 9, 6 met their NCE matched reading target (67%). This compares with a 35% NCE target attainment rate for schools that do not employ standards-based grading. Our examination of performance on the API provided similarly striking patterns. Eighty-five percent of the standards-based grading schools attained API goals and awards, compared to 54% of those schools not employing standards-based grading.

Teachers' perceptions of their *self-efficacy* matched the confidence parents have in their child's teacher's effectiveness. An example of an efficacy item asked of teachers is "Please indicate whether you have a strong working knowledge of methods to assess students." On our 5-point scale, all groups of teachers reported that they are very effective at teaching their current students—

elementary teachers (\bar{x} =4.39), middle school teachers (\bar{x} =4.36), and senior high teachers (\bar{x} =4.00).

School Climate and Culture. Regarding the presence of a *Safe and Orderly Campus*, the essential question, “Is this school a safe place?,” was answered in the affirmative by all respondent groups. Administrators are the most positive about school safety (\bar{x} =4.59 of 5.00, s.d.=0.49), followed by teachers (\bar{x} =4.30, s.d.=0.57), and then parents (\bar{x} =3.67, s.d.=0.54). Of all groups, students were most wary about school safety (\bar{x} =3.49, s.d.=0.65).

Collegiality also was examined. We expect instruction in schools where collegiality is integral to each teacher’s workweek to benefit from increased teacher interactions. The findings were mixed. Depending on both the schooling level and the type of collegiality assessed (for example, 3 hours or more spent during the past week on “collaborative lesson planning” versus “discussing individual students with other teachers” or some other activity), teachers reported varying degrees of collegiality. To continue the example, teachers at 80% of elementary and middle schools, compared with 33% of high schools, reported collaborative lesson planning. In contrast, 18% of elementary school teachers frequently discussed individual students with other teachers, whereas 29% of middle and high school teachers reported engaging in this type of collegiality.

Academic focus / high expectations and caring together make up what we term a strong professional ethic. Teachers at 79% of schools reported having a strong professional ethic. Reported behavior at levels this high may reflect a socially desirable response rather than actual behavior. While low expectations and the lack of caring may be underreported (only 7 of 34 schools), student performance at these schools is markedly lower, as expected. Whereas 60% of our sample schools met their API award targets, only 17% of the “low professional ethic” schools met their API award target.

Teachers also were asked to report on the extent of their colleagues helping, participatory, and constructive behaviors—all of which comprise *organizational*

citizenship behavior (OCB). The preponderance of schools where teachers exhibit citizenship behavior achieved their performance targets, both for the SAT/9 matched-NCE reading gains (64%) and the API targets (award-85%). Fewer schools with less than half of its faculty exhibiting OCBs attained their performance targets (SAT/9 matched-NCE reading gains-26%, API Award-39%).

Appropriate Use of Student Assessment. Teachers (86%) and principals (97%) overwhelmingly responded that their school had a plan for improvement. We found that a focus on four or more data sources measuring the progress of school goals (e.g., SAT/9 score, Student work samples, Performance Assessment Scores, Disaggregated Student Data, Parent Surveys, and Student Surveys) was associated with an increased likelihood of achieving API performance targets. There were only 7 of 33 schools that made extensive use of several data sources, but these 7 all achieved their API targets.

Site-level Commitment to School Improvement and Professional Development. For the pilot, we were most concerned with what schools are doing to monitor, reflect upon, and improve their own practice. One of the strongest indicators of commitment to school improvement is the presence of a high-quality *instructional monitoring system*. Such a system involves establishing goals, monitoring performance, and providing feedback.

Based on teacher-reports, the instructional monitoring systems were generally stronger at the elementary (\bar{x} =3.77 of 5.00, s.d.=0.62) and middle (\bar{x} =3.67, s.d.=0.41) schools than at the senior high schools (\bar{x} =3.31, s.d.=0.20).

Where instructional monitoring was strong (scoring at least four out of five), API performance followed. In the nine schools with strong instructional monitoring systems, all met their API school goals and eight (89%) received API awards. This compares with an award rate of 50% for schools without strong monitoring systems.

Regarding *professional development*, teachers reported that it was most frequently designed or chosen to support the implementation of state or local standards (\bar{x} =4.20, s.d.=0.56), the district's improvement goals (\bar{x} =4.09,

s.d.=0.54), and the school's improvement goals (\bar{x} =3.99, s.d. =0.59). While teachers clearly saw the connection between professional development and these large-scale improvement efforts, teachers were less certain that professional development was helpful (\bar{x} =3.28, s.d. =0.59), led to changes in classroom practice (\bar{x} =3.28, s.d. =0.55), or was supported by the school community (\bar{x} =3.31, s.d. =0.69).

Indicator 4: Classroom Assignments

Building on recent CRESST research on effective classroom practice, we focused on four separate dimensions of Classroom Assignments that 4th, 7th, and 10th grade language arts teachers assign: cognitive challenge, goal clarity, grading criteria, and overall quality. Our standard for each dimension was a scale score of 2.25 out of 4.00. As with the other indicators, our criterion for successful local district performance is 60% of schools meeting or exceeding the standard.

Analyses by grade level showed that in general the high school assignments received the highest ratings in terms of overall quality, with elementary school following and middle school receiving the lowest ratings. This pattern held true for each dimension. Details about individual dimensions follow.

Cognitive Challenge. This dimension focused on the degree to which students have the opportunity to apply higher order reasoning and to engage with academic content material. Scores on this dimension ranged from 1.75 for the middle school reading assignment to 2.69 for the high school writing assignment. Ratings were assigned on a 4-point scale where 4 was the highest quality.

Goal Clarity. This dimension measured the degree to which an assignment could be considered a purposeful, goal-driven activity focused on student learning. Here, scores ranged from 1.88, for the middle school reading assignment, to 2.46, for the high school reading assignment.

Grading Criteria. This scale captured the degree to which grading criteria are specific and have the potential for helping students improve their performance.

For grading criteria, assignments received ratings between 1.56, for the middle school reading assignment, and 2.46, for the high school writing assignment.

Overall Quality. This score assessed the degree of task quality in terms of level of cognitive challenge, clarity and application of learning goals and grading criteria. These ratings ranged from 1.50 for the middle school writing assignment to 2.28 for the high school reading assignment.

Implications and Future Directions

In this report we present one possible scoring system, an LDPM scorecard, as a useful way to summarize data related to the four indicators. This scorecard is offered only as an example—the actual scoring system is under development. Whatever emerges as the actual scoring system, it will be intended to provide local superintendents with useful feedback as to the performance of schools, teachers and students within their local districts.

Finally, this reports details several modifications to the LDPM design for the 2001-02 school year. Some of these changes are based on what we learned from the pilot administration (e.g., data collection modifications for the SOI and Classroom Assignment indicators), while others reflect changes in the District's adoption of the Principles of Learning.

1 Why the Local District Performance Measures?

The Local District Performance Measures were commissioned by the General Superintendent in May 2000 to track the performance of Local District Superintendents. The system is designed to report on and to influence those elements of the school system that have the greatest influence on student performance. As originally proposed, the system was designed for formative use during the 2000-01 school year, and summative use during the 2001-02 school year.

This system builds upon earlier systems of accountability undertaken by the LAUSD, but departs significantly from the earlier systems in two ways. First, the number of indicators is limited to four. Second, these indicators reflect both direct outcomes of student performance and the school processes expected to increase student performance. The Local District Performance Measures Accountability System is focused on tracking and directing attention toward increasing student performance and the characteristics of school organization that lead to increased performance.

Several experts have contributed to the initial design for this system and the recommended improvements detailed in this report. We have relied extensively on the literature surrounding the issues of accountability, school effectiveness and improvement, indicator design and measurement, and teaching quality. The research basis for the initial design is presented in Appendix B. We have also made extensive use of our university partnerships. One component of our system, which we call “Classroom Assignments,” originated in the work of two researchers affiliated with UCLA’s National Center for Research on Evaluation, Standards and Student Testing (CRESST)—Pamela Aschbacher and Lindsay Clare. Additionally, we have spent considerable time with Lauren Resnick and her colleagues from the University of Pittsburgh’s Institute for Learning refining another system component, the School Organization Index. The refinement, revision, and rigorous testing of our system against world-class standards has been and will continue to be a hallmark of the Local District Performance Measures.

2 A Balanced System for Accountability: Performance and Process Indicators

The Program Evaluation and Research Branch has undertaken a pilot study to test the use of four indicators to assess the performance of Local Districts. It is our aim to gauge the results and actions most salient to student achievement and school effectiveness. The four indicators include direct measures of student achievement, measures of how a school organizes itself for effective instructional delivery, and measures of how individual teachers plan and structure academic tasks for students.

The four indicators are the following:

1. Percent of schools meeting their API growth target
2. Percent of schools meeting their expected matched student reading gain on the SAT/9
3. Percent of schools reaching a satisfactory rating on the School Organization Index
4. Percent of schools reaching a satisfactory rating on the Classroom Assignments

For the pilot study, we have weighted each indicator equally. During this pilot year, the indicator results have been used to test the evaluation and reporting process. (In this report, based on pilot year data and existing standards, we suggest standards for the four indicators.) We have reviewed and revised these standards based on actual scores on the indicators during this formative year. This emphasis on review and revision reflects our belief that a necessary component of any accountability system is the on-going evaluation of its practicality, relevance and accuracy. During the 2001-02 school year the indicators will provide summative data to be used to assess overall local district effectiveness.

2.1 API Growth Target

All underperforming schools in California are expected to meet a 5 % growth target for 2000-01, based on the computational approach embedded in the API. We initially proposed a standard of 60 % of schools in each local district expected to reach this target. This number represents the projected statewide average percentage meeting target growth that was calculated by the Public Schools Accountability Act (PSAA) Technical Design Group, based on a

simulation of 1998-99 SAT/9 data for all schools in California. For the 1999-2000 school year, the actual statewide API award rate was 67%. The API growth target for the district has been adjusted to reflect the 67% statewide rate.

2.2 Matched Student Gains in Reading

Matched student gains (based on each student's test results from SAT/9 Reading in 1999 and 2000) is the second indicator of Local District performance.

Reading was chosen as the single subject area for the Local District Performance Measures in the initial year because reading improvement is currently the focal point of district improvement efforts. We anticipate that the subject matter focus of the student achievement gains component of the accountability system will continue to be consistent with district improvement efforts as directed by the General Superintendent.

For each school, the expected gain target will be calculated using the historic averages for their level of schooling (elementary, middle school, senior high) and 1999 performance grouping (quintile). Our analysis shows that the expected gains are highest for those schools with the lowest initial NCE scores¹ in reading. Increased expectations for lower performing schools, if met, will also serve the District's goal of narrowing the performance gap between the highest and lowest performing schools. Local District performance on this indicator will be based upon the percent of schools that have met or exceeded their expected gain targets (see Appendix A for a detailed discussion of alternative approaches considered for computing matched student gains, and the technical rationale for choosing this particular method).

2.3 School Organization Index

We have preliminarily identified five aspects of school organization that we believe are correlated with school effectiveness: 1) leadership; 2) quality of teaching; 3) school climate and culture; 4) appropriate use of student assessment; and 5) commitment to school improvement and professional

¹ A Normal Curve Equivalent (NCE) score is an equal-interval normalized-standard score with a mean of 50 and a standard deviation of 21.06. NCEs are frequently used in Title I reporting.

development. We regard these measures as process or intermediate outcomes that predict positive student outcomes when these processes are present at optimum levels in a school. We believe that the use of these elements in a local district performance measurement system will encourage schools to attend to these critical organizational features.

In Appendix B we present a detailed conceptualization of each of these school organization elements together with the research that supports selection of each construct. We will comment briefly on aspect #5 here since it includes measures that had a central role in previous accountability programs in the District. As a minimum criterion for evidence that schools are committed to improvement, we will ask schools to track a common list of site-level indicators. These indicators will include: AP enrollment, LEP redesignation rate, dropout rate, student enrollment in academic courses (e.g. high school A-G courses), pass rate for A-G courses, course marks, student attendance, and staff attendance. We expect these indicators to be joined by others that each school deems to be important. Furthermore, while we expect that schools will collect data across all eight of our basic site-level indicators, we do not expect that schools will focus attention simultaneously upon all eight. Instead, we believe it is in each school's interest to focus on a limited number of indicators as it seeks improvement. We expect that schools will begin to explore which of these eight indicators truly results in improved student performance and take action and make decisions according to their findings.

Data on degree of implementation of each School Organization Index construct will be gathered once a year in November through 20-minute telephone interviews of probability samples of certificated staff, parents and students from the schools in each local district.

The sampling procedure for this year's pilot study called for interviews with principals at 35 randomly selected schools (21 elementary, 7 middle, and 7 high schools). In addition to the 35 principals, we sampled 10% of the certificated teaching staff and 1% of the student and parent populations. We contracted with Harris Interactive, a nationally known polling firm, to administer the telephone

survey. Data from the pilot are being used to further develop the specific measures and approach to be utilized in subsequent years, and also undertake the needed technical analysis to develop the composite School Organization Index itself.

2.4 Classroom Assignments

We utilized another intermediate outcome as the fourth performance measure: the nature and quality of work assigned by individual teachers. Building on recent CRESST research, we measured four separate dimensions of Classroom Assignments: challenge of the task (cognitive demand and use of knowledge); clarity of the teacher's goals for student learning; clarity of the grading criteria; and alignment between the goals, tasks, and grading criteria.

As with the School Organization Index, we piloted the evaluation process during the 2000-01 school year to refine the data collection, measures and reporting procedures. During November 2000, we field tested the data collection procedures. All fourth, seventh, and tenth grade language arts teachers in 35 randomly selected schools² across all 11 local districts were asked to supply Classroom Assignments data from three lessons.

Subsequently, we tested scoring procedures in January 2001 with the assistance of an outside contractor (UCLA's Center for Research on Evaluation, Standards, and Student Testing) with the capacity and demonstrated experience to accurately score a large volume of these types of materials within our specified time period.

Our current plan calls for yearly collection of a probability sample of Classroom Assignments in each school during a six-week period beginning in October. The collected assignments will be scored in December/January to enable feedback to the individual schools and each Local District Superintendent by mid-March.

²Of these 35 schools, 15 participated in both the classroom assignments and the school organization index components of the LDPM.

2.5 Further Technical and Operational Considerations

The lead organization for the reporting of these performance indicators for Local District Superintendents will be the Planning, Assessment and Research Division.

Important technical issues have emerged as we have conceptualized and developed each of these performance indicators. First, we currently plan to calculate matched NCE gains for a given school on the basis of the student's presence in a given grade at the time of the posttest, irrespective of the student having been at the same school at the time of pretest. Some might argue that a better approach would be to require that the student had attended the same school at both pretest and posttest. Alternately, we could require that the student be listed as enrolled at the October norming date so as to be legitimately counted in the school's end-of-year results. We have intentionally chosen the current approach (enrollment in *district* at both pretest and posttest) to maximize the number of students for whom we have matched data from spring to spring.

A second technical issue pertains to the relatively low profile given within the Local District Performance Measures to oft-used indicators such as AP enrollment, LEP redesignation rate, dropout rate, enrollment in academic course (e.g. high school A-G courses), pass rates for A-G courses, student attendance, and staff attendance. Our considered judgment is that when these indicators are attached to incentives, they could be subject to manipulation. This in no way diminishes the importance we attach to these indicators. It does, however, change our approach to how they should be included within this accountability system. These indicators are best used as part of each school's self-assessment. It is imperative that each school understands how it serves or does not serve each of its various populations. We encourage the use of the school-level indicators for decision-making at the school site. This is a key aspect of each school's commitment to school improvement and professional development.

Third, we chose to use historical averages for our performance criteria because it is well-established that gains in standardized test scores increase at a decreasing rate following the adoption of a new testing instrument. Maintaining historical growth levels, therefore, represents gains above what we might otherwise anticipate. It is our intent to encourage growth over time. Our current performance criteria are based on an analysis of scores that represent one year of growth in reading achievement, namely the period between Spring 1999 and Spring 2000. Simple multiplication shows that what look to be modest growth targets in the short run become highly ambitious growth targets when forecasted over five years or more.

Fourth, we realize that pressures may be placed on all members of the school to meet identical gain targets at each grade. Such a practice would be technically unsound and violate some key assumptions that underlie our expected gain targets. Our performance expectations are valid only at the school and Local District levels. An attempt to use the same performance targets at individual grade levels would be inconsistent with their design.

Finally, we hope that this accountability system will encourage both the Local District Superintendents and their schools to reflect upon their results as a function of their practices and organizational characteristics. We believe that a sound performance monitoring system has the potential to inform and enrich the ongoing conversation about how all our schools can effectively serve all our students.

3 Local District P— *Demographic Comparisons with LAUSD*

Table 1 compares the characteristics of the student populations for both District P(iilot)³ and LAUSD. The most striking difference between our sample and the district at-large is that the ratio of elementary students to secondary students in District P is nearly the inverse of this same ratio within LAUSD. This is

³Throughout this report, we will refer to our study sample as District P. “P” stands for pilot. The use of the letter to signify a local district mirrors current district practice whereby the eleven local districts are designated A-K. Our intent is to increase the readability of the report and to mimic the comparisons that will be important to future Local District Performance Measures reports.

intentional. Our aim in the pilot was to test our instruments and processes, not to study the LAUSD population.

Except for hosting proportionately more secondary schools, District P closely resembles the demographic characteristics of LAUSD. District P also has slightly more Hispanic students proportionately than LAUSD, most likely due to our sampling of schools based on geographic, rather than demographic, stratification variables.

Table 1. Demographic Comparison of Local District P and LAUSD

	Local District P			Total District		
Elementary Students	29967 (43%)			396035 (65%)		
Secondary Students	39071 (57%)			214460 (35%)		
	EL	SEC	%	EL	SEC	%
Ethnicity						
American Indian	56	113	0.2	1068	689	0.3
Asian	924	1336	3.3	14294	10564	4.1
Black	4062	3768	11.3	47683	30096	12.7
Hispanic	22719	29334	75.4	290186	142581	70.9
White	1783	3500	7.7	35043	24826	9.8
Filipino	382	874	1.8	6512	4928	1.9
Pacific Islander	41	146	0.3	1249	774	0.3
Gender						
Female	14685	18831	48.5	193088	105337	48.9
Male	15282	20240	51.5	202947	109123	51.1
Specially Funded Programs						
Meal Program Participant	26157	29524	80.7	327400	148364	77.9
Title I	23393	24758	69.7	291590	112166	66.1
GATE	136	958	1.6	2809	7904	1.8
Special Ed	2004	1855	5.6	29175	9894	6.4
Language Program						
IFEP	2184	3229	7.8	29790	19961	8.1
LEP	16502	11406	40.4	209245	54697	43.2
RFEP	3147	14272	25.2	35768	70093	17.3
Migrant Ed	150	0	0.2	1407	0	0.2

4 Data Collection • Performance Indicators

The Local District Performance Measures (LDPM) document includes two indicators of student achievement test scores—API Growth Targets and Matched NCE Gains.

4.1 API Growth Targets

To reach the Local District Performance Measures expectation, at least 67% of the schools within a local district must meet API growth targets set by the California Department of Education. In order for a school to receive an award, two different growth targets must be achieved—a schoolwide target and a comparable improvement target. The schoolwide target states that a total school's API score must have grown by 5% of the distance between their baseline API score and the statewide goal of 800. The comparable improvement target specifies that each numerically significant subgroup achieve 80% of the schoolwide growth target.

Table 2 compares the performance of District P schools to the district as a whole on receiving API awards, and on meeting the schoolwide and comparable improvement targets. This table compares 592 LAUSD schools with 52 District P schools. These numbers are lower than the total number of schools in LAUSD and District P for several reasons. First, 32 schools (including 3 District P schools) were excluded from the awards process because insufficient numbers of students were tested. Additionally, there are other schools within LAUSD for which no API score could be calculated. Most likely, these exclusions are due to alternative configurations or lack of test data.

Overall, the schools within District P were less likely to meet state targets and win awards than were other schools across LAUSD. These findings can be attributed to our sample design. In order for us to test the full range of instrumentation, we over-sampled Middle and Senior High Schools. Proportionately, District P has one-quarter fewer elementary schools and twice as many senior high schools as the total district. Because elementary schools are more likely than secondary schools to meet their targets, it is not surprising that District P's awards percentage falls short of the LAUSD averages.

As Table 2 shows, District P fell short of the API Award goal— 58% of its schools met the API award target. A glance across the schooling levels reveals that District P did not exceed the district average at the elementary school level.

Table 2. Number and Percentage of Schools Meeting API Targets by School Level

	Total District	District P
	Met Goal/Target N Schools	Met Goal/Target N Schools
API Award	66% 372 of 560	58% 30 of 52
Elementary	78% 329 of 424	73% 22 of 30
Middle	44% 34 of 78	55% 6 of 11
Senior	15% 9 of 58	18% 2 of 11

4.2 Matched NCE Gains

LDPM guidelines suggest that at least 60% of local district schools meet or exceed their expected gains based on average initial NCE prior to the reorganization. Table 3 compares the performance of District P schools to the district as a whole on matched NCE gains.

In this comparison, District P did not meet its goal of having 60% of schools meeting their expected gains. Elementary and senior high schools roughly match the performance of the total district whereas middle schools exceed that of the total district.

Table 3. Number and Percent of Schools Meeting NCE Gain Expectations by School Level

	Total District	District P
	Met Goal/Target N Schools	Met Goal/Target N Schools
NCE Goal	47% 276 of 592	49% 27 of 55
Elementary	51% 226 of 447	55% 18 of 33
Middle	52% 44 of 85	73% 8 of 11
Senior	10% 6 of 60	9% 1 of 11

**5 Process Indicators •
School Organization Index & Classroom Assignments**

We refer to the School Organization Index and the Classroom Assignments indicators as process indicators since these are not direct measures of student performance. They are, however, indicative of those processes through which schools affect student performance, namely effective school organization and quality teaching. Process indicators represent factors over which schools and local districts have considerable and immediate control. This may not be true of student outcomes. The LDPM focus on process indicators encourages and rewards local districts that achieve those organizational features and exemplary teaching practices that research labels as indicative of successful schooling. The School Organization Index and Classroom Assignments capture these important process oriented results and provide process feedback to complement school test score results.

Regarding the use of this report, we urge caution and recommend that no attempt be made to generalize these findings to the Los Angeles Unified School District. In order for us to fully test the process indicators within the Local District Performance Measures it was necessary to over-sample secondary schools. While the data are real and carefully collected, the distribution of these schools is quite unlike the distribution of schools in LAUSD. We reiterate that District P is a fictional district, and even though it is made up of schools within LAUSD, generalizing these findings to LAUSD is unwarranted.

6 Data Collection • The School Organization Index

6.1 The Instrument

Drawing from the literature on best practices in school reform as well as our own expertise, the SOI team developed instruments for each of four respondent groups—teachers, principals, parents, and students. We asked respondents to report on aspects of school organization hypothesized to influence a school’s ability to function effectively. Our contractor, Harris Interactive, conducted 20-minute telephone interviews with respondents in each group. This data collection took place during November 2000 and the first week of December 2000.

6.2 The Sample

Thirty-five of the 55 District P schools were selected to pilot the School Organization Index— 21 elementary schools, 7 middle schools and 7 high schools. We surveyed the principal, 10% of the teachers, 1% of the parents and 1% of the students (secondary schools only) at each school. We selected each stakeholder group independently, except for students and parents, who are naturally linked. Table 4 presents the initial response rates for each participant group.

These response rates fell short of our goals. We established as our goal a 70% response rate from school-based personnel and a 50% response rate for parents, expecting some further attrition to the response rate for students. A 50% response rate is sufficient to eliminate most non-response bias, however some researchers advocate a more conservative 70% response rate to establish the generalizability of the collected data.

Table 4. Harris Interactive Data Collection November - December 2000

	Respondents	Completed Cases	Response Rate
Teachers	268	103	39%
Principals	35	24	69%
Parents	546	151	28%
Students	330	52	16%

Upon receiving these results from Harris Interactive, we considered whether we had sufficient data to make generalizable claims for each group of respondents and decided to undertake a second round of data collection to increase our response rates. This second round enabled us to meet our response rate goals for 3 of 4 respondent groups— teachers (70%), principals (86%), and parents (50%). Only students (29%) fell short of our response rate target (see Table 5).

Table 5. Overall Data Collection November 2000 – April 2001

	Respondents	Completed Cases	Response Rate
Teachers	268	188	70%
Principals	35	30	86%
Parents	546	270	50%
Students	330	95	29%

6.3 The Protocols

Teachers

Harris Interactive surveyed teachers by phone. Whenever the respondent could not be contacted directly by telephone at the school site, the interviewer left a message asking the teacher to call a toll-free number to conduct the interview at his or her convenience. Harris Interactive averaged four phone calls to the

school site for the potential respondents. Although fewer than 3% of the teachers refused to participate, it was difficult for Harris Interactive to contact teachers.

To increase our response rate with teachers, we mailed questionnaires to all non-respondent teachers. We timed this mailing to arrive just after the winter break, hoping that teachers might find this a convenient time to complete and return the questionnaires. Approximately one week after teachers received the questionnaires, we sent reminder postcards. Four weeks after the initial mailing, we sent follow-up letters to all remaining non-respondents. In this follow-up letter, we asked teachers to notify us if they were not willing to complete the questionnaire.

Principals

Harris Interactive used a calling protocol for school principals similar to that used for the teachers. Each principal received at least 2 calls with an average of 3.7 calls per school principal. No principal refused.

Our first round of data collection yielded a 69% response rate. While this rate was acceptable, the distribution of principals was unbalanced. We received a much higher response rate for the elementary principals than for the middle and high school principals. We were concerned that this uneven distribution would bias the sample.

Our second round of data collection focused on increasing the number of participating principals, especially at the secondary level. We made additional telephone calls from our office to non-respondent principals. These took place from January through March.

Parents

Using student names and home telephone numbers, Harris Interactive called and asked to speak with the parent or guardian most familiar with the selected student's school experiences. The parent was asked to answer questions based upon his or her personal experience with the sampled student's current school. Harris conducted interviews in both Spanish and English.

Harris Interactive telephoned each family at least once and averaged 4.1 calls. They did not contact nor try to find any additional information for parents whose telephone numbers were incomplete or inaccurate (2.6%). Harris Interactive did not offer parents a toll-free number, but they did accept appointment times if the parents wanted to schedule an interview. Nearly 10% of parents refused to complete the interview. Harris achieved a response rate of 28.3% by early December.

Our second round of data collection also used telephone interviews. PERB hired interviewers who conducted interviews weekday evenings and weekend afternoons. Due to limited resources, we restricted our sample and focused our efforts on a limited, but still representative sample of schools. This effort was successful. In the restricted sample, our parent response rate was over 70%. This increased the response rate of the full sample to 50%.

Students

At the end of each parent interview, the interviewer asked parents of secondary students to permit their son/daughter to be interviewed. If the student was currently available, the interview was conducted immediately. This happened less frequently than we had hoped. Fewer student interviews were completed because they were dependent upon 1) a successful contact with a parent, 2) parental consent, and 3) availability of the student. By early December, only 15.8% of the eligible students had completed interviews.

Our follow-up efforts for students were linked to the parent efforts. No students refused to participate. In two cases, a parent refused to offer consent. Students seemed quite willing to do the interview. The primary difficulty in completing student interviews was finding secondary students at home when they had time to be on the telephone for twenty minutes.

7 Measures and Results • School Organization Index

This overview provides both the conceptual map and the results of the School Organization Index survey administered to students, parents, teachers and school administrators during the Fall 2000 semester. Results, where applicable, follow the description of each domain or construct.

7.1 Domain 1: Leadership

School effectiveness research is best served by an expanded conception of leadership that includes staff (i.e. administrators and teachers) and guidance systems (i.e. goals, visions). These aspects are interconnected and are often, but not always, a reflection of choices made at the top of the organizational hierarchy. The three constructs measured within this domain were instructional leadership, trust, and shared governance.

Technical Note: The measurement model for leadership exhibited strong factorial validity. We combined several items within each scale to form two latent variables—instructional leadership and trust. These latent variables were created as first-order factors, which served as indicators for the second-order factor that we called leadership. Confirmatory factor analysis (CFA) results indicate a model with acceptable fit. On all fit indexes, the model exceeded the criteria for acceptable fit (Comparative Fit Index (CFI) and Goodness of Fit Index (GFI) above 0.90; Root Mean Residual below 0.05). Additionally, both scales were internally consistent—instructional leadership (Parents (P) $\alpha = 0.81$; Teachers (T) $\alpha = 0.88$) and trust (Parents (P) $\alpha = 0.84$; Teachers (T) $\alpha = 0.94$; Students (S) $\alpha = 0.82$; Administrators (A) $\alpha = 0.82$).

We found a strong relationship between leadership and API performance (see Table 6). Of those District P schools receiving an API award, 80% (16 of 20) were led by principals who scored at or above 4.00 on our leadership measure which ranged from 1 (low) to 5 (high). Conversely, of those schools not receiving an API award, 66.7% (8 of 12) were led by principals who scored below 4.00 on our leadership measure.

Table 6. Number and Percent of Schools Meeting Performance Goals by Leadership Level

		Leadership	
		LOW (<4)	HIGH (≥4)
Met Target?			
API Award	YES	4 (33%)	16 (80%)
	NO	8 (67%)	4 (20%)
API Schoolwide Target	YES	5 (42%)	17 (85%)
	NO	7 (58%)	3 (15%)
API Comparable Improvement Target	YES	5 (42%)	16 (80%)
	NO	7 (58%)	4 (20%)

7.1.1 Instructional Leadership

Instructional leadership refers to education-specific knowledge held by the principal. Although other individuals at the school site may provide significant instructional leadership, we intentionally focus on the principal. Instructional leadership includes knowledge of pedagogy, content standards, and specific knowledge of individual teacher’s practice. Examples of items (presented to teachers) include, “Your principal has accurate knowledge of your classroom practice” and “Your principal has influenced you to refine your teaching methods to improve student performance.”

Teachers in 16 schools out of 34 (47.1%) report their principal to be an effective instructional leader. Interestingly, teachers in senior high schools rated their principals lower on this dimension (\bar{x} =3.33, s.d.=0.48) than did teachers in middle schools (\bar{x} =3.86, s.d.=0.22) or elementary schools (\bar{x} =4.04, s.d.=0.56). All scores are based on a five point scale, five representing the high end of the continuum.

Principals' self-reports indicated a higher level of instructional leadership ($\bar{x}=4.35$, $s.d.=0.48$) than that reported by either teachers or parents. The principals' self-report corresponded with neither teachers' ($r=0.21$, n.s.), nor parents' ($r=0.03$, n.s.), assessments of the principal as an instructional leader. When assessing a principal's instructional leadership, parents and teachers had the highest levels of agreement ($r=0.42$, $p<0.01$). Parents' assessment of school leadership ($\bar{x}=4.15$, $s.d.=0.45$), however, was slightly more generous than that of the teachers ($\bar{x}=3.85$, $s.d.=0.55$).

7.1.2 Trust

We utilized a conceptualization of trust that has three bases: competence, consistency and benevolence (Mayer, Davis and Schoonhoven, 1995). Again, we focus on the role of the principal. Trust is made difficult if the principal cannot advise on matters important to instructional practice. Trust is undermined when principals are perceived to be unreliable, biased, or unfair, especially in personnel matters. Trust is buttressed when those whom the principal serves believe that the principal truly seeks what is good for them, good for the school, and good for students. The first basis of trust, competence, is represented in the instructional leadership construct. The other two bases for trust, consistency and benevolence, are addressed in this section. Items administered to teachers include, "Your principal wants to be known as someone who keeps promises and commitments" and "You can usually trust your principal to do what is good for you."

Teachers in 23 schools out of 35 (65.7%) report their principal to be trustworthy. As with instructional leadership, teachers in senior high schools rated their principals lower on this dimension ($\bar{x}=3.72$, $s.d.=0.35$) than did teachers at middle schools ($\bar{x}=4.22$, $s.d.=0.26$) or elementary schools ($\bar{x}=4.14$, $s.d.=0.56$).

Principals' self-reports indicated a higher level of trustworthiness ($\bar{x}=4.49$, $s.d.=0.42$). Although teachers reported that principals were generally trustworthy ($\bar{x}=4.07$, $s.d.=0.50$), teacher reports correlated only weakly with

principals' self-reports of trust characteristics ($r=0.19$; n.s.). Parents' assessment of the principal's trustworthiness was more generous ($\bar{x}=4.29$, s.d.=0.50).

7.1.3 Shared Governance

Within the literature, the effects of shared governance are equivocal. There is no clear body of research stating that increased shared governance has a positive impact on student performance. One possible explanation is that, unlike instructional leadership or trust where there is no clear upper bound, it may be possible to have too much shared governance. In order to test several hypotheses concerning shared governance, we asked questions about several discrete areas where shared governance may be considered at the school site. We also asked questions of several stakeholder groups in order to compare perceptions about governance across this range of activity. For example, teachers and parents were both asked about the extent to which, "The school administration encourages parents in the community to take part in developing school policies."

Shared governance did not correlate with either instructional leadership or trust. For this reason, it was excluded from our leadership latent variable. Even so, the data provided some insights regarding perceived power relationships within District P.

First, principals in District P see themselves as relatively powerful within the school environment ($\bar{x}=4.28$ out of 5.00, s.d.=0.44). Indeed, principals rank themselves as having more policy-making influence than students ($\bar{x}=2.51$, s.d.=1.16), parents ($\bar{x}=2.75$, s.d.=0.44), local district administrators ($\bar{x}=3.00$, s.d.=0.92), central district administrators ($\bar{x}=3.19$, s.d.=0.68) and teachers ($\bar{x}=3.88$, s.d.=0.65).

Second, we examined decisions concerning professional development activities as strong indicators of a school's priorities and the power(s) that dictate these priorities. Both administrators and teachers reported that the following were primary drivers of professional development at their school—standards

(principals: \bar{x} =3.66 of 5, s.d.=0.60; teachers: \bar{x} =3.42, s.d.=0.40), the school improvement plan (\bar{x} =3.22, s.d.=0.48), district-level initiatives (principals: \bar{x} =3.22, s.d.=0.79) and current problems (teachers: \bar{x} =3.10, s.d.=0.56). Even though principals view themselves as having power over several critical functions at their schools, they view their own preferences (\bar{x} = 1.72, s.d.=0.44) as the least important factor determining the professional development agenda. Teachers, likewise, saw their own preferences as the least important determinant of professional development planning (\bar{x} =2.73, s.d.=0.56).

7.2 Domain 2: Quality of Teaching

Few doubt the importance of quality teaching to school effectiveness. School effectiveness researchers draw heavily, if selectively, upon Brophy and Good's (1986) influential review of effective teacher traits. Our survey highlights four aspects of quality teaching: task and skill emphasis, classroom effectiveness, grading practice and self-efficacy.

7.2.1 Task and Skill Emphasis

The question underlying this construct is simply "How is learning structured and teaching carried out in classrooms within the district?" We ask teachers and their students to report on the instructional strategies used within the classroom during the past week. For example, we asked students to reflect upon their last English, math, science, and social science classes and report whether they were passive or active participants in the classroom. The prompts ranged from "You mostly listened" to "You made your own predictions, guesses, or hypotheses."

It appears that the lecture has given way to the class discussion. Less than one-quarter of the teachers reported that they had lectured to the class for more than half of the period during the five previous school days. Fifty-eight percent of the teachers reported using a mixture of brief presentations, followed by questions, answers and discussion. Other popular strategies included having students explain their reasoning (56%) and practicing study skills (55%). Indeed, these top three strategies were ranked identically across all levels of schooling.

Student reports of what goes on in the classroom differed between middle school and senior high school students. Students were asked to report the mode of instruction for their last class period across four different subjects. We then aggregated these responses to construct a composite measure of instructional practices. At the middle school level, 8 of 15 modes of instruction occurred in 80% of the students' classes. Middle school student reports largely confirmed the teacher reports. Students said that teachers most frequently “used models or examples” and “asked questions” (both at 87%). Rounding out the top five modes were “I took part in discussion” (84%), “I mostly listened” (84%), and “I asked questions” (84%).

At the senior high level, students reported that they “completed written assignments from the textbook” (82%) and “the teacher asked questions” (78%) during their classes. Others in the top five include “I mostly listened,” “I asked questions,” and “I followed step-by-step instructions” (all at 75%).

In general, teachers report modes of instruction that require students to be active learners. Students at the middle school generally confirm the teacher self-reports. At the senior high school, however, students report more passive forms of instruction. At both schooling levels, it was striking how seldom students were asked to connect what they were learning to the lives they live: Forty-three percent of students reported no instances of relating their learning to their lives during their past four core subject classes.

7.2.2 Classroom Effectiveness

It is difficult to assess classroom effectiveness without observing actual classroom practice⁴. The classroom effectiveness construct measures whether various stakeholders view teaching practices at a given school as effective. We asked principals to name the areas of teaching that they are currently focusing on and to rate how effective their teachers are in those areas of focus. We also

⁴ Another component of the Local District Performance Measures addresses this complicated task using a methodology far superior to surveys—refer to the Classroom Assignments section of this report for a detailed analysis of the quality of teaching practice within District P.

asked parents to report whether they believe that their children's teachers have the ability, motivation and training necessary to help their children learn. For example, "Your child's teachers have the ability to help your child learn."

Principals report that their primary emphasis is facilitating each student's progress in language and number skills (\bar{x} =3.94 of 4.00, s.d. = 0.25). This is followed closely by raising average performance on standardized tests (\bar{x} =3.72, s.d.=0.46). Not coincidentally, these are also the two areas where principals rate teachers as being the least successful (respectively, \bar{x} =2.94, s.d.=0.76; \bar{x} =2.97, s.d.=0.93).

Parents report that they are generally confident about how well teachers are prepared to help their child learn. Parents rated teachers as having sufficient ability (\bar{x} =4.08 of 5.00, s.d. = 0.25), resources (\bar{x} =3.98, s.d.=.053), and motivation (\bar{x} =3.89, s.d.=0.57). These elements were strongly correlated with each other.

7.2.3 Grading Practice

For grading practices we asked teachers which criteria they used to grade student work. Teachers reported that they rely most heavily upon individual improvement or progress over past performance (\bar{x} =3.53 of 4.00, s.d. 0.31). This is followed by effort (\bar{x} =3.50, s.d. 0.33), absolute level of achievement (\bar{x} =3.07, s.d. 0.43), and achievement relative to the rest of the class (\bar{x} =2.83, s.d. 0.52). As LAUSD considers itself a standards-based education system, it is curious that absolute level of achievement ranks below two of the three more relative bases for grades.

With this question in mind, we looked at the relationship between standards-based grading and student performance. Using 3.25 (out of 4) as our cutoff point, we looked at the performance of schools whose teachers reported using absolute level of achievement to determine student grades. We define these schools as having practiced standards-based grading. Only 9 of 35 schools met our 3.25 criterion for standards-based grading. Of these 9, 6 met their NCE

matched reading target (67%). This compares with a 35% NCE target attainment rate for schools that do not employ standards-based grading. Our examination of performance on the API provided similarly striking patterns. Eighty-five percent of the standards-based grading schools attained API goals and awards, compared to 54% of those schools not employing standards-based grading.

Table 7. Number and Percent of Schools Attaining an API Award by Level of Importance of Standards-based Grading

		Standards-based Grading	
		LOW ($\bar{x} < 3.25$)	HIGH ($\bar{x} \geq 3.25$)
API Award	Met Target?		
	YES	14 (54%)	6 (86%)
	NO	12 (46%)	1 (14%)

In addition, we asked students how often their teachers provided them with feedback on their work, how promptly teachers returned assignments, and whether students could anticipate the grade they received from teachers. We found that “most of the time” teachers returned students’ work with corrections and that the completed work was discussed. Students would correct one another’s work in class a little more than “half of the time.” Additionally, 80% of District P students report that they know what their grade will be before their report card is released.

7.2.4 Self-Efficacy

Self-efficacy is another perceptual measure of teaching effectiveness, this time from the perspective of the teacher. These questions asked teachers about the extent to which they were generally effective at teaching their students and had specific competence in several areas related to instruction, such as knowledge of standards and child development, facility with techniques, curriculum-development expertise, and the ability to reflect on their practice. An example

of an efficacy item asked of teachers is “Please indicate whether you have a strong working knowledge of methods to assess students.”

Teachers’ self-assessment matched the confidence parents have in their child’s teacher’s effectiveness (section 7.2.2). On our 5-point scale, all groups of teachers reported that they are effective at teaching their current students—elementary teachers (\bar{x} =4.39), middle school teachers (\bar{x} =4.36), and senior high teachers (\bar{x} =4.00).

7.3 Domain 3: School Climate and Culture

The ideal school climate, as captured by school effectiveness studies, is safe, orderly, welcoming to parents, free from conflict, and achievement oriented. Several studies have noted the apparent requirement of a safe, orderly climate for teaching and learning to be effective (Edmonds, 1979; Levine & Lezotte, 1990; Rutter et al., 1979; Sammons et al., 1995; Teddlie & Stringfield, 1993). Also, the association of effectiveness with a “salient” level of parental involvement (Levine, 1992) recognizes the benefits of parental involvement (Mortimore et al., 1988) without the presumably negative consequences of ceding complete control to lay persons. Third, teaching and learning are personal and relational processes. Conflict between students and adults or between adults at the school site threatens the establishment of meaningful relationships necessary to productive teaching and learning (Goodlad, 1984; Lightfoot, 1983). Finally, “a press for achievement” commonly distinguishes the more effective schools (Oakes, 1989). As with high expectations, an achievement orientation may motivate students and teachers toward increased performance. The constructs used within our survey instruments reveal a similar logic: safe and orderly campus, collegiality, caring, academic focus/high expectations, and organizational citizenship behavior.

7.3.1 Safe and Orderly Campus

A safe and orderly campus does not likely contribute to student achievement so much as its opposite, a dangerous and chaotic campus, impedes student achievement. Serious learning requires sustained, concentrated effort: Fear and

disorder thwart learning. Our questions were inclusive of general safety, maintenance, administrative support for a disciplined environment, student adherence to school rules, and specific violations of law. We asked all respondents whether they feel “The school is a safe place” and whether “Teachers deal effectively with most disruptions or behavior problems in the classroom.”

The essential question, “is this school a safe place?” was answered in the affirmative by all respondent groups. Administrators are the most positive about school safety (\bar{x} =4.59 of 5.00, s.d.=0.49), followed by teachers (\bar{x} =4.30, s.d.=0.57), and then parents (\bar{x} =3.67, s.d.=0.54). Of all groups, students are most wary about school safety (\bar{x} =3.49, s.d.=0.65). Parents, teachers and students are increasingly less optimistic about school safety as the schooling level increases⁵. Administrators, however, report higher levels of safety at the senior high level.

7.3.2 Collegiality

We expect instruction in schools where collegiality is integral to each teacher’s workweek to benefit from increased teacher interactions. We asked questions about recent activity where teachers collaborated on lesson planning, curriculum development, and strategizing instruction for their students. Teachers reported having spent 3 or more hours in collaborative lesson planning during the past week at 80% of elementary and middle schools, compared with 33% of high schools. Teachers reported having spent 3 or more hours in curriculum development “during the past week” at 44% of elementary schools, 50% of middle schools, and 17% of high schools. Additionally, teachers reported having spent 3 or more hours discussing individual students with other teachers during the past week at 18% of elementary schools, and 29% of middle and high schools.

⁵ The correlation between teacher and parent ratings, $r=0.46$, was significant at 0.01 level. The correlation between administrators and the other respondent groups was less than 0.06 and non-significant.

7.3.3 Caring

Caring for students is considered part of the professional ethic for teachers (McLaughlin and Talbert 1994). We assessed the level of caring at a school through questions about teacher behaviors, and student relationships with adults at the school and immediate environment. For example, we asked students whether they agree, “Most students at this school have a personal relationship with at least one adult in the school.” Caring has been combined with Academic Focus/High Expectations to form a single construct called professional ethic. The results are reported in the next section.

7.3.4 Academic Focus / High Expectations

Achievement orientation, also part of the professional ethic for teachers, is a strong predictor of effective schooling. We asked questions about the academic expectations of students held by multiple stakeholders. It is possible that the high expectations need to be consistent across school, class, and home environments to have the desired impact on student performance. At the school level, our questions targeted general expectations, the organization of time and the delivery of courses as indicative of the level of expectations about student learning capacity. At the class level, we asked teachers how high their expectations are and whether their students are capable of meeting standards. We asked parents and students questions about their future plans, the importance of learning, and whether they believe the school is equipped to help their student achieve academically. In one question, we asked teachers to what extent they agree, “There is really very little you can do to ensure that most of your students achieve at a high level.”

Technical Note: The measurement model for professional ethic exhibited strong factorial validity. We combined several items within each scale to form two latent variables—academic focus and caring. These latent variables were created as first-order factors, which served as indicators for the second-order factor that we called professional ethic. Confirmatory factor analysis (CFA) results indicate a model with acceptable fit. On all fit indexes, the model exceeded the criteria for acceptable fit (Comparative Fit Index (CFI) and Goodness of Fit Index (GFI) above 0.90; Root Mean Residual below 0.05). Additionally, the scales were internally consistent for academic focus (Parents (P) $\alpha = 0.68$; Teachers (T) $\alpha = 0.88$; Administrators (A) $\alpha = 0.76$) but not for caring (Parents (P) $\alpha = 0.51$; Teachers (T) $\alpha = 0.50$; Students (S) $\alpha = 0.36$; Administrators (A) $\alpha = 0.46$).

Teachers at 79% of schools reported having a strong professional ethic—high expectations of and a caring disposition toward their students. Reported behavior at levels this high may reflect a socially desirable response rather than actual behavior. While low expectations and the lack of caring may be underreported (only 7 of 34 schools), student performance at these schools is markedly lower, as expected. Whereas 60% of our sample schools met their API award targets, only 17% of the “low professional ethic” schools met their API award target.

7.3.5 Organizational Citizenship Behavior

Organizational citizenship behavior (OCB) is a measure of extra-role activity (Podsakoff & Mackenzie, 1994). Teachers were asked to report on the extent of their colleagues helping, participatory, and constructive behaviors. OCB is likely to correlate strongly with commitment to school goals since pursuit of such goals often requires effort beyond the pursuit of individual classroom goals. These behaviors go beyond formal job descriptions. They indicate a supportive environment and a willingness to go above and beyond the call of duty for the sake of the school, its goals, and the students it serves. For example, we asked teachers what portion of their school faculty “Willingly gives of his or her time to help other teachers who have work-related problems.”

Technical Note: The measurement model for organizational citizenship behavior (OCB) exhibited strong factorial validity. We combined several items within each scale to form three latent variables—helping behavior, participatory behavior, and constructive behavior. These latent variables were created as first-order factors, which served as indicators for the second-order factor that we called school climate and culture. Confirmatory factor analysis (CFA) results indicate a model with acceptable fit. On all fit indexes, the model exceeded the criteria for acceptable fit (Comparative Fit Index (CFI) and Goodness of Fit Index (GFI) above 0.90; Root Mean Residual below 0.05). Additionally, the entire OCB scale was internally consistent (Teachers (T) $\alpha = 0.89$, 14-items).

Teachers at 42% of schools reported that at least one-half of their faculty exhibited behaviors that typify organizational citizenship. The preponderance of schools where teachers exhibit citizenship behavior achieved their performance targets, both for the SAT/9 matched-NCE reading gains (64%) and the API targets (award-85%). Fewer schools with less than half of its faculty exhibiting OCBs attained their performance targets (SAT/9 matched-NCE reading gains- 26%, API Award- 39%).

Table 8. Number and Percent of Schools Meeting Performance Goals by OCB

		Organizational Citizenship Behavior	
		LOW (<1/2)	HIGH (\geq 1/2)
Met Target?			
API Award	YES	7 (39%)	11 (85%)
	NO	11 (61%)	2 (15%)
NCE Target	YES	5 (26%)	9 (64%)
	NO	14 (74%)	5 (36%)

7.4 Domain 4: Appropriate Use of Student Assessment

Frequent monitoring of student progress is often cited as a correlate of effectiveness (Edmonds, 1979; Teddlie & Stringfield, 1993). Some cite “appropriate monitoring” (Levine, 1992), presumably to guard against those cases where frequent monitoring interferes significantly with the time allocated to learning. Others note that effective schools are more likely to pay attention to

disaggregated test results (Reynolds & Packer, 1992), suggesting that the quantity of monitoring matters less than the effective use of test results. Our survey instruments probe the use of student assessment data to inform decisions at both the school and teacher level.

7.4.1 Data-driven decision-making at the school level

Our survey focused on the uses of data at the school site. First, we asked teachers and principals if their school had a plan for improvement. Teachers (86%) and principals (97%) overwhelmingly responded that this was the case.

Teachers reported that the following data sources are used to measure progress toward goals set forth in their school plan: SAT/9 Scores (55% of schools); Student work samples (53%); Performance Assessment Scores (51%); Disaggregated Student Data (41%); Parent Surveys (37%); and Student Surveys (32%). We found that a focus on four or more of these data sources is associated with an increased likelihood of achieving API performance targets. There are only 7 of 33 schools that made extensive use of several data sources, but these 7 all achieved their API targets.

7.4.2 Data-driven decision-making at the teacher level

At the teacher level, we focused on how teachers use different sources of data to inform aspects of their daily practice. We used the term “data” in its broadest terms. Certainly, we included SAT/9 scores, but we also considered performance assessments, student grades, student work samples, satisfaction surveys, principal evaluations, peer reviews and any other form of data teachers use to inform decision making.

Ninety percent of teachers reported using student test score data to assess areas where they need to strengthen their content knowledge and/or pedagogy and to adjust their curriculum in areas where students encountered problems. Fewer than half the teachers (47%) report having used student test score data to group students into different instructional groups. Not surprisingly, teachers used student grades (\bar{x} =3.88 of 5.00, s.d.=1.0), above all other data sources, to inform

how they teach. All teachers were also likely to use performance assessment scores (\bar{x} =3.52, s.d.=1.13). Only elementary school teachers made frequent use of SAT/9 scores (\bar{x} =3.48, s.d.=1.3). Senior high teachers appeared to make less use of any type of feedback, be it from parent or student surveys (\bar{x} =2.30, s.d.=1.2); principal evaluations (\bar{x} =2.24, s.d.=1.26); or peer evaluations (\bar{x} =2.48, s.d.=1.2).

7.5 Domain 5: Site-Level Commitment to School Improvement and Professional Development

School improvement refers to improvement efforts developed, modified and/or embraced at the site-level. This does not include state or district attempts to coerce schools to change, although it does not preclude the state or district from encouraging or developing a school's interest in improvement. Site-level professional development (in-service training) serves as an indicator of a school's autonomous commitment to improvement (Levine & Lezotte, 1990). Second, effective improvement efforts are those that focus on classroom practice (Reynolds & Packer, 1992) as opposed to schedules or relationship maintenance. Third, effective schools track their improvement using a limited number of indicators (Peters, 1987) such as pupil academic progress, pupil satisfaction, and the presence of a meaningful pupil/teacher relationship (Gray & Wilcox, 1995).

For this pilot, we were most concerned with what schools are doing to monitor, reflect upon, and improve their own practice. As this project scales up from pilot to full implementation, we will ask schools to describe the indicators they use to determine whether or not their practices are effective. Tracking a few key indicators will be a minimum criterion for evidence that schools are committed to improvement.

7.5.1 Presence of an Instructional Monitoring System

One of the strongest indicators of commitment to school improvement is the presence of a high-quality instructional monitoring system. Such a system involves establishing goals, monitoring performance, and providing feedback.

7.5.2 Teacher Trust in Instructional Monitoring System

Teachers should see all school policies, but especially the instructional monitoring system, as contributing to the education of students. Otherwise, the policy has the unintended effect of encouraging behaviors aimed at manipulating the system rather than improving teaching and learning. Examples of this type of question include, “Established procedures prevent me from doing what I think best serves the needs of students” and “Teachers are encouraged to suggest more effective policies.”

The performance monitoring systems were generally stronger at the elementary (\bar{x} =3.77 of 5.00, s.d.=0.62) and middle schools (\bar{x} =3.67, s.d.=0.41) than at the senior high schools (\bar{x} =3.31, s.d.=0.20). Trust in the instructional monitoring system was strongly correlated with strength of the system (r =0.75, p <0.01).

Where instructional monitoring was strong (scoring at least four out of five), API performance followed. In the nine schools with strong instructional monitoring systems, all (100%) met their API school goals and eight (89%) received API awards. This compares with an award rate of 50% for schools without strong monitoring systems.

Table 9. Number and Percent of Schools Meeting Performance Goals by Presence of an Instructional Monitoring System

		Instructional Monitoring System	
		LOW (<4)	HIGH (≥4)
API Award	Met Target?		
	YES	12 (50%)	8 (89%)
	NO	12 (50%)	1 (11%)

7.5.3 School Goals

School goals overlap with constructs discussed earlier in this overview. They are the explicit manifestation of a school’s high expectations for students.

Furthermore, they serve to guide what is expected both of students and teachers in terms of focus and performance.

7.5.4 Professional Development toward School Goals and Best Practices

Consistent with district practice, we believe that professional development time is a valuable resource and should be used solely for purposes that advance student achievement. Our questions target which class of improvement goals the professional development is designed to address. Additionally, we are interested to learn who planned and presented professional development sessions, whether follow-up occurred, and the impact of professional development upon classroom practice.

Teachers reported that professional development was most frequently designed or chosen to support the implementation of state or local standards ($\bar{x}=4.20$, s.d.=0.56), the district's improvement goals ($\bar{x}=4.09$, s.d.=0.54), and the school's improvement goals ($\bar{x}=3.99$, s.d. =0.59). While teachers clearly saw the connection between professional development and these large-scale improvement efforts, teachers were less certain that professional development was helpful ($\bar{x}=3.28$, s.d. =0.59), led to changes in classroom practice ($\bar{x}=3.28$, s.d. =0.55), or was supported by the school community ($\bar{x}=3.31$, s.d. =0.69).

8 Data Collection • The Classroom Assignments

8.1 The Instruments

Classroom assignment packets. In early November 2000, selected language arts teachers were asked to submit samples of their reading and writing assignments. Classroom assignment (CA) packets containing materials for submitting these assignments were sent to teachers via school mail. Teachers were asked to select a writing assignment and two reading assignments. They were then asked to describe these assignments using a sheet calling for information about the reading material used for the assignment, the unit into which the assignment fit, the learning goals of the task, and the teacher's grading criteria. Teachers were also asked to attach four samples of their students' work. The information sheet,

or “cover sheet” was based on the cover sheets developed at UCLA’s National Center for Research on Evaluation, Standards, and Student Testing (CRESST) (Clare et. al, 2001).

Observations. Observations were conducted in order to obtain validity data for classroom assignment measures and test the assumption that personal contact encourages teacher submission of assignments. A set of 4-point scales developed at CRESST (Clare et. al, 2001) was used to rate the quality of observed instruction. The scales paralleled those used to rate the assignments and specifically measured the cognitive challenge posed to students by the lesson’s activities, the clarity of the teacher’s learning goals, alignment of goals and tasks, the quality of the feedback that the teacher provided to students about their progress toward the learning goals, the quality of the teacher’s implementation of the lesson (essentially classroom management), and the level of participation of students in the learning activities. In addition to these analytic scales, researchers also assigned a holistic score factoring in all of the above dimensions.

Teachers were observed once for approximately one hour during a language arts lesson. Six researchers trained in the use of the observation instrument (described below) conducted the observations. Observations took place within a four-week span during November and December 2000. Teachers received a memo in their CA packets notifying them that they had been randomly selected for the study and that they would be contacted within two weeks following their receipt of the memo.

8.2 The Sample

Classroom assignments. The classroom assignment component of the study involved 35 randomly selected schools from across the 11 local districts. Each of the fourth, seventh, and tenth grade language arts teachers at these schools were asked to submit samples of their classroom assignments. A total of 193 packets were mailed to teachers, but it was eventually determined that no more than 181 of these teachers were eligible to participate. (Twelve teachers notified us that they had been misidentified either because they did not teach students

from the targeted grade levels or because they were off-track at the time of data collection.) Fifty teachers submitted the requested assignments for a total of 139 assignments.

Observations. The design of the study called for observations of teachers at 15 of the 35 CA schools. Fifty teachers from across these 15 schools were randomly chosen for observation. Because of scheduling problems, 40 teachers were ultimately observed.

8.2.1 Efforts to Increase Participation

In order to increase teacher participation, various techniques were employed to encourage teachers to submit their assignments.

Reminder Letters

All of the 181 teachers were mailed letters every other week for six weeks reminding them of the upcoming deadline for submission of their materials. One of the letters informed them that they would receive \$75 to be used to purchase instructional materials for their classrooms as compensation for their time. Another letter informed them that they would be given time beyond the deadline to submit their packets.

Observations

Observations were conducted in part to test the assumption that personal contact encourages teacher submission of assignments. Among the 40 observed teachers, there were 21 who were exclusively observed. That is, they received no other treatment such as phone support.

Telephone Support

Roughly a third of the CA teachers were targeted to receive phone calls to provide information and support, and to monitor their progress toward completion of the packets. Fifty-four of these teachers were eventually contacted. Thirty-five of the teachers received telephone support exclusively. The remaining 19 teachers received phone support and were also observed.

A part-time, temporary employee was oriented on the goals of the CA component, the contents of the CA packets, and the types of questions that concerned teachers might ask about the CA data collection effort. The goals of this “phone monitor” were to provide information, to prompt and encourage teachers to submit their samples, and to note teachers’ reactions to the project. The phone monitor worked from her home and called schools before school, during lunchtime, and after school. She left messages for teachers, making herself available for support. For some teachers, whose schools had voice mail systems, the phone monitor was able to leave detailed messages about the project. In some cases, the phone monitor was able to speak with teachers directly. In all, the phone monitor spoke with 20 teachers. The remaining teachers received messages.

In-person Reminders

After the deadline for submission of CA packets, 64 teachers received in-person visits in their classrooms by researchers who reminded them that they had not turned in packets and who also asked them for feedback about what had kept them from submitting the packets. The schools of these teachers were randomly selected, and all of the teachers in the schools were visited in their classrooms (if they had not yet submitted a packet).

Table 10 shows the response rates of the teachers by the type of follow-up treatment they received. The largest proportion of teachers responded to a combination of telephone support and observation. “Telephone support exclusively” also yielded a high response rate.

TABLE 10. Numbers of Teachers Targeted and Respondents

Teacher Samples	Targeted N	N of Respondents	% Response
CA packets	181	52	29
Observation	40	16	40
Observation exclusively	21	5	24
Observation and phone support	19	12	63
Telephone support	64	28	52
Telephone support exclusively	35	16	46
Telephone support and Observation	19	12	63
In-person reminders	64	1	2

9 Measures and Results • Classroom Assignments

9.1 Classroom assignment scales⁶

The scales used to rate the quality of the assignments were developed at CRESST (Clare et. al, 2001) and consisted of the following six dimensions, each of which was rated on a 4-point scale (1 = *poor* to 4 = *excellent*):

Cognitive challenge of the task. This dimension describes the level of thinking required of students to complete the task. Specifically, this dimension describes the degree to which students have the opportunity to apply higher order reasoning and engage with academic content material.

Clarity of the learning goals. This dimension describes how clearly a teacher articulates the specific skills, concepts, or content knowledge students are to gain from completing the assignment. The primary purpose of this dimension is to describe the degree to which an assignment could be considered a purposeful, goal-driven activity focused on student learning.

⁶ Sections 9.1 and 9.2 are taken from Clare, Lindsay (2001) *A Preliminary Report Describing the Quality of Classroom Assignments from the LAUSD Local District Accountability Measures Project*. Los Angeles: University of California, Center for Research on Evaluation, Standards and Student Testing.

Clarity of the grading criteria. The purpose for this dimension is to assess the quality of the grading criteria for the assignment in terms of their specificity and potential for helping students improve their performance. How clearly each aspect of the grading criteria is defined is considered in the rating, as well as how much detail is provided for each of the criteria.

Alignment of the learning goals and task. This dimension focuses on the degree to which a teacher's stated learning goals are reflected in the design of the assignment tasks students are asked to complete. Specifically, this dimension attempts to capture how well the assignment appears to promote the achievement of the teacher's goals for student learning.

Alignment of goals and grading criteria. This dimension is intended to describe the degree to which a teacher's grading criteria support the learning goals, i.e., the degree to which a teacher assesses students on the skills and concepts they are intended to learn through the completion of the assignment.

Overall quality. This dimension is intended to provide a holistic rating of the quality of the assignment based on its level of cognitive challenge, the clarity of the learning goals, and the clarity of the grading criteria, the alignment of the learning goals and the assignment task, and the alignment of the learning goals and the grading criteria. As described earlier, scores were not separately recorded for the alignment dimensions, though these two dimensions were considered in the overall quality ratings.

The rubric used for scoring the classroom assignments is reproduced on the next two pages. For each dimension, the qualities associated with each rank are detailed and, in most cases, examples are provided to aid scoring. Indeed, scoring agreement across raters (within one scale point) was very high for both elementary and secondary classroom assignments (91.7% to 100%).

Language Arts Assignment Rubric

<p>COGNITIVE CHALLENGE</p>	<p>4 Task requires strongly complex thinking as an extensive, major focus of task. Student also engages with substantive content material.</p> <p>E.g. Student may be asked to synthesize ideas; analyze cause and effect; identify a problem and pose reasonable solutions; hypothesize; speculate with details or justification; defend opinions or argue a position with evidence; evaluate; analyze (distinguishing important or irrelevant); determine bias, values, intent.)</p>	<p>3 Task requires complex thinking. Student may also engage with substantive content material.</p> <p>E.g. Student may be asked to synthesize ideas; analyze cause and effect; identify a problem and pose reasonable solutions; hypothesize; speculate with details or justification; defend opinions or argue a position with evidence; evaluate; analyze (distinguishing important or irrelevant); determine bias, values, intent.)</p>	<p>2 Task requires some moderately complex thinking. Some substantive content area material may be covered.</p> <p>E.g. Student may be asked to summarize straightforward information, infer simple main idea, or simply apply the appropriate format for a given genre).</p>	<p>1 Task does not require any degree of complex thinking and/or does not engage students with substantive content material.</p> <p>E.g. Student may be required to recall basic information, or recall definitions. Or student may be asked to answer simple reading comprehension questions or write on a topic with little focus or structure.)</p>
<p>FOCUS OF THE GOALS ON STUDENT LEARNING</p>	<p>4 Goals are very focused on student learning. Goals are very clear and explicit in terms of what students are to learn from the assignment. Additionally, all the goals are elaborated.</p>	<p>3 Goals are mostly focused on student learning. Goals are mostly clear and explicit in terms of what students are to learn from the assignment.</p>	<p>2 Goals are somewhat focused on student learning. Goals are somewhat clear and explicit in terms of what students are to learn from the assignment. Goals may be very broadly stated (e.g. reading comprehension). Or there may be a combination of learning goals and activities.</p>	<p>1 Goals are not focused on student learning, goals are not clear and explicit in terms of what students are to learn from the assignment OR all goals may be stated as activities with no definable objective ("activity for activity's sake").</p>
<p>CLARITY OF THE GRADING CRITERIA</p>	<p>4 Teacher's grading criteria are very clear, explicit, and elaborated.</p> <p>E.g. Teachers' rubric or guidelines are detailed and elaborated. Additionally a model of good work may be provided to the students.</p>	<p>3 Teacher's grading criteria are mostly clear and explicit with regard to what is expected with little or no question.</p> <p>E.g. Teacher may use a rubric or a very elaborated and specific list of dimensions.</p>	<p>2 Teacher's grading criteria are somewhat clear and explicit. Teacher provides some general directions or a rudimentary rubric.</p> <p>E.g. a list of dimensions such as "style, creativity, and organization," but some dimensions are undefined or vague.</p>	<p>1 Teacher does not specify a grading criteria. OR it is not possible to determine the grading criteria from</p>

ALIGNMENT OF LEARNING GOALS AND TASK	<p>4 There is exact alignment between teacher's stated learning goals for students on that assignment and what the task asks students to do, AND task fully supports instructional goals.</p> <p>E.g. goal is being able to summarize several points and activity entails summarizing; tasks and goals overlap completely—neither one calls for something not included in the other.</p> <p><i>Note: This dimension cannot be rated a '4' if the goals are unclear, broadly stated, or stated as activities.</i></p>	<p>3 There is good alignment between teacher's stated learning goals and what the task asks students to do, —AND the task supports instructional goals.</p>	<p>2 There is only some alignment between teacher's stated goals and what the task asks students to do. The task only somewhat supports the instructional goals.</p> <p>E.g. goal is to be able to write an essay, but task calls for completing a concept map and making an outline for an essay (but NOT actually writing an essay).</p> <p>—OR the goal may be so broadly stated that the task and goal are aligned at a very general level.</p>	<p>1 There is very little or no alignment between teacher's stated goals and what the task asks students to do. The task does not support the instructional goals.</p> <p>E.g. goal calls for writing an essay, but task calls for giving an oral report.</p>
ALIGNMENT OF LEARNING GOALS AND GRADING CRITERIA	<p>4 There is exact alignment between teacher's stated learning goals for students on that assignment and teacher's stated grading criteria.</p> <p>E.g. goal is to write a persuasive essay, and criteria include appropriate dimensions such as stating a point of view and providing relevant supporting evidence; do not include dimensions not mentioned in goals (e.g. creativity).</p> <p><i>Note: This dimension cannot be rated a '4' if the goals are unclear, broadly stated, or stated as activities.</i></p>	<p>3 There is good alignment between teacher's stated learning goals and the stated criteria for grading.</p> <p>E.g. goal is to write a persuasive essay, and criteria include appropriate dimensions but also extraneous ones.</p> <p>Or, fails to include critical dimension (e.g., support for assertions or point of view).</p>	<p>2 There is only some alignment between teacher's stated learning goals and the stated grading criteria.</p> <p>E.g. goal is to write a business letter, but criteria include <u>mostly</u> extraneous dimensions, e.g. participation in class discussion is given more weight than letter format.</p> <p>Or, criteria given are not very appropriate, e.g. slang is acceptable in a business letter.</p>	<p>1 There is very little or no alignment between teacher's stated learning goals and the stated grading criteria.</p>
OVERALL TASK QUALITY (Consider all previous dimensions.)	<p>4 Excellent quality in terms of level of cognitive challenge, clarity and application of learning goals and grading criteria.</p>	<p>3 Good quality in terms of level of cognitive challenge, clarity and application of learning goals and grading criteria.</p>	<p>2 Limited quality in terms of level of cognitive challenge, clarity and application of learning goals and grading criteria.</p>	<p>1 Poor quality in terms of level of cognitive challenge, clarity and application of learning goals and grading criteria.</p>

Most of the assignments were rated a 2 across the different dimensions at each grade level. The corresponding means ranged from 1.38 to 2.67. In general, tenth grade scores were highest across the dimensions for each assignment type. In terms of overall quality, tenth grade assignments scored higher than seventh and fourth grade assignments for writing (2.46, 2.00, 1.96), reading comprehension #1 (2.40, 1.50, 2.08), and reading comprehension #2 (2.17, 1.50, 1.96). Specific results for each grade level regarding the quality and nature of the classroom assignments with attendant examples are presented in Tables 11 through 19.

Table 11. Fourth-Grade Writing Assignment Scores by Dimension

Dimension	Rating										\bar{x} (SD)
	1		2		3		4				
	<i>N</i>	<i>n</i>	%	<i>n</i>	%	<i>N</i>	%	<i>n</i>	%		
Cognitive challenge	26	5	19.2%	18	69.2%	3	11.5%	0	0.0%	1.92 (0.11)	
Clarity of goals for student learning	26	0	0.0%	18	69.2%	7	26.9%	1	3.8%	2.35 (0.11)	
Clarity of grading criteria	25	2	8.0%	9	36.0%	10	40.0%	4	16.0%	2.64 (0.17)	
Overall task quality	26	5	19.2%	17	65.4%	4	15.4%	0	0.0%	1.96 (0.12)	

Note. Assignments rated on a 4-point scale (1 = *poor*, 4 = *excellent*).

Table 12. Fourth-Grade Reading Comprehension Assignment (#1) Scores by Dimension

Dimension	N	Rating								\bar{x} (SD)
		1		2		3		4		
		n	%	n	%	n	%	n	%	
Cognitive challenge	25	3	12.0%	13	52.0%	9	36.0%	0	0.0%	2.24 (0.13)
Clarity of goals for student learning	25	1	4.0%	14	56.0%	9	36.0%	1	4.0%	2.40 (0.13)
Clarity of grading criteria	24	11	45.8%	8	33.3%	4	16.7%	1	4.2%	1.79 (0.18)
Overall task quality	25	3	12.0%	17	68.0%	5	20.0%	0	0.0%	2.08 (0.11)

Note. Assignments rated on a 4-point scale (1 = poor, 4 = excellent).

Table 13. Fourth-Grade Reading Comprehension Assignment (#2) Scores by Dimension

Dimension	N	Rating								\bar{x} (SD)
		1		2		3		4		
		n	%	n	%	n	%	n	%	
Cognitive challenge	25	7	28.0%	10	40.0%	8	32.0%	0	0.0%	2.04 (0.16)
Clarity of goals for student learning	25	2	8.0%	13	52.0%	9	36.0%	1	4.0%	2.36 (0.14)
Clarity of grading criteria	25	10	40.0%	11	44.0%	3	12.0%	1	4.0%	1.80 (0.16)
Overall task quality	25	7	28.0%	12	48.0%	6	24.0%	0	0.0%	1.96 (0.15)

Note. Assignments rated on a 4-point scale (1 = poor, 4 = excellent).

Table 14. Seventh-Grade Writing Assignment Scores by Dimension

Dimension	N	Rating								\bar{x} (SD)
		1		2		3		4		
		n	%	n	%	n	%	n	%	
Cognitive challenge	10	0	0.0%	9	90.0%	1	10.0%	0	0.0%	2.10 (0.10)
Clarity of goals for student learning	10	0	0.0%	8	80.0%	2	20.0%	0	0.0%	2.20 (0.13)
Clarity of grading criteria	10	3	30.0%	5	50.0%	2	20.0%	0	0.0%	1.90 (0.23)
Overall task quality	10	1	10.0%	8	80.0%	1	10.0%	0	0.0%	2.00 (0.15)

Note. Assignments rated on a 4-point scale (1 = poor, 4 = excellent).

Table 15. Seventh-Grade Reading Comprehension Assignment (#1) Scores by Dimension

Dimension	N	Rating								\bar{x} (SD)
		1		2		3		4		
		n	%	n	%	n	%	n	%	
Cognitive challenge	8	3	37.5%	4	50.0%	1	12.5%	0	0.0%	1.75 (0.25)
Clarity of goals for student learning	8	2	25.0%	5	62.5%	1	12.5%	0	0.0%	1.88 (0.23)
Clarity of grading criteria	8	2	25.0%	6	75.0%	0	0.0%	0	0.0%	1.75 (0.16)
Overall task quality	8	4	50.0%	4	50.0%	0	0.0%	0	0.0%	1.50 (0.19)

Note. Assignments rated on a 4-point scale (1 = poor, 4 = excellent).

Table 16. Seventh-Grade Reading Comprehension Assignment (#2) Scores by Dimension

Dimension	N	Rating								\bar{x} (SD)
		1		2		3		4		
		n	%	n	%	n	%	n	%	
Cognitive challenge	8	2	25.0%	6	75.0%	0	0.0%	0	0.0%	1.75 (0.16)
Clarity of goals for student learning	8	1	12.5%	7	87.5%	0	0.0%	0	0.0%	1.88 (0.13)
Clarity of grading criteria	8	5	62.5%	3	37.5%	0	0.0%	0	0.0%	1.38 (0.18)
Overall task quality	8	4	50.0%	4	50.0%	0	0.0%	0	0.0%	1.50 (0.19)

Note. Assignments rated on a 4-point scale (1 = poor, 4 = excellent).

Table 17. Tenth-Grade Writing Assignment Scores by Dimension

Dimension	N	Rating								\bar{x} (SD)
		1		2		3		4		
		n	%	n	%	n	%	n	%	
Cognitive challenge	13	0	0.0%	7	53.8%	3	23.1%	3	23.1%	2.69 (0.24)
Clarity of goals for student learning	13	0	0.0%	8	61.5%	4	30.8%	1	7.7%	2.46 (0.18)
Clarity of grading criteria	13	2	15.4%	5	38.5%	4	30.8%	2	15.4%	2.46 (0.27)
Overall task quality	13	2	15.4%	5	38.5%	4	30.8%	2	15.4%	2.46 (0.27)

Note. Assignments rated on a 4-point scale (1 = poor, 4 = excellent).

Table 18. Tenth-Grade Reading Comprehension Assignment (#1) Scores by Dimension

Dimension	N	Rating								\bar{x} (SD)
		1		2		3		4		
		n	%	n	%	n	%	n	%	
Cognitive challenge	10	0	0.0%	6	60.0%	3	30.0%	1	10.0%	2.50 (0.22)
Clarity of goals for student learning	10	1	10.0%	5	50.0%	4	40.0%	0	0.0%	2.30 (0.21)
Clarity of grading criteria	10	6	60.0%	1	10.0%	3	30.0%	0	0.0%	1.70 (0.30)
Overall task quality	10	0	0.0%	7	70.0%	2	20.0%	1	10.0%	2.40 (0.22)

Note. Assignments rated on a 4-point scale (1 = poor, 4 = excellent).

Table 19. Tenth-Grade Reading Comprehension Assignment (#2) Scores by Dimension

Dimension	N	Rating								\bar{x} (SD)
		1		2		3		4		
		n	%	n	%	n	%	n	%	
Cognitive challenge	12	1	8.3	7	58.3	3	25.0	1	8.3	2.33 (0.22)
Clarity of goals for student learning	12	0	0	8	66.7	3	25.0	1	8.3	2.42 (0.19)
Clarity of grading criteria	12	5	41.7	5	41.7	2	16.7	0	0	1.75 (0.22)
Overall task quality	12	2	16.7	7	58.3	2	16.7	1	8.3	2.17 (0.24)

Note. Assignments rated on a 4-point scale (1 = poor, 4 = excellent).

9.2 Sample Fourth-Grade Assignments

The following examples of fourth-grade writing assignments illustrate 1) those typically submitted by our teachers (i.e., assignments scored a 2 for overall quality) and 2) exemplary assignments from previous CRESST data collections

(i.e., assignments rated a 4 for overall quality) to illustrate the range in assignments captured by the scoring dimensions.

We focus on writing examples because it is likely that a single writing assignment will be collected from teachers next year. This will be done in order to reduce the burden on teachers, and this recommendation is based on findings from this year's data. We focus on a single grade for brevity's sake. Writing examples of other grade levels as well as reading examples are included in Appendix C.

Basic-quality writing assignment. For this assignment students were instructed to write a paragraph on a topic of their choice. As a preliminary activity, students completed a worksheet where they wrote down the main idea of their topic, four details to support it, and a concluding sentence. Students then strung these together to create a paragraph.

This assignment was scored a 2 for cognitive challenge because it required only moderately complex thinking on the part of the students. While learning to write a paragraph is certainly an important activity, students were not required to engage with any degree of substantive content material to complete the task. For example, students wrote about going to a concert, the zoo, etc. Writing on a more challenging topic would have increased the level of cognitive challenge for this assignment. Additionally, the directions to students regarding how to write a proper paragraph were fairly surface-level in that they focused only on there being a main idea and supporting details. For example, students were not given information about how to link ideas or include interesting or compelling details in their writing.

The clarity of the teacher's stated goals for this assignment also was scored a 2.

The teacher's goals for this assignment were as follows:

To write a paragraph using good sentences, spelling, and punctuation.

The goal for this assignment was considered to be of basic quality because the teacher did not specify what specific skills students were to learn from

completing the task, beyond carrying out the activity. Additionally, it was not clear what the teacher meant by a “good” sentence. To receive a higher score, the goal should be more elaborate, explicit, and focused on student learning.

The teacher’s stated grading criteria were:

Content, sentence structure, punctuation and spelling.

These grading criteria were scored a 2 because the teacher only described the broad categories she would be attending to when grading students’ paragraphs, but did not provide details of what she was looking at within these broad categories.

Sample student work. The following is a sample of student work for this writing assignment considered by the teacher to be of medium quality for the class.

1. Choose an idea for your paragraph. Write the title, main idea, and details. Retell the main idea at the end.

Title of Paragraph: A Sleep Over
Main Idea: you go to somebodys
Details: 1. house a you sleep at
 2. there house till morning
 3. and then you go
 4. home.

2. Use the information to write a paragraph. Include the main idea and details then retell the main idea. Indent the first sentence. Use capitals and periods. Remember to give the paragraph a title.

A Sleep Over

Last Saturday I had a sleepover. We had a party. We had 4 big boxes of pizza. We watched movies. We had a lot of music and we stayed up till midnight and they stayed for 2 days. We played lots of games.

Excellent-quality writing assignment. The following writing assignment scored a 4 for overall quality. Students used Laura Ingalls Wilder’s descriptive, sensory writing

style as a model for their own descriptions of a setting. The students chose a familiar place to describe, wrote notes about the place paying particular attention to sensory details, and then wrote several drafts of their descriptions. Students also drew a picture of their setting as part of their brainstorming exercises.

The following were questions that students were to answer about their settings before they started to write their description.

- What do I see?
- What do I hear?
- What do I smell?
- What other details do I want to include?
- Will I use lots of details in my description?
- Will I paint a believable picture with my words?
- Will readers feel as if they have visited the place I describe?

This assignment was scored a 4 for cognitive challenge because of the elaborate preparation required of students prior to writing, their involvement with high-level reading content, and the length and sophistication expected in their descriptions. As part of their prewriting, students read and analyzed Wilder's *Little House on the Prairie* and used her writing style as a model for their own writing. They completed a worksheet that structured their prewriting notes.

The teacher's stated goals for the assignment were:

Practice the writing process, create their own sensory style. To learn to write descriptively, to learn to describe a setting with sensory language that makes the place believable and helps the reader feel present. I wanted them to be able to learn from a good writer and to be able to use her writing as a model for their own writing. This was accomplished by having a class discussion where we talked about Wilder's specific style of writing in describing the setting of the story. After students chose the place they were to describe, they used a prewriting organizer to help plan their description.

The grading criteria (scored a 3) were as follows:

Proficient	<p>You described the setting with specific sensory language that makes the place believable and helps the reader feel present.</p> <p>You understood and followed instructions of assignment. Your work is neat and organized.</p>
Apprentice	<p>The setting is described, but not always in vivid detail. The place may not be believable. The reader may not feel present.</p> <p>You understood and followed the basic point of the assignment, but missed some details. Your work was fairly neat and organized.</p>
Novice	<p>The description may be vague or generalized rather than detailed and vivid. It may not transport the reader to the scene.</p> <p>You did not understand the assignment. Your work was unorganized and messy.</p>

Scorers rated this assignment a 4 overall because it demonstrated high quality on all the dimensions. Students were engaged with substantive content material and were required to use complex thinking skills. The teacher’s goals were clear in terms of what students were expected to learn by completing the assignment, and these goals were reflected in the grading criteria which were fairly elaborated and provided students with information for what they would need to do to successfully complete the task.

Sample student work. The following is a sample of student work for this writing assignment considered by the teacher to be of medium quality for the class.

The beach is the best place to be! Some people walk around the beach, and many come to swim. There are kids playing in the sand and making sand castles, and sometimes the waves destroy their creations. Waves crash against the shore, and slimy seaweed land on the sand. The white and naughty seagulls steal peoples sandwiches while they are dozing under their umbrellas, or playing in the water. It is very hot at the beach, but it has very cold water. There are dolphins hardly jumping out of the water because they are afraid of the humans, and people are riding on waves, either on their stomachs or on their boogie boards. The bright yellow sun shines over the large wide salt water sea. Airplanes and helicopters fly over the great and adventurous ocean. The beach is such a lovely place and it has so much adventure.

10 Scaling-up the Local District Performance Measures – 2001-2002

10.1 Scoring Considerations

The LDPM design committee envisioned a four point system whereby local districts received one point each for meeting the established targets for four indicators. This scoring system, while easy to understand, may unintentionally sacrifice precision for parsimony.

Upon further consideration, we will delay the release of a scoring system until we are confident that it reflects an optimal balance of psychometric and pedagogic qualities. These two qualities frequently conflict, but we believe that with due consideration, we will be able to reconcile our scoring system so that it is accurate and provides useful feedback to local district management.

10.1.1 The Scorecard

The scorecard below (Table 20) summarizes the performance of District P across all four indicators. A detailed report of District P scoring follows this scorecard. Again, this scorecard represents just one possible scoring system, not the actual system that will be used in 2001-02.

Table 19. Scorecard for Local District P

Indicator	% of Schools Meeting Target	Target	Met Goal
API Growth Target	58%	67%	NO
Matched Student Reading Gains	49%	60%	NO
School Organization Index	60%	60% above Composite Score	YES
Classroom Assignments	Above 60% for one indicator, below 60% for other indicators	60% above 2.25 for each indicator	NO
Total Score		4	1

10.1.2 API Growth Target

The LDPM framing document set the standard for each local district that 60% of its schools would have to meet or exceed the API growth targets and NCE gains. We have adjusted this target upward to reflect the intent of the framing document—that the API growth target for the district would reflect the statewide API award rate. For the 1999-2000 school year, this rate was 67%. As Table 21 shows, District P fell short of the API Award goal— 58% of its schools met the API award target. The failure to meet this criterion means that District P receives no points for this component of the Local District Performance Measures (Table 20).

10.1.3 SAT/9 Matched NCE Reading Gains

Fewer than 60% of the schools within Local District P met the Matched NCE Reading Gain target. No point is awarded for this indicator (Table 20).

10.1.4 School Organization Index

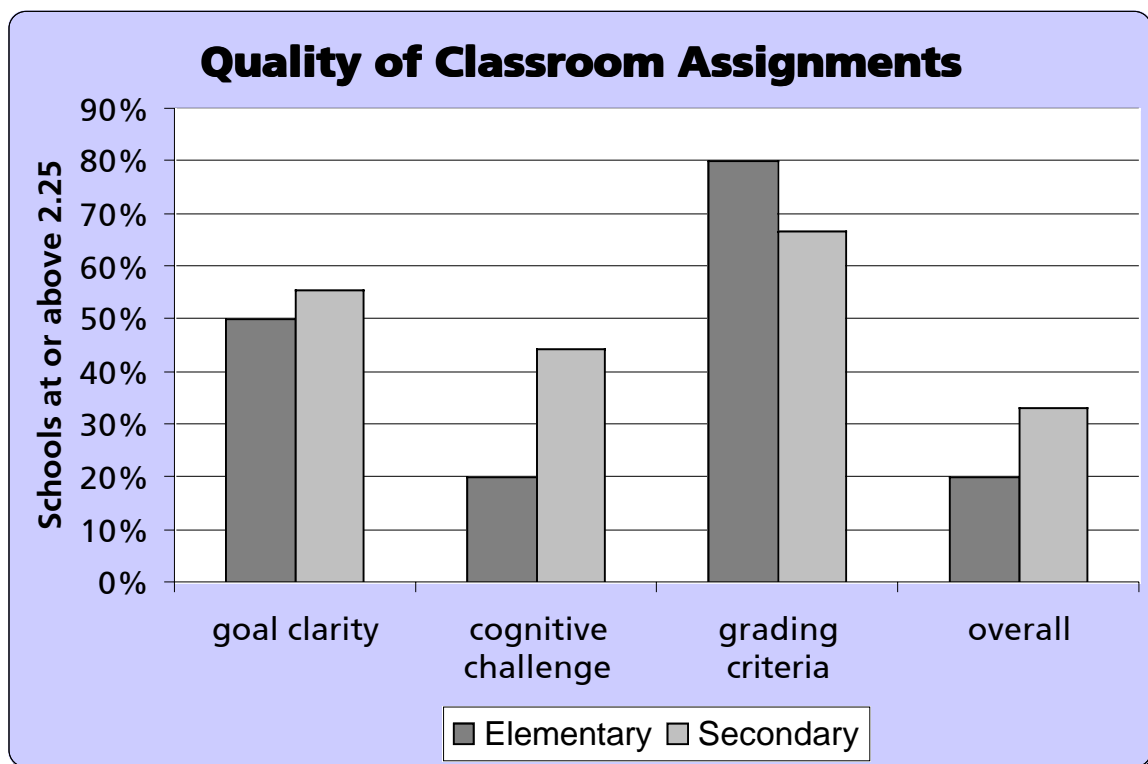
As reported above, we found several of the domains measured by the School Organization Index to be correlated with school performance, especially as regards attainment of API growth targets. Although the scoring system has not been developed, one possibility is that the score will reflect a weighted average of the School Organization Index domains. In such a system, the weights would be relative to each domain's ability to predict student performance. For each domain, our standard will be agreement (4.0 on a 5.0 scale) that the domain is present and healthy in 60% of a local district's schools.

In District P, 62.5% of its schools reported their principals as having desirable leadership traits, and 60.5% of its schools reported a healthy school climate and culture. If these indicators were weighted most heavily (although further analysis may suggest that other indicators merit higher weights), then District P would most likely meet the standard of 60% for the entire School Organization Index. For illustrative purposes, we will assume that this is the case and award District P one point for meeting the School Organization Index target.

10.1.5 Classroom Assignments

We assessed the quality of classroom assignments across the four key dimensions: Goal Clarity, Cognitive Challenge, Grading Criteria, and Overall Quality. Our standard for each dimension was a scale score of 2.25 out of 4.00. As with the other indicators, our criterion for successful local district performance is 60% of schools meeting or exceeding the standard. Schools in District P met our criteria for one dimension of four. Table 23 reports the percent of schools at both elementary and secondary levels scoring above the cut point of 2.25 across all four dimensions. District P will be awarded no points as it did not meet the Classroom Assignments target.

Table 20. Quality Ratings of Classroom Assignments by Dimension and School Level



11 Modifications to LDPM Design for 2001-02 School Year

In addition to developing a comprehensive scoring system, we anticipate several changes to the research design. Some of these changes are based on what we learned from the pilot administration of the LDPM, others reflect changes in the district's reform emphasis that have arisen since the development of the LDPM design.

The API growth target and matched student gains in reading indicators will continue relatively unchanged. In addition to the scoring system modifications mentioned above, we anticipate adjusting the NCE targets to reflect what we have learned from examining several years of student SAT/9 matched reading gain scores.

The School Organization Index will undergo several changes in the administration of the instruments. Instead of beginning our data collection with computer-aided telephone interviews, we will start with self-administered questionnaires. We will also improve the quality of our data by confirming the accuracy of our staffing lists, expanding our surveys to include home languages in addition to English and Spanish, increasing the visibility of the project, and limiting the burden to respondents. A more detailed explanation of these changes can be found in Appendix D.

The content of the School Organization Index will also be modified to better reflect the District's emphasis on standards-based learning, including professional development based on the Principles of Learning provided by the University of Pittsburgh's Institute for Learning.

The Classroom Assignments will be modified to reduce respondent burden. The number of collected assignments will be reduced from three to one. The emphasis will be on the writing assignment, rather than the reading assignments. Reducing the number of assignments is likely to increase the response rate as well as increase the relevance of our data collection for elementary teachers whose reading assignments are typically prescribed by the adopted reading program at the school. Also, we will emphasize as our fourth

category (replacing overall quality) the alignment among goals, tasks, and grading criteria.

We are confident that the changes we have made to the Local District Performance Measures will result in more accurate data and less intrusive data collection.

Appendix A: Ranking Schools Using SAT/9: Six Approaches

In this analysis we attempt to answer the question: To what extent do various statistical manipulations of student performance scores advantage or disadvantage individual schools? We will consider several different statistical approaches. Among the issues we seek to explore are the following:

Should we use simple matched gain scores to assess relative performance or should we estimate expected scores to serve as the criterion for successful school performance?

Should expected growth be constant across schools or vary by initial performance levels?

Should expected growth values be derived from observations at the individual level or from observations at the school level (using aggregated individual scores)?

Our approach is practical. We will examine the effect of the method on the relative ranking of school-level SAT/9 reading performance. This will allow us to answer several questions regarding the impact of a given method on specific schools. More importantly, we will be able to assess whether schools of a given characteristic are systematically advantaged or disadvantaged by a given method. For example, a given method may favor schools with high initial NCE scores. A school with high-performing students, such as Topanga El. (64.08), may be subject to far lower growth expectations than a school with low-performing students, such as Grape St. El. (25.96). At this point, we do not make normative judgements regarding a given method, rather we seek to describe how each method changes the ranking of schools.

Matched Gain Scores (MGS)

Matched gain scores reflect school growth as a function of the change in individual student scores. Table 1 (see below) shows that the use of MGS without any adjustment for starting value appears to advantage schools with

lower initial NCE scores. When using MGS as the basis for ranking schools, 56 schools with initial NCE scores in the bottom third of the distribution are ranked among the top 100 with respect to gains. Conversely, 59 schools with initial NCE scores in the top third of the distribution are ranked among the bottom 100 with respect to gains. We believe there is no compelling reason to expect that the schools with the lowest initial NCE scores are LAUSD's best schools and that the schools with the highest initial NCE scores are LAUSD's worst schools. Thus, the uneven distribution of schools throughout the highest and lowest ranks suggests that the use of matched gain scores would bias the rankings so as to favor those schools with the lowest initial NCE scores.

Adjusted Gain Scores (AGS)

Adjusted gain scores reflect the difference between the matched gain score and the expected gain score. The expected gain score is derived from the average school-level matched gain scores for those schools with initial NCE scores (1998) in a given performance band. Table 1 shows that AGS does not appear to advantage schools in any given performance band. Schools with high, medium, and low initial NCE scores are close to being equally represented among the top and bottom 100 ranked schools.

Residual Scores

Residual scores are derived from the regression equation that predicts an individual's current performance from past performance. The residual is the difference between actual performance and predicted performance. In practice, predicted scores are lower for higher performing students⁷ and higher for lower performing students. For example, a student whose pretest score is 65 might have a predicted posttest score of 62. If this student's actual posttest score is 63, the residual gain would be a +1 even though the actual (unresidualized) gain is a -2 NCE points.

⁷ Students will be used for ease of exposition. Any unit of analysis, including school, could substitute for student in this example.

We residualized test scores at both the individual student and school levels. Individual-level residual scores were derived from regression equations that calculated a residual score for each individual. These individual-level residuals were then aggregated to the school level for our analysis. School level residual scores were derived from regression equations that used school-level aggregate data. Each school's pretest and posttest score was derived from the mean of individual student scores. As with all analyses noted herein, only data from students with scores in both 1998 and 1999 were used for this analysis. School-level residuals differ from individual-level residuals because the order of operations differs. School-level residuals are based on data that is first aggregated, then residualized. Individual-level residuals are computed for individuals first, then aggregated. Both methods represent attempts to discern school performance.

Residual Scores – Individual Level

We examined two different estimation methods at the individual level. RIG predicted scores based on the entire population of students. RIT predicted scores based on school type, whether the school is elementary, middle or secondary. Table 1 shows the impact of these two residual methods on schools in different performance bands. Whether the residual was derived from a global or a type-specific sample did not appear to influence the results. In both cases, schools with high initial scores are disproportionately represented in the top 100 and schools with low initial scores are disproportionately represented in the bottom 100. Residualizing test scores at the individual level appears to disadvantage schools in lower performance bands.

Residual Scores – School Level

As with individual level analyses, we examined estimates that were based on the entire population of schools (RSG) and estimates based on regression equations derived separately for each type of school (RST). Neither treatment appears to have disadvantaged schools according to initial NCE scores.

Table A-1. Distribution of Rankings by Initial NCE scores (Thirds)

Rankings	Top 100 Gains		Bottom 100 Gains	
	Low Initial NCE Scores	High Initial NCE Scores	Low Initial NCE Scores	High Initial NCE Scores
Mean Gain (MGS)	56	10	14	59
Adjusted Gain (AGS)	24	31	33	44
Res. Individual (RIG)	9	67	46	19
Res. Individual (RIT)	8	72	61	9
Res. School (RSG)	36	30	31	34
Res. School (RST)	36	32	31	37

School-level Residuals or Adjusted Gain Scores?

Both school-level residuals and adjusted gain scores have the desirable property of not disadvantaging schools based on initial performance level. We have come to the conclusion that adjusted gain scores are to be preferred to school-level residuals for several reasons. First, adjusted gain scores are more readily interpretable. The practice of comparing schools within performance bands mirrors the API. Additionally, the concept of a gain is more readily understood than the concept of the residual.

Second, residual scores could have distorted consequences. For any student or school above the mean, the predicted gain will be negative. A simple example will help. Take a school with a high enough initial score, say 65. This school could lose a point per year for each of the next seven years and still be rewarded for positive performance *because in each of the seven years, the predicted loss exceeds the actual loss*. The converse is true as well. A low scoring school could gain for several years consecutively and suffer adverse consequences each and every year because the predicted gain for each given year exceeds the actual gain.

Table A-2. Spring 1998 to Spring 1999 School-Level Gains by Initial NCE Score

Quintiles (Initial NCE)	Quintile 1 (17.0-20.1)	Quintile 2 (20.2-33.5)	Quintile 3 (33.6-38.3)	Quintile 4 (38.4-45.8)	Quintile 5 (45.9-73.4)
Elementary (N of schools)	3.1 (91)	2.9 (87)	2.3 (87)	1.8 (83)	1.0 (95)
Middle	3.5 (17)	3.3 (17)	3.4 (20)	2.8 (15)	1.9 (16)
Senior	0.3 (6)	0.5 (15)	0.4 (11)	-0.5 (19)	-0.3 (9)

Thirds (Initial NCE Range)	Low (17.0-32.9)	Med (33.0-39.4)	High (39.5-73.4)
Elementary (N of schools)	3.0 (163)	2.3 (121)	1.3 (159)
Middle	3.4 (32)	3.3 (26)	2.3 (27)
Senior	0.6 (17)	0.1 (17)	-0.4 (26)

METHODOLOGY: To compute this data, like grade levels were combined into elementary, middle, and senior high school levels. Elementary schools were defined as grade levels 1-5 (or 1-6 if an elementary school also had 6th graders). Middle schools were defined as grade levels 6-8. Senior high schools were defined as grade levels 9-11. Schools with multiple school levels were grouped into separate levels using the same criteria mentioned above. Thus, in some cases, the same multi-level school could be considered in elementary, middle, and senior high school levels. It should be noted that this analysis does not include opportunity, continuation, special education schools, or schools with less than 30 students assessed.

Table A-3. Spring 1998 to Spring 1999 Grade-Level Gains by Initial NCE Score (Quintiles)

Grade Level (Initial NCE)	Quintile 1 (17.0-20.1)	Quintile 2 (20.2-33.5)	Quintile 3 (33.6-38.3)	Quintile 4 (38.4-45.8)	Quintile 5 (45.9-73.4)
2 (N of schools)	1.7 (18)	-0.7 (22)	-0.5 (71)	-0.9 (172)	-2.3 (162)
3	2.7 (91)	2.5 (72)	1.7 (96)	1.0 (89)	0.7 (94)
4	4.7 (159)	4.2 (73)	3.7 (71)	3.6 (60)	2.9 (80)
5	3.9 (136)	3.5 (71)	3.0 (82)	2.5 (65)	0.9 (90)
6	4.6 (25)	4.7 (19)	4.0 (24)	3.1 (19)	2.4 (20)
7	0.9 (17)	1.3 (15)	1.7 (23)	1.5 (15)	1.8 (16)
8	6.6 (22)	5.0 (14)	4.4 (20)	3.4 (13)	2.4 (17)
9	-0.2 (6)	-0.8 (10)	-1.3 (15)	-2.8 (17)	-2.8 (12)
10	-0.5 (8)	-0.3 (17)	-0.5 (9)	-0.7 (20)	-0.3 (6)
11	4.7 (9)	3.8 (10)	3.7 (14)	3.0 (19)	3.5 (8)

Note: Tables by grade level exclude continuation, opportunity, and special education schools.

Table A-4. Spring 1998 to Spring 1999 Grade-Level Gains by Initial NCE Score (Thirds)

Grade Level (Initial NCE)	Low (17.0-32.9)	Med (33.0-39.4)	High (39.5-73.4)
2 (N of schools)	0.6 (35)	-0.7 (108)	-1.6 (302)
3	2.7 (151)	1.7 (122)	0.8 (169)
4	4.5 (226)	3.8 (88)	3.2 (129)
5	3.8 (197)	3.1 (105)	1.4 (142)
6	4.5 (41)	4.3 (30)	2.6 (33)
7	1.2 (29)	1.5 (28)	1.7 (29)
8	6.1 (32)	4.5 (27)	2.7 (27)
9	-0.5 (15)	-1.3 (17)	-2.9 (28)
10	-0.4 (23)	-0.8 (15)	-0.4 (22)
11	4.3 (18)	3.5 (18)	3.2 (24)

Note: Tables by grade level exclude continuation, opportunity, and special education schools.

Appendix B: Indicators of Effective School Organization

Characteristics: What are the organizational characteristics of effective schools? Admittedly, much of the effective schools literature boils down to an examination of high performance schools to examine if specific characteristics of high performance schools are unique to high performing schools (e.g. Levine & Lezotte, 1990). The causal linkage, methodological or theoretical, in these studies is nearly non-existent. Thus the term “correlates” is widely used within the literature to describe what are otherwise referred to as “factors of effective schooling.” Researchers have identified anywhere from five (Edmunds’ (1979) famous “five-factor” model of school effectiveness) to twenty or so correlates of effective schooling identified across various studies. Based on an examination of several reviews of the literature (Gray & Wilcox, 1995; Levine & Lezotte, 1990; Reynolds, 1992; Sammons et al., 1995; Zigarelli, 1996), we find that the correlates from nearly every study can be classified into one of five groups: leadership, teaching, climate, student assessment, and improvement.

Within each group it is the qualitative differences and emphases placed that distinguish the effective school from its less able counterpart. Also, despite the nominal similarity, researchers’ views of the factors differ widely. For example, effective leadership was conceptualized either as “strong” or “participative” depending upon the researcher. Therefore, this discussion of each of the correlate groups will highlight the meaning, the differences, and the possible contradictions within the empirical work on school characteristics.

One small note: Recently, researchers have taken to the notion of contextualization— what is known in organization theory as a contingency model— whereby effectiveness is a function of the fit between the factor of importance and the context/environment.⁸ A contingency approach changes the interpretation of what may otherwise seem to be conflicting effectiveness

⁸ This is referred to, in the school effectiveness literature, as generalizability vs. situational dependency (Scheerens & Creemers, 1989).

characteristics. When possible, contextual detail from the empirical studies will be provided in order to facilitate contingency-type analyses.⁹ What follows is an extended discussion of each category, including the empirical support for or against specific factors of effectiveness, and other research relevant to the debate.

Correlate One: Leadership

School effectiveness research is best served by a conception of leadership that expands beyond the principal to include both other people (i.e. administrators and teachers) and guidance systems (i.e. goals, visions). These aspects are interconnected and are often, but not always, a reflection of choices made at the top of the organizational hierarchy.

Many studies note the importance of the school principal. Characteristics of the principal that are associated with school effectiveness include instructional leadership or academic focus (Edmonds, 1979; Mortimore, Sammons, Ecob, Stoll, & Lewis, 1988; Teddlie & Stringfield, 1993), strong promotion of school goals or vision (Hallinger & Murphy, 1985; Rutter et al., 1979; Sammons et al., 1995), participative or representative leadership style (Louis & Miles, 1992; Mortimore et al., 1988; Rutter et al., 1979), provision of resources for teachers (Hallinger & Murphy, 1986; Levine & Lezotte, 1990; Scheerens & Creemers, 1989), and significant control of personnel decisions (Levine & Lezotte, 1990; Louis & Miles, 1992).

None of the studies includes all of these aspects of leadership, and most emphasize different aspects. Academic focus and vision were the most universally agreed upon characteristic of effective school principals. Participatory leadership– the level of staff participation and other forms of shared decision-making– was one of the least agreed upon characteristics of

⁹ Although the call to examine school effectiveness using contingency models has been sounded, research using a contingency approach is scant.

effective principal leadership. Studies focusing primarily upon the effectiveness of urban, low-SES schools supported a less-participatory leadership style (Edmonds, 1979; Hallinger & Murphy, 1985), although both Rutter et al. (1979) and Mortimore et al. (1988) in studies of low-SES primary schools suggest “representative leadership”– defined by moderate participation, neither autocratic nor overly-democratic– to be most effective. To complete an already confusing picture, Miles and Louis (1992), again studying a low-SES population, showed high levels of participation to correlate with effective schooling.

The conflicting findings across different studies suggest that, at least for this aspect of leadership, a simple, direct effect does not exist. Recent research in school leadership shows this to be a complex construct. Total leadership, an additive measure of leadership across the organization, was found to impact two organizational functions, goal achievement and commitment (Pounder, Ogawa, & Adams, 1995). These functions were associated with student achievement and perceived achievement, respectively. Thus, leadership was not only shared, but also bifurcated in its effect. This is broadly consistent with theories of organizational leadership that emphasize both task and social aspects of leadership (Likert, 1967).

Correlate Two: Quality of Teaching

Few doubt the importance of quality teaching to school effectiveness. School effectiveness researchers draw heavily, if selectively, upon Brophy and Good’s (1986) influential review of effective teacher traits. This review examines the administrative structure of teaching, teaching methods, and teacher attitudes. As the following survey makes clear, Brophy and Good’s influence can be seen throughout the effective schools literature.

One aspect of the administrative structure– providing students with ample time for academic learning and sufficient learning material (“the opportunity to learn”)– correlates with student achievement across several studies (Creemers, 1992; Levine & Lezotte, 1990; Mortimore et al., 1988; Teddlie & Stringfield,

1993). While this is not a teacher trait or method, it appears to be a necessary condition for effective teaching and learning.

Teaching methods associated with effective schools include a structured approach to teaching, which includes both clear objectives and rationale for each aspect of the lesson taught (Mortimore et al., 1988; Scheerens & Creemers, 1989). Other teaching methods associated with effective schools include direct instruction— the time spent in direct contact with students— either in whole group instruction or one-to-one tutoring/monitoring (Mortimore et al., 1988). One assumes that the antithesis of direct instruction is a host of “negative” teacher behaviors including over-reliance on video, excessive seatwork, or under-managed project-based work.

Other teacher methods correlated with student achievement include an emphasis on basic skills (Edmonds, 1979), intellectually challenging work (Mortimore et al., 1988), and corrective feedback (Creemers, 1992). An emphasis on basic skills may seem to contradict intellectual challenge. Indeed, Teddlie et al. (1989) have framed this contradiction in contingency terms, suggesting that an emphasis on basic skills may be more suitable for schools serving low-SES populations. An alternative explanation suggests that a tautology exists: Students taught basic skills do better on standardized tests of basic skills (Scheerens & Creemers, 1989).

“Teaching to the test” is not a problem (i.e. does not threaten test validity) if the test assesses what is actually valued (Wiggins, 1993b). Most standardized tests, however, serve as a weak proxy for what is valued. It may be true that students who have grasped higher-order concepts exhibit greater facility with basic skills. However, an emphasis on basic skills does not necessarily facilitate a greater understanding of higher-order concepts (Perkins & Blythe, 1994). Teaching to a test of basic skills may increase student performance on these tests, but it is not likely that students have increased their understanding of what is valued within the discipline. Successful students’ failure to grasp key concepts within a discipline is well documented (Gardner, 1991). This provides evidence that the

content of the curriculum lacks conceptual rigor. Researchers concerned about school effectiveness need not support any further decline by highlighting an emphasis on basic skills and reinforcing what is most likely a tautological relationship between basic skills and test performance.

One more problem exists with trying to assess teaching quality through an examination of teacher methods. If, as expected, teaching methods (actions) reflect teacher attitudes toward students and their perceived abilities, then the methods may be less important than the messages received by the students. Any teaching method, if communicated to the students in a way that raises their self-efficacy or motivation to achieve, could favorably impact students. This is supported in several studies, where teacher attitudes, especially high-expectations for students, figure prominently as correlates to student achievement (Creemers, 1992; Edmonds, 1979; Levine & Lezotte, 1990; Rutter et al., 1979).

Correlate Three: School Climate and Culture

The ideal school climate, as captured by school effectiveness studies, is safe, orderly, welcoming to parents, free from conflict, and achievement oriented. Several studies have noted the apparent requirement of a safe, orderly climate for teaching and learning to be effective (Edmonds, 1979; Levine & Lezotte, 1990; Rutter et al., 1979; Sammons et al., 1995; Teddlie & Stringfield, 1993). This comes as no surprise, for as Gray and Wilcox (1995:21) note: “In 20 years of reading research on the characteristics of effective schools we have only once come across a case of an ‘excellent’ school where the physical environment left something to be desired.”

Second, the association of effectiveness with a “salient” level of parental involvement (Levine, 1992) recognizes the benefits of parental involvement (Mortimore et al., 1988) without the presumably negative consequences of ceding complete control to lay persons. Reynolds and Packer (1992) address the issue of parental involvement as increasingly important in an era where many

countries are utilizing market-based approaches. To the degree full parental choice of schools exists, schools must increasingly satisfy parents, the consumers of their “product.”

Third, teaching and learning are personal and relational processes. Conflict between students and adults or between adults at the school site threatens the establishment of meaningful relationships necessary to productive teaching and learning (Goodlad, 1984; Lightfoot, 1983).

Finally, “a press for achievement” commonly distinguishes the more effective schools (Oakes, 1989). This is generally consistent with an atmosphere of learning, marked by order. A press for achievement, like several aspects of culture, may be an artifact of leadership, or an artifact of prior student performance at the school. Like high expectations, an achievement orientation may motivate students and teachers toward increased performance.

Correlate Four: Appropriate Use of Student Assessment

Frequent monitoring of student progress is often cited as a correlate of effectiveness (Edmonds, 1979; Teddlie & Stringfield, 1993). Some, cite “appropriate monitoring” (Levine, 1992), presumably to guard against those cases where frequent monitoring interferes significantly with the time allocated to learning. Others note that effective schools are more likely to pay attention to disaggregated test results (Reynolds & Packer, 1992), suggesting that the quantity of monitoring matters less than the effective use of test results.

This last point deserves some comment. Whether the monitoring is frequent or simply appropriate, its effect on effectiveness is probably indirect. Monitoring may serve as a proxy for a school’s concern to understand how well it functions, but this interpretation is complicated by state and/or district testing mandates. The school may not care to monitor, yet do so anyway to remain in compliance and not threaten its supply of resources. Furthermore, no theoretical linkage exists between monitoring and performance. Common sense dictates otherwise:

Wiggins (1993a) offers the analogy of the doctor who frequently takes the patient's temperature expecting this to aid recovery.

Although some may argue that the frequency of testing increases the motivation to perform. Goal theory suggests that for motivation to increase, the measurement instrument must provide meaningful feedback that allows for correction of errors and increased performance (Locke & Latham, 1990). For students, standardized testing provides none of these conditions. Although the students receive a score, it does not serve any motivational function. First, students often do not see a connection between what they are learning and the test. Most tests of basic skills (those used most frequently by researchers to assess school performance) are not designed to correlate strongly with content taught in the class. Second, students have difficulty interpreting test scores. These scores are typically reported in statistical language, either as a percentile or a stanine. Third, the corrective aspect of feedback is missing. Seldom are items included in score reports, so even the student motivated to improve on the next iteration of testing does not know where to exert more effort. With many tests, some form of topic analysis is provided, but translating this into a mechanism for improvement would require significant work on the part of teachers, who themselves do not fully understand the test reports.¹⁰

Correlate Five: Site-Level Commitment to School Improvement and Professional Development

The correlate 'school improvement' differs from school improvement as a research topic as follows: The school improvement literature is concerned mainly with providing ways for schools to understand and facilitate organizational change efforts. School improvement as a correlate merely suggests that schools engaged in improvement efforts will likely fare better than those schools not involved with school improvement.

¹⁰ This observation is based on the first author's experience working directly with teachers as a workshop facilitator intent on helping teachers make more effective use of school-level data.

Several caveats are suggested within the school effectiveness literature. First, school improvement refers to improvement efforts developed, modified and/or embraced at the site-level. This does not include state or district attempts to coerce schools to change, although it does not preclude the state or district from encouraging or developing a school's interest in improvement. Site-level professional development (in-service training) serves as an indicator of a school's autonomous commitment to improvement (Levine & Lezotte, 1990). Second, effective improvement efforts are those that focus on classroom practice (Reynolds & Packer, 1992) as opposed to schedules or relationship maintenance. Third, effective schools track their improvement using a limited number of indicators (Peters, 1987) such as pupil academic progress, pupil satisfaction, and the presence of a meaningful pupil/teacher relationship (Gray & Wilcox, 1995). This last item is really more of a process indicator than a characteristic of school improvement. It is broadly suggestive of the types of recommendations contained within the school improvement literature and in this way represents the increasingly overlapping concerns of the two approaches.

We will ask schools to track a common list of indicators as a minimum criterion for evidence that schools are committed to improvement. These indicators will include: AP enrollment, LEP redesignation rate, dropout rate, course marks, student enrollment in academic courses (e.g. high school A-G courses), student attendance, and staff attendance. We expect these indicators to be joined by others that each school deems to be important. Furthermore, while we expect that schools will collect data across all eight of our basic school-level indicators, we do not expect that schools will focus attention simultaneously upon all eight. Instead, we believe it is in each school's interest to focus on a limited number of indicators as it seeks improvement. We expect that schools will begin to explore which of these eight indicators truly results in improved student performance and take action and make decisions according to their findings.

References

- Aitken, M. and N. Longford. 1986. "Statistical modeling issues in school effectiveness." *Journal of the Royal Statistical Society* 149:1-43.
- Brophy, Jerome and Thomas L. Good. 1986. "Teacher behavior and student achievement." Pp. 328-375 in *Handbook of research on teaching (3rd ed)*, edited by M. Wittrock. New York: Macmillan.
- Creemers, Bert. 1994. "The History, Value, and Purpose of School Effectiveness Studies." in *Advances in School Effectiveness Research and Practice*, edited by D. Reynolds. Oxford: Pergamon.
- Creemers, Bert P.M. 1992. "School Effectiveness, Effective Instruction and School Improvement in the Netherlands." Pp. 48-70 in *School Effectiveness: Research, Policy and Practice*, edited by D. Reynolds and P. Cuttance. London: Cassell.
- Cuttance, Peter. 1992. "Evaluating the Effectiveness of Schools." Pp. 71-95 in *School Effectiveness: Research, Policy and Practice*, edited by D. Reynolds and P. Cuttance. London: Cassell.
- Daly, P. 1991. "How Large are Secondary School Effects in Northern Ireland?" *School Effectiveness and School Improvement* 2:305-23.
- Fitz-Gibbon, Carol Taylor. 1996. *Monitoring Education: Indicators, Quality and Effectiveness*. London: Cassell.
- Fitz-Gibbons, Carol Taylor. 1992. "School Effects at A Level: Genesis of an Information System?" Pp. 154-163 in *School Effectiveness: Research, Policy and Practice*, edited by D. Reynolds and P. Cuttance. London: Cassell.
- Gardner, Howard. 1991. *The Unschooled Mind*. New York: Basic Books.
- Goodlad, John. 1984. *A place called school: Prospect for the future*. New York: McGraw-Hill.
- Gray, John, D. Jesson, and N. Sime. 1990. "Estimating differences in the examination performances of secondary schools in six LEAs: a multilevel approach to school effectiveness." *Oxford Review of Education* 16:137-58.
- Gray, John and Brian Wilcox. 1995. *Good school, bad school: evaluating performance and encouraging improvement*. Buckingham: Open University Press.
- Hallinger, P and J Murphy. 1985. "Instructional leadership and school socio-economic status: a preliminary investigation." *Administrator's Notebook* 31:1-4.
- Hallinger, P and J Murphy. 1986. "The social context of effective schools." *American Journal of Education* 94:328-355.

- Levine, Daniel U. 1992. "An Interpretive Review of US Research and Practice Dealing with Unusually Effective Schools." Pp. 25-47 in *School Effectiveness: Research, Policy and Practice*, edited by D. Reynolds and P. Cuttance. London: Cassell.
- Levine, Daniel U. and Larry W. Lezotte. 1990. *Unusually Effective Schools: A Review and Analysis of Research and Practice*. Madison, WI: National Center for Effective Schools Research and Development.
- Lightfoot, Sarah L. 1983. *The Good High School: Portraits of Character and Culture*. New York: Basic Books.
- Louis, Karen Seashore and Mathew Miles. 1992. *Improving the Urban High School: What Works and Why*. London: Cassell.
- Mortimore, Peter. 1992. "Issues in School Effectiveness." Pp. 154-163 in *School Effectiveness: Research, Policy and Practice*, edited by D. Reynolds and P. Cuttance. London: Cassell.
- Mortimore, P., P. Sammons, R. Ecob, L. Stoll, and D. Lewis. 1988. *School Matters: The Junior Years*. Salisbury: Open Books.
- Nuttall, D., H. Goldstein, R. Prosser, and J. Rasbash. 1989. "Differential school effectiveness." *International Journal of Educational Research* 13:769-76.
- Oakes, Jeannie. 1989. "School Context and Organization." Pp. 40-65 in *Indicators for Monitoring Mathematics and Science Education*, edited by R. J. Shavelson, M. L. McDonnel, and J. Oakes. Santa Monica: Rand Corporation.
- Perkins, David and Tina Blythe. 1994. "Putting Understanding Up Front." *Educational Leadership* February:4-7.
- Peters, Tom. 1987. *Thriving on Chaos: Handbook for a Management Revolution*. New York: HarperCollins.
- Pounder, Diana G., Rodney T. Ogawa, and E. Ann Adams. 1995. "Leadership as an Organization-Wide Phenomena: Its Impact on School Performance." *Educational Administration Quarterly* 31:564-588.
- Reynolds, David and Anthony Packer. 1992. "School effectiveness and school improvement in the 1990's." Pp. 171-188 in *School Effectiveness: Research, Policy and Practice*, edited by D. Reynolds and P. Cuttance. London: Cassell.
- Rutter, M., B. Maughan, Peter Mortimore, and J. Oustoe. 1979. *Fifteen Thousand Hours: Secondary Schools and their Effects on Children*. London: Open Books.
- Sammons, Pam, Josh Hillman, and Peter Mortimore. 1995. "Key Characteristics of Effective Schools." University of London Institute of Education, London.

- Scheerens, Jaap and Bert P.M. Creemers. 1989. "Towards a more comprehensive conceptualization of school effectiveness." Pp. 265-278 in *School Effectiveness and School Improvement*, edited by B. Creemers, T. Peters, and D. Reynolds. Amsterdam: Swets & Zeitlinger.
- Teddlie, Charles and Sam Stringfield. 1993. *Schools Make a Difference*. New York: Teachers College Press.
- Wiggins, Grant. 1993. *Assessing student performance: exploring the purpose and limits of testing*. San Francisco: Jossey-Bass.
- Zigarelli, Michael A. 1996. "An Empirical Test of Conclusions from Effective Schools Research." *The Journal of Educational Research* 90:103-110.

Appendix C: Indicators of Classroom Practice

The proposed indicators of classroom practice are based on the work of Pamela Aschbacher and Lindsay Clare at UCLA's Center for Research on Evaluation, Standards, and Student Testing (CRESST). This work has resulted in the refinement of a framework for the evaluation of classroom practice and the development of corresponding instrumentation (i.e. rubrics). Initial results from a pilot test of the instruments are encouraging. Teaching assignments were rated on the framework's six dimensions: challenge of the task, clarity of the teacher's goals for student learning, clarity of the grading criteria, alignment between goals and task, alignment between goals and grading criteria, and overall quality. These results show reasonable inter-rater reliability and generalizability.

We believe that the proposed indicators reflect aspects of teaching practice that are highly valued within Los Angeles Unified School District. We are confident that this framework can be utilized within our district to assess the quality of classroom assignments.

Abstract from the Technical Report:

This report describes the development of indicators of classroom practice for monitoring and improving the quality of school reform. The work entailed development of a rubric to rate key facets of classroom practice based on assignments and samples of student work. This approach was used to describe the intellectual challenge of class assignments, the alignment of tasks with learning goals and grading criteria, clarity of criteria for success, and provision of informative feedback to students. It also compared teacher judgments of student work with external rater judgments using a school district's standards-based rubric. The study demonstrated use of this methodology within an evaluation of a complex urban reform initiative. Inferences from the data were analyzed for their technical quality and usefulness. Overall, the technical quality of the approach was reasonable, but anchor papers have been selected and the rubric refined to improve future generalizability. The indicators show promise for use in school or district self-evaluation efforts, not only in monitoring progress but identifying areas for administrative attention, professional development, and teacher reflection.

Appendix D: Changes to the SOI Protocol for 2001-2002

We are using the pilot data collection for the 2000-2001 school year to guide us in implementing stronger protocols for the full project implementation in 2001-2002. During the pilot several issues arose for individual respondents that could be addressed fairly directly given the relatively small scale of the project and the direct involvement of the Program Evaluation and Research Branch during the second set of data collection efforts. Next year the magnitude of the project will make customized responses much more difficult. It is, therefore, important to have fully developed procedures that will deal with most situations likely to arise.

These issues have been generally grouped into issues involving community support, coordination between respondent groups, data collection coordination, respondent eligibility, language access, and respondent burden. Each topic is reviewed in more detail below. This review is not intended to be a guide to the data collection plans for 2001-2002 as much as a review of the issues and intended responses to those issues.

Community Support

Parents are bombarded in their homes with calls to choose long distance companies and to help with marketing research. Teachers are asked to complete documentation of various sorts for special projects or record-keeping efforts. Principals often find it difficult to monitor academic issues given the level of administrative duties they must perform. Without community support, gaining cooperation from these stakeholder groups becomes more difficult. With an understanding of the value of the project and a clear understanding of the effort requested from each individual, respondents with full, busy lives and their own personal priorities are more likely to make the effort to accommodate our data collection.

Community members typically become aware of school events or programs through regular means of communication, such as parent newsletters, staff

meetings, mailings and school announcements. In order to discuss the project in these ways, we should develop principal awareness and support. Principals are not only likely to be contacted directly by teachers and parents with concerns, but they can also help to disseminate information prepared by the Program Evaluation and Research Branch for inclusion in various newsletters and memos going to the school community.

Data Collection Coordination

Data collection for each group of respondents will be generally independent from each other. However, as all of the stakeholder groups interact and have formalized relationships within the school community, it will be important for our survey implementation to acknowledge that fact. For example, students and parents should learn about our project in an appropriate sequence. Another example is noted above; principals should have a broad understanding of our data collection efforts before they are likely to be contacted by concerned parents and teachers.

We discovered that the one-month data collection period used was insufficient, a major factor in the low response rate encountered by Harris Interactive. We also plan to use multiple data-collection methods to increase response rates. Multiple data collection methods will help to accommodate the time constraints of our various respondent groups and to limit the expense of extending the data collection effort. In each case, we will begin with a written version of the survey followed by telephone interviews.

Respondent Eligibility

All of our respondents were identified through school district records held at the Central District offices. Since the source of this information is distant from the school site, inaccuracies are difficult to identify. The lives of the school members are constantly in flux. Teachers move between schools, so do families or individual children. The Program Evaluation and Research Branch used the

best information available. Even so, there were inaccuracies. Some of our efforts in the pilot involved identifying the kinds of inaccuracies found in the data sets used to draw the sample for this project. As we are trying to develop measures for specific schools, it is important to have eligibility rules in place so that we identify all of the individuals to interview for each school. We will design our initial mailings with these issues in mind in order to enhance our ability to correctly determine respondent eligibility.

Other eligibility issues arise from scheduling complexities within LAUSD. Because of the various school calendars and tracks, timing our survey is difficult. We cannot simply assume that an October data collection means that we are getting responses from classes that have been together for roughly the same length of time. For student and teacher respondents in particular, the timing of the survey relative to the individual's track calendar can affect their responses. For example, questions about recent interactions in the classroom may differ for classes early in the year, late in the year or during testing. Rather than disregarding or under-representing particular tracks by only selecting those running during a limited data collection period, we want to include those randomly selected into our samples as fully as possible by acknowledging their different circumstances and making appropriate adjustments.

Language Access

LAUSD schools are generally managed and taught in English (only one teacher in our pilot asked for a non-English survey). However, most of our students use other languages at home. Only about 30% have English as their home language. Approximately 63% of LAUSD students with an identified home language are from Spanish-speaking homes. Armenian, Korean, and Tagalog are each the home language of approximately 1% of the LAUSD students according to the SIS data. The primary impact of our linguistically diverse community on the School Organization Index is on the parent interview component. Vendor capacity in Spanish and any other languages selected for interviews will influence the selection process substantially.

It is unclear how many of our parents are sufficiently fluent in English to complete the survey in English. As neighborhoods are frequently comprised of individuals with the same language background, even the small percentages of the overall district using these other languages may be quite significant in selected schools. The Program Evaluation and Research Branch will consider language-clustering effects when determining if the parent survey needs to be conducted in other languages in addition to Spanish and English. We will be looking for information from schools and local districts to help us identify whether this is a significant need.

In general, students from homes where English is not the primary language are sufficiently fluent in English by the time they are in a secondary school program to complete an English survey. However, we did identify 5% of the students interviewed by PERB interviewers whose ability to understand the survey was limited by their lack of fluency. They were all native Spanish-speakers. As with the parents, PERB will be looking at District data on language use and will welcome input from schools and local districts to determine if the student survey should be translated into Spanish.

Respondent and School Burden

Our objective is to benefit the education of students in the Los Angeles Unified School District. If our data collection burdens the schools or the school community in a way that interferes with education, our efforts are disrupting the very process we want to benefit.

Our concern for the effect of our process on the school community impacts all of our procedures. When developing the various survey instruments, we have focused on limiting the response burden for each respondent group. We did this by selecting questions that were most likely to provide information that directly assesses or impacts the quality of teaching and learning. The questions were also selected for the relevance toward District policies in order to provide

information that will be useful to schools and local districts for their own planning.

The initial mode of data collection will be a self-administered survey delivered by mail or available on the web and completed at the respondent's convenience. We will use the schools' calendars to determine the timing for our mailings. However, we are also going to include interviews in order to provide a mode of data collection some individuals find more convenient. We will notify adult respondents that telephone interviews will follow for those who do not complete the initial survey. This will allow them to choose the method they find most convenient and least intrusive for themselves.

Respondent burden for the student survey is of a different nature. Students tend to have very busy schedules, but often not under their own control. They have classes and school activities, and sometimes work schedules. At home, still other duties compete for their time. With the students, we plan to distribute the survey during non-academic time periods at the school (lunch, homeroom, nutrition, etc.), allowing ample time to complete the survey at their leisure and a convenient means to return the survey. Since the students are generally very willing to complete the survey, our goal will be to minimize the burden on the school and classrooms for the distribution and collection of the surveys.

As the Local District Performance Measures will be implemented annually, we hope to develop positive respondent relations and build upon them year to year.