Lessons Learned in Using Data to Support School Inquiry and Continuous Improvement: Final Report to the Stuart Foundation

CSE Technical Report 535

Joan Herman and Barry Gribbons

Center for the Study of Evaluation (CSE)
University of California, Los Angeles

February 2001

Center for the Study of Evaluation
National Center for Research on Evaluation,
Standards, and Student Testing
Graduate School of Education & Information Studies
University of California, Los Angeles
Los Angeles, CA 90095-1522
(310) 206-1532

Joan Herman and Barry Gribbons, Project Directors, CSE/CRESST, University of California, Los Angeles

The research contained herein was supported by grant number 97-167 from the Stuart Foundation to the Center for the Study of Evaluation. The findings and opinions expressed in this report are those of the authors and do not necessarily reflect the position or policies of the Stuart Foundation or the Los Angeles Educational Partnership (LAEP).

Preparation of the report was supported in part under the Educational Research and Development Centers Program, PR/Award Number R305B60002, as administered by the Office of Educational Research and Improvement, U.S. Department of Education.

The findings and opinions expressed in this report do not reflect the positions or policies of the National Institute on Student Achievement, Curriculum, and Assessment, the Office of Educational Research and Improvement, or the U.S. Department of Education.

LESSONS LEARNED IN USING DATA TO SUPPORT SCHOOL INQUIRY AND CONTINUOUS IMPROVEMENT
FINAL REPORT TO THE STUART FOUNDATION

Joan Herman and Barry Gribbons

Center for the Study of Evaluation (CSE)
University of California, Los Angeles

Introduction

Accountability and assessment are at the heart of today's reform efforts. Strong accountability mandates, with heavy stakes attached, are sweeping the country at every level: federal, state, and local. No longer satisfied with business and student achievement as usual, the public and its policymakers are demanding that schools focus on achieving high standards for *all* children, and they are requiring evidence of progress toward those standards. The mandates reflect new expectations for learning, as well as renewed commitment to closing the traditional achievement gap between rich and poor, between students entering school with and without full English proficiency, and for other special subgroups.

Committed to enabling *all* students to achieve, schools are expected to use data to understand their students' academic standing, and to establish improvement plans accordingly. They also are expected to chart the effectiveness of their strategies, and to use assessments to monitor and assure progress. No longer are decisions made based solely on anecdotal experience, personal preference, or historical precedent. Rather, data-based decision making and continuous improvement are the current operating concepts.

Yet, despite both the mandates and the rhetoric, schools are woefully underprepared to engage in such inquiry. There may be a wealth of potential data available to schools, a growing number of tests mandated at the state, local, and school levels for various purposes: dropout statistics, attendance figures, course enrollments, SAT and ACT scores, and results of teacher and parent surveys. But such data are not easily accessible or available for comprehensive analysis. Looking at the whole picture and understanding what the results mean for the progress of individual students or the performance of special subgroups is difficult. The practice of integrating and applying large-scale data with that from classroom practices for analysis is virtually nonexistent. Furthermore, despite the precepts of current

standards-based reforms, the power of decentralization and the importance of teacher empowerment, there is little in most educators' background or training that prepares them to engage in systemic inquiry. Rarely does teaching rhetoric include program planning, performance-based decision making, or the intricacy of data collection, analysis, and interpretation. These are new principles in the culture of most schools.

How can schools move ahead in meaningful ways and use assessment and accountability to improve education for *all* students? How can schools serving large proportions of economically disadvantaged and ethnically diverse populations use data to break the cycle of poor performance and educational failure? This paper reports on preliminary answers to these questions. It is based on our work, in a number of schools and districts in southern California, facilitating and developing schools' capacity to use information for planning and decision making. These efforts included:

- technical assistance to the management teams from three urban school reform projects in support of their inquiry processes. The assistance included helping the teams to articulate data-based questions, access data, and generate reports that could inform their planning;
- direct training to school teams composed of administrators and teachers, to deepen their capacity to engage in school-based evaluation, and aid in continuous improvement;
- collaboration with administrators and teachers in several schools to help them better use data to improve their programs for students.

While the settings and context were varied, each project involved a similar process of data introduction and use, which we describe briefly below. We focus on the results of our work in two settings. Following a brief description of these two "case studies," we discuss general lessons learned, based on our entire experience, and conclude with our analysis of next steps, including core tensions that must be alleviated to move ahead.

Our Data-Based Inquiry Process

In each instance, our data inquiry process started with currently available data, both because of their accessibility, and because these are the data that districts and schools are required to use and base their progress upon. Furthermore, we were interested in better understanding how such data can inform improvement and their limits in doing so.

Available Data

The large districts with which we worked had abundant data available in their student information systems. Although the content of these databases varied slightly, and the ease of access varied substantially, the essential contents of current-year district databases were the same: demographic data, program participation data (e.g., Title I, gifted, special education), scores from state- or district-required tests, and language proficiency data. At the secondary level, districts also tended to have attendance data (both excused and unexcused) and course history data. Table 1 shows the common, current-year data elements we found in our two districts. By using a common student identification number, we were able to merge current data with archived files from prior years to construct longitudinal databases that enabled us to examine important issues of student progress and performance over time. We underscore that it is possible to fairly easily construct such longitudinal databases from available electronic data in most school districts with reasonably recent computer systems, as long as the district has a unique ID number for every student. Problems arise in districts without digitized information systems or those operating with older systems. More difficult is bridging confidentiality and security issues and gaining sufficient trust and credibility for districts to release the data.

Typically, the schools we worked with also had a range of various kinds of other formal and informal data about students that tended not to be electronically stored. These included classroom data, such as quizzes and tests, assessments of student work, and special projects. Schoolwide test or other data are often also present. For example, one school had administered diagnostic placement tests to all entering freshmen. Presumably the tests had been scored and used for placement purposes, but results had never been entered into student records. Another school systematically collected and discussed teacher assignments and student work in order to examine the alignment between standards and teaching, the strengths and weaknesses of student performance, and the ways to improve instruction, but these data were never formally summarized or recorded in any way.

There were a number of reasons why we were not able to use these sources of data in our work. First, obviously data in nonelectronic form are not easily accessible for the types of analysis we were interested in investigating (without additional cost and time-intensive data entry). This is a tractable problem. Second, and more important, is the issue of data quality. The reliability and validity of school-based measures are often suspect because they are not the product of a careful

Table 1

Sample Database

Field no.	Field name	Descriptions
1	STUID	Student's ID
2	SCODE	School the Student Is Currently Attending
3	GENDER	Student’s Gender
4	ETHNIC	Student’s Ethnicity
5	BCOUNTRY	Student's Birth Country
6	USSCH	Date First Entered US School
7	GRADE	Current Student Grade Level
8	MEAL	Meal Program Eligibility (Free, Reduced, Full Pay)
9	GATE	Gifted and Talented Education
10	SPED	Special Education Class Code
11	LCODE	Leave Code
12	LDATE	Date Student Left School
13	TOSCH	New School Location Code
14	TVPROG	Participant in Traveling Program
15	RESSCH	Student's Resident School Location
16	TRRE	Student’s Transfer Reason
17	HLC	Home Language
18	LANG	Language Proficiency
19	BILASS	Bilingual Program Assignment
20	RDESIG	Date Redesignated
21	RTPTLE	Reading Total Percentile (English)
22	MTPTLE	Math Total Percentile (English)
23	LTPTLE	Language Percentile (English)
24	SCPTLE	Science Percentile (English)
25	SOPTLE	Social Science Percentile (English)
26	SPPTLE	Spelling Percentile (English)
27	RTPTLS	Reading Total Percentile (Spanish)
28	MTPTLS	Math Total Percentile (Spanish)
29	LTPTLS	Language Percentile (Spanish)
30	ATTEND	Number of Days Attended
31	APNABS	Number of Days Approved Absences
32	NAPABS	Number of Days Non-approved Absences
		(Repeat Fields 33 to 40 Ten Times)
33	CRSNO	Course Number
34	CRSTIT	Course Title
35	MARK	Grade Received
36	SEMES	Semester
37	SCODE	School Location Code
38	AF	A-F Course Flag
39	AP	AP Course Flag
40	TCHRID	Fictitious Teacher's ID

development and review process. Much of what is available is idiosyncratic to individual teachers, and so cannot be aggregated and compared across classrooms. As a result, most of our inquiry and subsequent analyses focused on the uniform data, which admittedly left them at some distance from actually classroom curriculum.

Initial Empirical Questions

Where to start in raising questions of available data that could inform improvement? The number of possible questions is seemingly infinite. However, we identified some initial basic questions that are foundational for most school inquiry. We started with these questions and then moved to additional data collection and analyses to try to discover why things were as they were and to inform subsequent improvement plans.

Available data enabled our schools to initiate inquiry on three basic kinds of questions:

- How are we doing?
- Are we well serving all students?
- What are our relative strengths and weaknesses?

Data from these three questions inform a fourth set of questions at the core of the inquiry process: *Why* are things the way they are? *How* can we make them better? *What* are the implications of the data for improving teaching and learning? Note that the answers to these questions can only be informed by empirical analysis, but not directly answered by it. Good answers to the "why?" and "what next?" questions require the integration of data with the working knowledge, observations, and professional judgments of the school community.

How are we doing? "How are we doing?" questions ask how schools, and most importantly, the students within them, are performing on a range of learning and other important student outcomes. Available indicators of such outcomes typically include standardized test scores (reading, mathematics, language, spelling, science, and social science), other content assessments, attendance rates, redesignation as English proficient, completion of courses required for admission to the University of California, and completion of Advanced Placement courses.

Schools can get a sense of how they are doing on these various indicators by analyzing their students' performance relative to several comparison points:

- How are we doing compared to pre-established goals or expected standards? For example, have we achieved the goals we set for this subject area last year? Have we made adequate progress, based on the state or district formula?
- How are students doing relative to how they have done in the past? For example, is student performance getting better, worse, or staying the same? Are more students scoring at the proficient level compared to last year?
- How are students doing relative to others' performance? For example, how are students in our school doing compared to similar students in other schools, compared to the district as a whole, and compared to the state?

Each of these comparison points provides a slightly different view on student performance, and together they may provide schools with perspective for establishing realistic goals for the future—goals that require stretch, but that also are feasible and can be accomplished realistically.[1]

Note that the issue of whom to include arises in all of these analyses. Should all students be included or should the analysis be limited to only students who have attended the school for some minimum period of time (e.g., one or two years)? Certainly, schools are responsible for all their students and for purposes of needs assessment, all students should be included. However, when the question is "How effective is our school program?" there is merit in limiting the analysis to those students who have attended the school long enough to be influenced by the school program.

Are we well serving all students? The second set of questions takes another look at the student performance question by examining results by subgroup—ethnic, gender, language status, poverty, and special program participation (Title I, students with disabilities, etc.). Comparing the absolute level of performance across various subgroups provides one essential set of analyses—for example, are boys and girls achieving standards at similar rates; is their performance similar in various subject areas?

Examining subgroup progress over time represents another critical set of analyses. Are we closing the gap between poor and economically advantaged students in the various subject areas? Are limited English proficient (LEP) students making progress at rates similar to those of fully English proficient students?

[1] This, of course, assumes that schools are free to establish their own goals. The California Academic Performance Index essentially establishes goals for schools—unless they want to set their sights higher than those established by the state.

Comparing the performance of various subgroups in a particular school to those in similar schools, to the district as a whole, and to the state also provides additional perspective for judging the school's performance relative to equity goals.

What are our relative strengths and weaknesses? In the process of looking at how well students are doing, schools naturally can look for areas of relative curriculum strength and weakness—both across curriculum subject areas and within each subject area—that can help establish priorities for school improvement. Across subject areas, one looks to see whether there are consistent patterns, across grade levels and over the years, of students doing relatively better in one subject area than another. For example, are students' scores in language arts consistently higher than those in mathematics? Within each subject area, performance by grade level is of interest: Do some grades stand out as relatively high or low performing? Again, such comparisons can help establish priorities, for both future inquiry and future improvement efforts.

Similarly, once having identified a curriculum area as a relative priority, it is wise to delve into greater detail in that area. Where subscale performance is available (e.g., computation or problem solving in math, phonics or comprehension in reading), schools can get a diagnostic reading on particular subject areas in need of boosting.[2] Comparing performance on multiple measures within a subject area also can provide a sense of relative strengths and weakness. For example, if results on the direct writing sample measure are low and the Stanford Achievement Test, Version 9 (SAT-9) results are in the mid-range, that might suggest written communication as an area of relative weakness. Exploring performance in writing in greater detail might reveal that students' writing was judged low in focus and grammar, suggesting these as priorities for the future, which in turn could bolster the SAT-9 scores.

Why? Delving into detail about performance within a subject area is one way of discovering the reasons for overall performance. There are innumerable other analyses that could inform the "why" question, for example:

- Is it an attendance problem? What is the relationship between attendance and performance? Do students who are absent more than 12 days per year show a markedly different level of performance than other students?

[2] In examining subscale performance, it's important to consider the reliability of the subscale. If composed of only a few items, subscales can be quite unreliable and thus provide a false sense of diagnostic information.

- Are programs for English language learners (ELL) an issue? What proportion of ELL students are redesignated after one year? After two years? How many of the limited English proficient (LEP) students who have been at the school and have had strong attendance are never redesignated?
- Are all students being given the opportunity to learn and excel? Are all students from all subgroups equally likely to be enrolled in college preparatory course work? What proportion of minority students versus White students are enrolled in Advanced Placement (AP) courses? How do rates of enrollment in A-F courses compare for Black and Latino students who perform well on the SAT-9 as opposed to White students who perform similarly well?
- Are some of our special programs working better than others? For example, how do students who participate in the afterschool program perform compared to those who are tutored, or to similar students who do not participate in any program? Does the performance of students who attend summer school improve after summer school? Do they maintain the improvement over time?

These questions are difficult to prespecify or establish, arising as they do from educators' hypotheses about specific, important issues at their schools. Similarly, the data results are rarely prescriptive in informing schools what to do next. The questions and the data tend to be instead good starting points for understanding the status quo and engaging key constituents in further discussion and inquiry. Most useful are questions that encourage schools to focus on their school programs and curriculum as they relate to identified priorities for student learning. More important still are the plans and actions that follow from the inquiry.

The Process of Inquiry in Two School Groups

In the following section, we describe our work with two school groups that were involved in similar data inquiry processes, resembling that described above (in superficial features). In both cases, we started the data inquiry process with available data and attempted to merge our efforts with ongoing school improvement planning. The two groups moved in quite different directions, however, and experienced different levels of success.

Although school group A was located in an affluent community, because of overcrowding in other parts of the district and a voluntary bussing program, it served a diverse student population. The group was composed of a high school and its feeder middle and elementary schools. School group A's test scores on the state assessment program were relatively high. School group A's team was led by the

high school assistant principal, who was an expert in curriculum and instruction and enthusiastic and knowledgeable about the use of data to improve the schools' offerings. The team, composed primarily of teachers, but also including other administrators, was instituted to serve overall school planning and improvement functions. It met regularly.

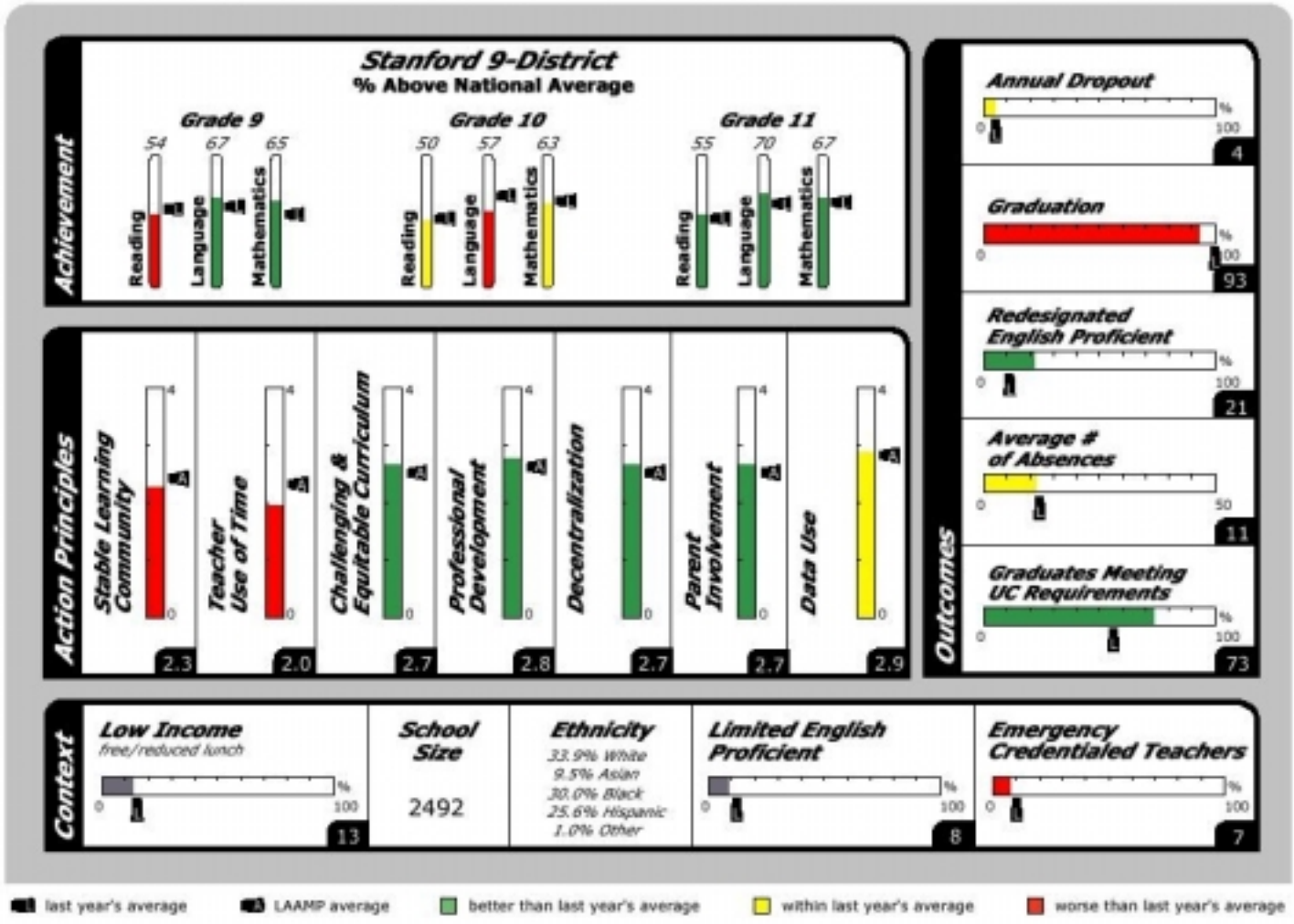
School group B, in contrast, was located in an economically impoverished area and served primarily Latino and African American students. The school management team was headed by the principal and composed of administrators, although other subject area teams composed of teachers and administrators were part of the planning and improvement process. Test scores in school group B were low, and the principal felt under enormous pressure to improve them. Management team meetings were less frequently scheduled in school group B, and the first round of data inquiry and analysis involved only the principal.

Initial Data Analysis and Reporting: A Comprehensive School Report Card

Initial data analysis and reporting for both groups addressed the first two questions described above: How are we doing? Are we well serving all subgroups? Answers were provided using simple descriptive statistics derived from the district's longitudinal student database, and in some cases, these data were combined with information available from the state's CBEDs database, as well as other available data. In presenting this first-cut of the data, we used a school report card format that would be engaging and communicative for the entire school community, including parents and community members. Our intent was to provide a report that schools could use to communicate their progress to their community, and that would initiate discussions of needs and goals with representative school site councils.

Through the school report cards, we invited the schools to first examine in one glance a variety of indicators of student performance. These included SAT-9 results, SAT results, attendance rates, dropout rates, percent completing challenging courses (courses required for eligibility to be admitted to the University of California), and graduation rates (see Figure 1). The reports showed performance compared to the previous year and to the district (or to special programs within it). SAT-9 results were a particular focus of attention, since they are the principle measure of student learning used by the public and policymakers.

FIG1. Sample High School



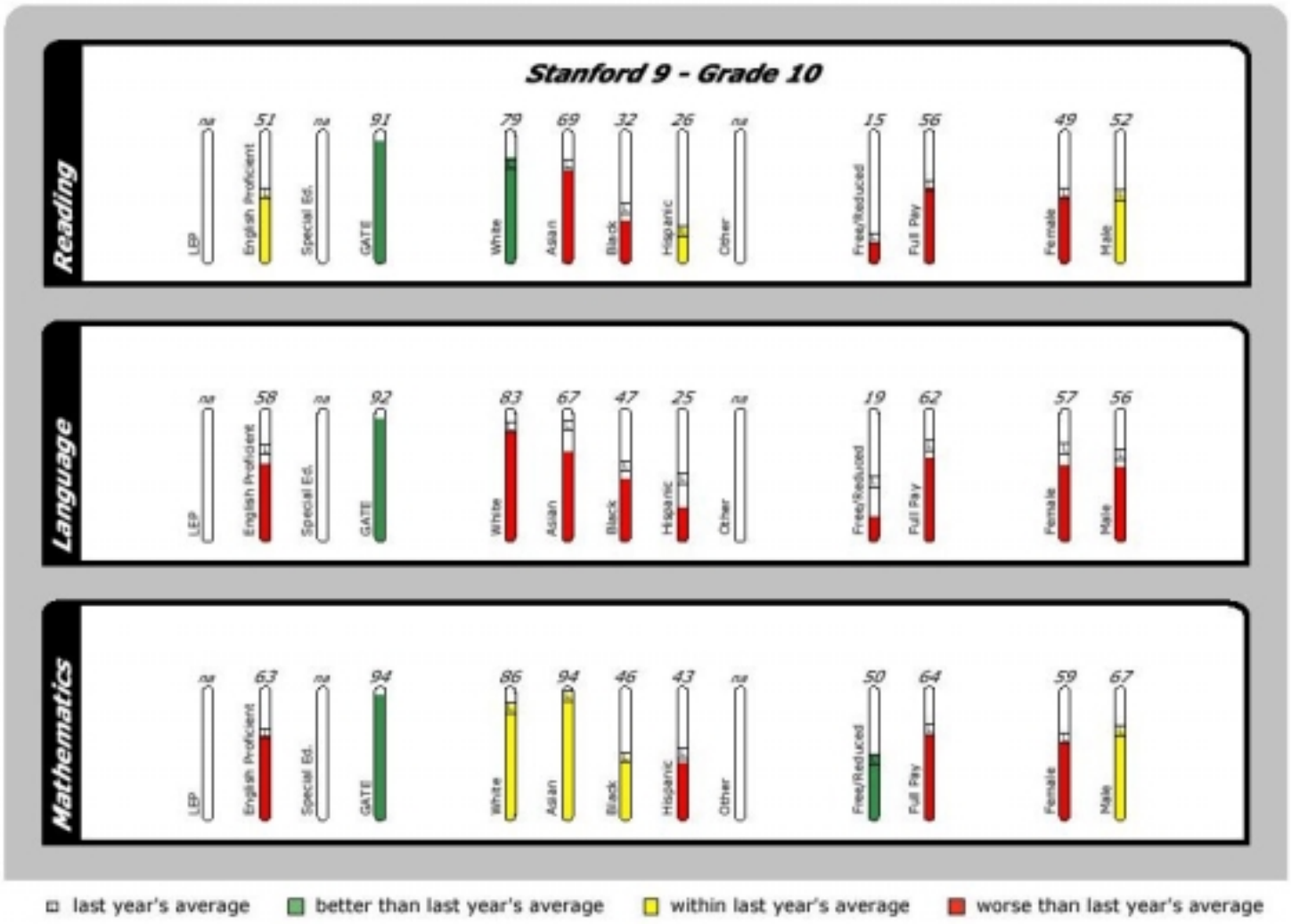
Note that in addition to performance indicators, the reports provide information on contextual indicators useful in interpreting results. Examples of these indicators currently available include demography, school size, ethnic distributions, percent of low-income students, percent of limited English proficient students, and percent of credentialed and experienced teachers. Because the schools participating were also part of the Los Angeles Annenberg Metropolitan Project (LAAMP), indicators of school processes related to reform were also available. These included

- Stable Learning Community—the degree to which there is a well-articulated program within and across grades K–12;
- Teachers' Use of Time—a measure of reallocation of teacher time so that teachers could collaborate in developing curricula and tune instructional strategies by learning from one another;
- Challenging and Equitable Curriculum—teachers' expectations for students vis-à-vis a rigorous curriculum;
- Professional Development—the extent to which teachers viewed professional development as useful and attuned to student needs;
- Decentralization of Decision Making—the extent to which teachers felt involved and shared a vision of reform at the school level;
- Parent Involvement—the extent to which parents participated in school and classroom activities; and
- Using Data to Inform Planning and Decision Making—the extent to which data were involved and used for planning and monitoring progress.

Similarly, Figure 2 shows how various subgroups performed on the statewide assessment, using one of several possible approaches to disaggregating data. All of the categories included in this figure—limited English proficiency, Special Education, Gifted and Talented Education (GATE), ethnicity, income level, gender, and migratory status—are required by Title I of the Improving America's Schools Act (IASA).

When all of these indicators—contextual, process, and performance—are included in a school report card, educators and the community have a fuller picture of the quality of schooling at their sites. By grouping the information, and using color strategically, schools can answer a number of basic questions about their status and progress on available measures, in the space of a page or two. This "information-at-a-glance" strategy has limits, of course; and it is the discussions, questions, and subsequent analysis it stimulates that are most important.

FIG2. Sample High School



School Group A

To no one's surprise, school group A (composed of a high school and feeder middle and elementary schools) found that students across their schools were performing well relative to the national average and to the district. When examining the subgroup data, however, the team found that some groups of students, particularly economically disadvantaged minority students and limited English proficient students, were performing poorly—confirming prior preconceptions. These poor-performing subgroups in general corresponded to students who were being bussed into the school from other parts of the city.

Exploring school-level trends. Having found large differences between students bussed to the school and "local" students, the school posed additional questions to assess reasons for the differences and what they might do to respond. For example, first they examined the actual feeder pattern of students (see slide 1 in Appendix A). This slide clearly shows that many students at the high school and middle school do not attend the traditional feeder schools, raising questions about the consistency of curriculum for these students across grade levels. Focusing on the high school, they examined the performance of local students compared to bussed students on standardized achievement tests (see slides 2 through 4 in Appendix A). Clearly, students not from the local community performed more poorly than did local students.

With this information, the team decided to do additional analyses. They hypothesized that students who had attended the local middle school might perform better than students from other areas. They reasoned that the quality of education might be better at the local middle school compared to other middle schools in the district, and that the curricula in the two schools might be more closely aligned. Implicitly, the overall nature of the student body was different at the local versus other middle schools. For comparison, the team created four groups of students:

- Local, local middle school (students living in the local area who had attended the local middle school);
- Local, non-local middle school (students living in the local area, who had *not* attended the local middle school);
- Non-local, local middle school (students bussed to school, who had attended the local middle school); and

- Non-local, non-local middle school (students bussed to school, who had *not* attended the local middle school).

Across these four groups, the team engaged in a series of analyses of different performance indicators that moved them progressively closer to curriculum and teaching issues.

These other indicators were intended to help account for or explain the large differences in test scores the team observed between groups, and their selection was based on team members' hypotheses about what the source problems might be. The team started with absenteeism (is there a relationship between attendance and performance?) and continuity at the school (how long do students continuously enroll at the high school?), reasoning that non-local students traveling considerable distances to the high school might be absent more often than local students. Students living long distances and requiring a long bus ride to school daily might also be less likely to remain at the school. They might enroll for only a year or two, rather than the full four years, and thus not benefit from a continuous curriculum. As illustrated in slides 5 through 8 of Appendix A, local students who had attended the feeder middle school were absent less often and attended the school longer than students who traveled long distances to the school. Interestingly, students who were bussed to the school, but also attended the feeder middle school persisted at the high school longer and had fewer absences than local students who did not attend the feeder middle school. Although this alone does not indicate why some students performed more poorly than others, this information was useful in encouraging the planning committee to probe further into what happens to students as they moved through the high school.

Moving to a focus on mathematics teaching and learning. Taking another step closer to the classroom, the group began looking at course-taking patterns and examining math progress at the high school level. The group was particularly interested in improving students' learning and having more students successfully complete the mathematics coursework requirements for University of California admissions and Advanced Placement (AP) courses. They wondered whether students were making regular progress in the required coursework (algebra, geometry, calculus, etc.) or whether particular courses were causing problems.

These questions required the development of new curriculum progress indicators using course history information in the district longitudinal database.[3]

First, they assessed the percentage of students who had taken Advanced Placement (AP) courses (see slides 9 through 11 in Appendix A). As illustrated in these graphs, local students were taking more challenging AP courses compared to non-local students. Furthermore, this gap increased over time. Considering only calculus, the team examined patterns of enrollment in math courses over time (slides 12 through 23 in Appendix A).

They requested a more in-depth look at how students were progressing through the mathematics curriculum. They found that in students' freshman year at the high school, many more local students were enrolling in challenging math courses, such as Integrated Math 1 and 2 (analogous to algebra and geometry in a traditional math sequence). Students who were bussed to the high school, in contrast, were more likely to be enrolled in lower level courses. Over time, the non-local students did not catch up in the math sequence to enable them to take calculus by their senior year. By their junior year, many more non-local students were still taking Integrated Math 2 or lower level courses.

These curriculum progress reports showed not only differential progress for bussed and local students but particular problems at the 10th-grade level. Investigating further, the team looked for an explanation at the feeder school level—that is, were students from the local junior high school and the bussed students who had attended that school performing differently than students from other non-local junior high schools? The answer was yes, and while the question provokes concerns about whether the high school team was attempting to place responsibility elsewhere, the team ended up discovering an important curriculum alignment problem. The high school offered a nontraditional math sequence. Although students from the local middle school took Integrated Math (the first course in the sequence), students from other areas might have taken algebra in their

[3] These indicators required more complex data management than the standard outcome indicators. They required special programming using standard statistical packages such as SPSS or special programs written in languages such as FORTRAN. Essentially, these programs were used to create new indicators that identified the number of AP courses enrolled in and passed, the math courses passed, and how many of the six (A-F) requirements necessary to be eligible for admission to the University of California were passed. For example, we created an indicator for math analysis. Initially, it is set to zero (not passed). Then, we move through the entire course history for each student, checking to see whether both math analysis courses had been taken, and whether the grade was a "C" or better. This was repeated for other courses. Once these indicators were created, we simply counted the percent of students who passed at least one AP course, passed math analyses, etc.

middle schools. Thus, there was a disjunction for these students when they entered the high school math curriculum. The team then asked how well bussed students who attended the local middle school (and were exposed to an aligned curriculum), and who had participated in a special mathematics program at the junior high school, performed compared to similar students not participating in the program. Results were promising and stimulated the school to think about additional supports entering students might need to make the transition from one type of math sequence to another. Although this line of inquiry is specific to a single school, it raises general questions that may be of interest in many schools: Where do the students come from? Are there many students from schools not traditionally considered feeder schools? To what degree are the curricula from all of these schools aligned with the curriculum of the receiving school?

The mathematics improvement team also probed further to look at what happened to students who experienced difficulty with a course. Is the course offered again in the following term, and if so, do these students take it? Do they pass on the second try? Third try? What percentage of students never passes? Results were disappointing. Data brought the mathematics team to a new question about the value of one possible action. If a two-semester (spring and summer) course was developed for students who fail the first semester, instead of retaking the more traditional single semester course, would more students pass and move to more advanced math and science courses? Or might it be possible to identify students beginning to show signs of academic difficulty early in the fall term and enroll them in a two-semester course initially? The team thus focused on what the school could do to modify courses to help students who were not performing well. They agreed on a course of action and agreed to follow up on the effectiveness of the new course.

In school group A, data did not "fix" the inequities faced by some groups of students at the school. However, data were useful in engaging key groups of constituents and helped a smaller set of school planners trace problems to the classroom experiences of students. Using a variety of data—initially, these student-level performance data; later, surveys of teachers and classroom assessments—they began to make the data a key component in decision making. Considerable effort, both internally and externally, was needed to progress. As a result, though, many of the efforts began to be institutionalized. The district, now having the design and specifications for the longitudinal database, assumed responsibility for maintaining the data districtwide. The district also began to support further training for others

wanting to initiate data use in school-level decision making. Appendix B contains the course materials, including a course syllabus, a description of course projects, homework assignments, and presentation slides designed for teachers and administrators currently involved in supporting data use for school-level planning and decision making. As software such as SPSS, or the locally developed Quality School Portfolio, continues to become more intuitive and flexible, these efforts will be easier to replicate and support on a large scale.

School Group B

School group B, as mentioned earlier, served an economically impoverished community. In this school, nearly all children were eligible for free or reduced-price lunch, and many were limited English proficient. Many of the students were transient, moving from school to school. Many of the teachers had emergency credentials. To no one's great surprise, when the administrators reviewed the initial data report, all of the indicators showed that student achievement was low, and that it was consistently low for all demographic groups. No clear priority or relative need emerged from any of the analyses.

Since literacy had been established as a school priority and the school had recently committed to additional assessments to better support and monitor student progress, the inquiry process focused on the use of assessment data to support improvement in this area. As part of their school improvement process, school administrators agreed to assemble a volunteer team of English teachers, led by a school coordinator, to examine results across available measures. Among the available measures were a new writing assessment developed by some teachers at the school, and a commercially developed diagnostic reading test, as well as the standardized test results for which the school (and particularly the administrator, at that point) was being held accountable. The intent was for the group to analyze student performance and use their analysis to improve teaching and learning in their classes. To better understand student learning and how to improve it, the group agreed to

- examine available data, including SAT-9 scores, student grades, and the district's newly developed performance assessments;
- apply rubrics from the district performance assessment to student work from their classrooms; and
- develop and analyze core classroom assignments aligned with the Los Angeles Unified School District (LAUSD) performance standards.

Through this process it was hoped that teachers would develop a common understanding of both district standards for student learning and what was expected of students on the various assessments. In addition to becoming familiar with district rubrics and applying them to classroom assignments, teachers were given early opportunities to evaluate students' performance on tasks similar to the Los Angeles Unified School District (LAUSD) performance assessments. This would enable teachers to better understand how to help students improve, while also providing direct opportunities for students to develop the knowledge and skills needed for success on the various measures.

The proposed group thus had both assessment and professional development goals and was designed to try to begin to bridge the gulf between external district assessments, which teachers viewed as insensitive to their efforts, and teachers' own classroom teaching and assessment. It was also hoped that the experience would help teachers come to a consensus on what to expect of their students.

Laudable as these intents might have been, however, progress toward them was limited. Despite being volunteers who were advised of the intended process in advance and led by "one of their own,"[4] a former teacher, whom the participant teachers presumably trusted, the group balked at almost every step of the process. They almost totally discounted the results of the SAT-9 and preferred to discuss the reasons why the test was inappropriate for their students and why improvement on it, despite pressure from the principal, was unlikely. The writing assessment developed by some of the teachers at their school was similarly dismissed as inappropriate. While there was limited discussion of the type of assessments that might be appropriate, the group had difficulty coming to consensus. They did agree on a core assignment to give in their classes, but when the time came to bring in students' work for joint analysis, most came empty-handed. The experience provided more lessons in the importance of the special expertise needed by facilitators of such groups than it did lessons in the use of assessment. The importance of trust and efficacy were underscored. Combating a siege mentality and getting beyond blame and defensiveness to action are problems that go far beyond technical and mundane aspects of data use.

[4] The school facilitator was considered an expert in language arts and a veteran in reform projects that integrated writing across the curriculum, particularly in urban schools. Two advisors assisted the facilitator, one an expert on assessment, data analysis, and use, and the other on reform processes.

Lessons Learned

We learned many lessons in our work with these two schools as well as with the other data and training projects. The good news is that easily available technology can facilitate relatively easy access to student level and other available data, and these data can be effectively represented and communicated in ways that support school inquiry. The technical aspects of data access and communication are only a small part of the challenge of school-based inquiry, however. The social-political context determines the ultimate value of the inquiry and its contribution to school improvement.

Data Access and Analysis

Data-based inquiry does not require complex analysis. All of the questions described above and in almost all of our work with schools could be answered using simple descriptive statistics, which are available through any number of easily available programs (e.g., Excel). In no instances were complex multivariate analyses needed.

Our analyses did depend on the availability of student-level data. As noted above, the districts with which we worked kept extensive databases, which they stored annually. Thus, we were able to create longitudinal databases by linking annual records through student identification numbers. The result was a flat database similar to that depicted in Table 1. A similar study—albeit containing fewer variables of interest—could be accomplished for schools in California by simply merging SAT-9 results (delivered to schools on a CD) for each year.

While the data merging was a straightforward process for demographic variables, test scores, attendance data, and the like, the course history information was more involved. In our analyses, as we noted earlier, we were interested in determining the number of AP courses that students enrolled in and passed, how far along the math curriculum students progressed, and whether or not students had successfully completed the courses required for admission to the University of California. Creating the appropriate variables for analysis required more complex data management options than are available with standard statistical packages such as SPSS, or using computer programs written in languages such as FORTRAN. Essentially, these programs were used to create new variables that counted for each student the number of AP courses enrolled in and passed, the math courses passed, and the number of courses required for eligibility to the University of California (A-

F requirements) that were passed. For example, we created a variable for math analysis. Initially, the value was set to zero (not passed). Then we moved through the entire course history for each student, checking to see whether s/he had taken math analysis and whether the grade was a "C" or better. This was repeated for other courses. Once these variables were created, we simply calculated the percent of students who passed at least one AP course, passed math analysis, etc.

Analysis Choices

Is student performance improving? The question is a basic one and seems very simple. Indeed, schools, districts, and newspapers typically take a simple approach to the answer. They compare this year's test scores with those from previous years. If the scores go up, student achievement is thought to be improving. But how and which scores are used in the calculation can greatly influence the outcome. Some of these analytic choices are dictated by available data, whereas others depend on the subtle differences in the questions being asked.

Basically, there are three reasonable choices in looking for progress or improvement in student test scores: longitudinal, cross-sectional, and quasi-longitudinal analyses. In a longitudinal analysis, which methodologists view as the purest way to examine progress, the same students are followed for several years. For example, you might examine test score performance for students currently in the school (1999) who were in the third grade in 1997, in the fourth grade in 1998, and the fifth grade in 1999. Only students enrolled in the school who completed the test for all three years would be included in the analysis. If a student left the school after 1997, came to the school after 1997, or didn't take the test in one of the years, the student's scores would be excluded and not counted in the analysis. The advantage of this approach is that it controls for students; that is, results are not biased by changes in the nature of the student body. Such changes may account for changes in performance. We want to know whether students are learning or improving in scores, and want to trust that the results are not caused by irrelevant factors, such as changes in whose scores are being counted. The longitudinal approach also provides the strongest test of the question "Is the school helping students improve, or at least increase their test scores?" because it focuses on the progress of students who have attended that school for a sufficient period of time to have been affected by the school program. On the down side, the true longitudinal approach requires annual testing in each grade (which is costly in student time and financial resources) with a test adequately scaled across years. It also requires the availability of a student-level

database that includes information on the years each student has attended the school. Furthermore, the true longitudinal analysis ignores the performance of more transient students, and thus may be inappropriate for schools with high transiency rates. It does not answer the question ”How are students at this school doing?” but rather the more specific question “How are students who have been at this school continuously for ‘x’ number of years doing?”

In the second type of analysis, cross-sectional, all students in a single grade this year are compared to all students in that same grade in prior years—for example, you might compare test scores for third graders this year with test scores for third graders last year. A major problem with this approach, as any teacher can tell you, is that this year’s third graders may be very different from last year’s—for example, in ability, motivation, preparation, etc. Or there may be dramatic changes in the student population from one year to the next—for example, a change in school boundaries. If these differences exist, higher test scores may not be the result of improvements in teaching and learning at the school, but instead an artifact of changes in students. Nonetheless, when test results are available only for selected, noncontiguous grades, a cross-sectional approach is the only available option. It does have the advantage of including all students—so if the issue is the strengths and weaknesses of this year’s third graders, it is the appropriate analysis. Further, with trends *over a number of years*, schools may begin to see consistent patterns that bear on the quality of the school’s program at selected grade levels.

We refer to the last approach as quasi-longitudinal because it combines elements from the previous two. Test scores for a grade level in one year are compared with those in previous grade levels in previous years. For example, you could compare the average test score for all fourth graders this year with test scores for all third graders last year. This approach is not as precise as the longitudinal method, but tends to provide results more similar to the latter than the cross-sectional approach. It answers the question “Are our students making progress as they move through the grades?” but is confounded by transiency. That is, where there is high transiency—for example, many of the students in this year’s fourth-grade cohort were not at the school, nor tested last year in the third grade—the pattern of results may be as much an artifact of changes in students as the result of school curriculum and instruction. However, this result is the only possible option when schools do not have access to an individual student database—which is often the case—and have only aggregate results from one year to the next.

Although the decisions about which approach to use may seem arcane and subjective, they do have a significant impact on most schools' average test scores. Optimally, schools should look at results from multiple perspectives, each providing a somewhat different story. However, it is also the case that many schools will show improvement while using one of the methods and decline using another, opening possibilities for misrepresentation.

Need for Better Learning and Other Performance Indicators

Unfortunately, looking at course-taking patterns was as close as we were able to get to classroom curriculum and practices with available data, and the alignment of available test data (SAT-9) with district standards and classroom curriculum was problematic. Even given a better match, schools and teachers need data more than once a year to support student learning. They need better assessment data to understand and monitor their students' progress, take appropriate action, and use that data wisely for school decision making. Assessments that are closely connected to the curriculum, diagnostically useful, and that can transverse the gap between classroom and external assessment are necessary. In the absence of such assessments, there are severe limits to what can be accomplished with available data and to how much they can inform curriculum planning and school reform.

There is a need to improve other commonly used performance indicators, as well. For example, dropout rates should be calculated within a single year, and over several years, by tracking students over time. Attendance rates should be calculated as the percent of days the student attended the school, rather than the average number of absences, since schools differ somewhat in number and length of days. Regarding indicators based on courses (AP, A-F, algebra, etc.), our experience working in vastly different school contexts (with vastly different relationships between course-taking and standardized test scores) suggests the need for standardizing the meaning of these courses. We believe a strong argument can be made for developing common, end-of-course exams that assess students' mastery of course expectations. Then we could determine the percentage of ninth graders (or any other grade) that has passed the corresponding grade-level course exams. In high schools, such measures should be emphasized over those based on mastery of college-level material (e.g., AP tests) since it is more appropriate to hold high schools accountable for students mastering high school material.

Finally, the ultimate measure of K-12 success is whether students successfully transition into the workforce or higher education. Constructing such indicators would require tracking students after completion of their K-12 education. Although such a proposal would be cost-prohibitive for an individual school, the information already exists at the state level for those employed in California or enrolled in public colleges and universities in California. Were there a common statewide student ID system—such as social security number—the information could be gleaned from these other databases. Important issues of privacy aside, such a common ID system would enable the state to better monitor student progress, even as our mobile population crosses districts, and to facilitate the timely and easy transfer of student information from one district and school to the next.

Data Representation

Throughout our work with schools, we found that a key factor in data use was how data were reported. (In the foregoing sections, we have sampled the much larger pool of reports created in order to illustrate a range of possibilities.) Several principles emerged from our work:

- Graphs and other iconic representations communicate more effectively and efficiently than words and tables of numbers. We found that teachers and schools had limited tolerance for reading explanatory materials. It thus is important that displays be as intuitive and self-explanatory as possible.
- Similarly, any unnecessary detail or clutter should be removed. Backgrounds were made clear (as opposed to gray) and information was arranged to eliminate repeating labels.
- Graphical representation should be kept as simple as possible to make the information easily comprehensible to school communities—people not trained in statistical analyses. Simple bar charts, pie charts, and line diagrams are good examples. Box and whisker plots, stem and leaf plots, and other complex displays that educational researchers like to use to depict data are counter examples. Stacked bars, simple three-dimensional displays, and histograms are workable, but require special explanation. (Stacked bars are much more workable with color than without.)
- Color is an important element in effective reporting. It can be effectively used to engage audiences and communicate key ideas. For example, bright red, yellow, and green can be used to reinforce the idea that the levels of performance depicted are “needing improvement,” “ok,” or “good.” Shading of the same color (for example, a range from dark blue to light blue) can be used to illustrate and reinforce the idea of a continuum. Users also find it easier to see dimensions differentiated by color than by shading, cross-hatching, etc. The down side of using color, however, is that most schools do not yet have access to color printers, making it difficult and

somewhat expensive to copy and share results widely. Posting on the web can be an effective alternative.

- Statistical maps, made using easily available software, can display relationships between location and variables of interest. These are particularly important for schools serving wide and diverse neighborhood areas.
- Consistency in the use of color and graphical design provides users with a unified framework for comprehending the information. Similarly, sticking to a few different kinds of displays is helpful.
- Design matters. Schools and their communities are attracted to, and thus more willing to engage with, data displays and reports that are visually appealing. Investing in the services of a graphic artist can be important to the process.
- Confirming the findings of others, we have found the following kinds of representations useful for different purposes:
    a. Bar charts encourage users to make comparisons—among subgroups, for different subjects, or for different years; they also are useful for showing relationships.
    b. Line graphs are useful for depicting progress over time, but are only appropriate for true longitudinal data.
    c. Pie charts are useful for displaying demographic data.
    d. Stacked bars, if color is used, can communicate the distribution of performance across different levels (e.g., percent of students scoring at different proficiency levels or percent of students scoring in different quartiles).

Socio-Political Perils

It is more than apparent from our two case examples that data use in schools is more than a technical/technological issue. Ostensibly the same technological processes were initiated in both sites, but our attempts at technology transfer achieved vastly different results. What differentiated the two? Certainly, leadership and culture were important issues. In the one setting, there was a strong leader: well trusted, well versed and committed to outcomes-based curriculum planning and decision making, and knowledgeable about evaluation and the use of data for improvement. He led the process and was instrumental in creating a team context that was focused on improving student learning. The culture supported data use, in part because data were a natural element in the team's view of its role and of the nature of improvement planning. Over a number of years, with ongoing support from the assistant principal, the team had adopted an outcomes-based model of

planning and saw data as critical in defining needs and assessing progress. At the foundation, they agreed on the goals of their planning activities and committed to the enterprise.

But the differences between the two examples were more than cultural. Differences in the political context abounded. In the one case, the school team felt empowered by data and the planning process; they felt confident they could use the data to improve opportunities for their students. In the second case, the teachers felt disenfranchised and devalued by the data. They feared that the data would be used against them, and felt little sense of efficacy. Furthermore, the teachers with whom we worked were distrustful of the administration, and at odds with other teachers in their department about the nature of language arts learning they should be promoting. Struggles over ideology, competing interests, "we-they" mentality—all these conflicts overwhelmed any continuing interest or motivation to use data. An obvious point that is worth underscoring is that a school must come together first as an empowered community, seeking agreed-upon goals and committed to change, before any kind of meaningful use of data to support change is possible.

Misuses of Data: Statistics Don't Lie, But Liars Use Statistics

The growing emphasis on the use of test results to judge schools and teachers not only can make teachers nervous, as the above scenario suggests, but the pressure can also spur on districts, schools, and/or programs to less-than-desirable action. We would be remiss not to point out that our work provided a few examples where data could be used to misrepresent, rather than to support change and improvement. As statisticians well know, data can be manipulated to satisfy the needs or interests of anyone, and the temptation to manipulate grows as greater and greater consequences are applied to schools for showing improvement, or lack thereof. The instances we cite below admittedly are exceptions, but worth guarding against in policy and practice.

Selective reporting of results. We noted above the different legitimate approaches for examining improvement in performance, and the possibilities of getting different answers from different methods. Similarly, schools may face decisions about which grade levels and content areas to include in their reporting, and how to combine data. In fact, if we were to run the analyses using all of the different ways described above, the vast majority of schools would show improvement in at least one analysis and decline in others. Thus, through selective

reporting, a district, school, or program can dramatically change whether student achievement appears to be improving or declining.

Manipulating those tested. Two major special groups of students tend to score worse on assessments: students with disabilities and students who are not proficient in English. In the absence of strong and enforced state or district rules in this area, there is little consistency in testing such students; sometimes they are included in school assessments, sometimes they are given accommodations, and sometimes they are excluded. When the proportion is high for students in a school who are designated as disabled or limited English proficient, differences in who is included from year to year can have a significant impact on the overall test scores. In addition, schools have some flexibility in who is specially designated, and changes in such designations from year to year can influence scores. Similarly, without requirements for testing absent students, schools can subtly encourage some students to be absent on test day, again with impact on observed performance. Carefully checking the number and proportion of students who are tested from year to year is one potential safeguard in this area.

Manipulating dropout rates. The primary means of manipulating dropout rates arises from how that statistic is defined. Ideally, a dropout indicator would track students throughout the school year and represent the proportion of students who were there at the beginning of the year, but no longer attending the school at the end, and who had not transferred to another school. But because of concerns for ease of calculation and accuracy of school tracking systems, dropout rates typically are computed using a different formula. Commonly, the rate is based on the difference between the number of students attending the school at the beginning of the year and the number at the end of the school year. The dropout statistic is calculated by dividing this difference by the number of students attending the school at the beginning of the year.

Although this indicator is typically a fair proxy for the true dropout rate, it does not account for students enrolling in a school after the first time point at the beginning of the year. If a school were pressed to show improvement, they might be tempted to wait to officially enroll some of their students—say a few hundred—until after that first date. Then a couple hundred students could drop out of the school, without any difference in scores showing up, and the computed dropout rate would still be zero. Even if a school does not consciously manipulate the data, determining whether a student has truly dropped out of school is difficult,

since there is no way of tracking students across districts. Who is considered a dropout may be as much a function of follow-up effort and clerical conjecture as anything else.

A-F requirements and AP courses. Another set of measures of school performance, which were included in the school report cards discussed above, look at students' enrollment and performance in challenging courses. Two such types of courses are A-F courses, which are required for eligibility in getting into the University of California, and Advanced Placement (AP) courses, which are intended to be rigorous college-level courses offered to high school students. If a student passes the College Board's exam after completing an AP course, the student typically receives college course credit when he or she enrolls in college.

The primary limitation of these measures of school quality is again based on how the indicators are composed. Typically, the indicators are derived from the numbers or proportions of students enrolled in each course and course grades (passing or not passing). In practice, any high school course can be *called* an A-F course or an AP course. Moreover, the teacher of the course can assign any letter grade he or she feels is appropriate without much external scrutiny, leaving the indicators totally under the control of the school and as potentially prime candidates for manipulation.

Conclusions: Challenges in Moving Toward School-Based Inquiry

Our work shows the potential for, but also illustrates some important challenges that continue to impede, data-based inquiry in many schools. In this final section, we step back and consider the broad issues that must be addressed in moving forward.

A first set of issues has little to do with the methodology or technology of data use. Instead these issues center on leadership and commitment to standards-based reform and real improvement in student learning. Our two case studies illustrate relative extremes on these issues, with consequent differences in results. As we noted above, data use is a natural activity in schools that are a community, that agree on goals, and that focus on results. Authentic use in school-level decision making is unlikely, and even irrelevant, in schools that do not share a common vision of goals or that, in truth, are not interested in change. Leadership is an intangible quality and difficult to truly develop, but it is essential to school reform and within reform, to productive use of data. This is not to say, however, that

principals or administrators are the only ones who can assume the mantle; indeed teachers can and do fill the void. Most importantly, there is a distinct need for someone (or several people) to marshal the commitment of the school community and bring the school to a common vision of its goals, common understandings of expectations and standards for student learning, and a shared commitment to action. Building the technical capacity of schools to respond to standards-based reform is necessary, but not sufficient. Building leadership to develop and nourish a culture conducive to moving forward on the reform agenda is the more difficult endeavor.

Having said this, our work suggests the importance of continuing to build schools' capacity to respond productively to the requirements of standards-based reform. Again, understanding and the capacity to engage in standards-based education is a first, essential step, within which effective data use may be embedded. Teachers and administrators in many of the teams and schools we dealt with had an incomplete understanding of the standards they were supposed to be addressing, and what the standards meant for school practice. California's current accountability system contributes to the misunderstanding. The Stanford Achievement Test, Version 9 (SAT-9), is the only indicator of student learning, yet the test is not aligned with state and district standards. To many, teaching to the standards means teaching to the test, an issue we return to below.

Beyond understanding of standards-based reform, most schools need to develop the capacity to collect, analyze, and use information to inform their efforts. In our project, *we* were the technical capacity for the schools, and through modeling actual examples, templates, and direct training, we were able to leave some capacity to continue. However, it was a time- and resource-intensive endeavor that is unlikely to be available to every school.

How can schools' capacity be more developed in a less resource-intensive, more feasible way? We believe there are multiple avenues to pursue jointly. It continues to be the case that administrators and teachers receive little pre-service training in assessment and evaluation, and when their enthusiasm for data use is stimulated, they may engage in analyses that are methodologically flawed and therefore likely to provide inaccurate inferences. Including additional training in evaluation and assessment during pre-service programs and mandating such coursework for credentialing seems an obvious step in beginning to address the capacity problem at its source.

At the same time, teachers and administrators should not be expected to be experts in research/evaluation design and analysis, given the other pressing demands on their time and energy—notably teaching children and fostering their development.[5] Rather, teachers and administrators should be knowledgeable about the basics. If schools are going to move ahead sensibly in this area, they need to have such expertise available to them. One option would be assuring that every school had an administrator, coordinator, teacher, or other person with the necessary expertise to provide help and support in this area. A specialist credential or certificate in assessment/evaluation might be considered. Such specialists could be school based—for example, an administrator whose portfolio includes leadership and support in data-based inquiry—or districts and/or county offices could consider full-time specialists in this area of service for every school.

Technology offers another potential solution as "smart" system options become more available. For example, the National Center for Research on Evaluation, Standards, and Student Testing's (CRESST) Quality School Portfolio (QSP) can convert available student data from multiple sources into an easy-to-use longitudinal database and can enable schools to easily perform the kinds of analyses cited above. With QSP, schools can answer basic questions about their progress relative to goals, performance over time, subgroup performance, and the relationships between various demographic and performance variables. Safeguards can be and are being built into QSP to protect against inappropriate analyses and to prompt important questions. QSP also will soon have a tool kit to help schools administer additional measures to monitor their progress and inform their decision making.

The latter capability responds to a continuing problem in supporting schools' use of data: the limits and sensitivity of the data currently available. As we noted above, there are definite limits to what can be learned with commonly available, standardized test data or other large-scale assessment data from the state or district level. It is worth underscoring that these data are at considerable distance from the real issues of curriculum and instruction or from providing teachers and schools with the kinds of detailed and qualitative understanding they need to help move students ahead. There is a continuing need for better curriculum-embedded and

[5] Certainly assessment of student learning is an important component of the teaching process, but here we refer to issues related to appropriate research design and analysis strategies, interpretation, and reporting.

classroom measures of student progress that are linked to standards and broader assessments. As with the challenge of school inquiry process, moving to better classroom assessments is a problem containing capacity, resource, and socio-political dimensions. The capacity dimension derives from the fact that pre-service programs rarely provide teachers and administrators with experience in being good evaluators or being informed consumers of available assessments. Despite the need for measures that are tailored to ongoing objectives and curriculum emphases, schools and teachers have little capacity to create such measures themselves. Good classroom measures indeed are hard to find (assessments accompanying texts and other learning materials are of notoriously poor quality), not to mention assessments that are aligned to local or state standards and measures. Furthermore, the hesitation to accept and use classroom measures not "invented here" adds an additional socio-political complication.

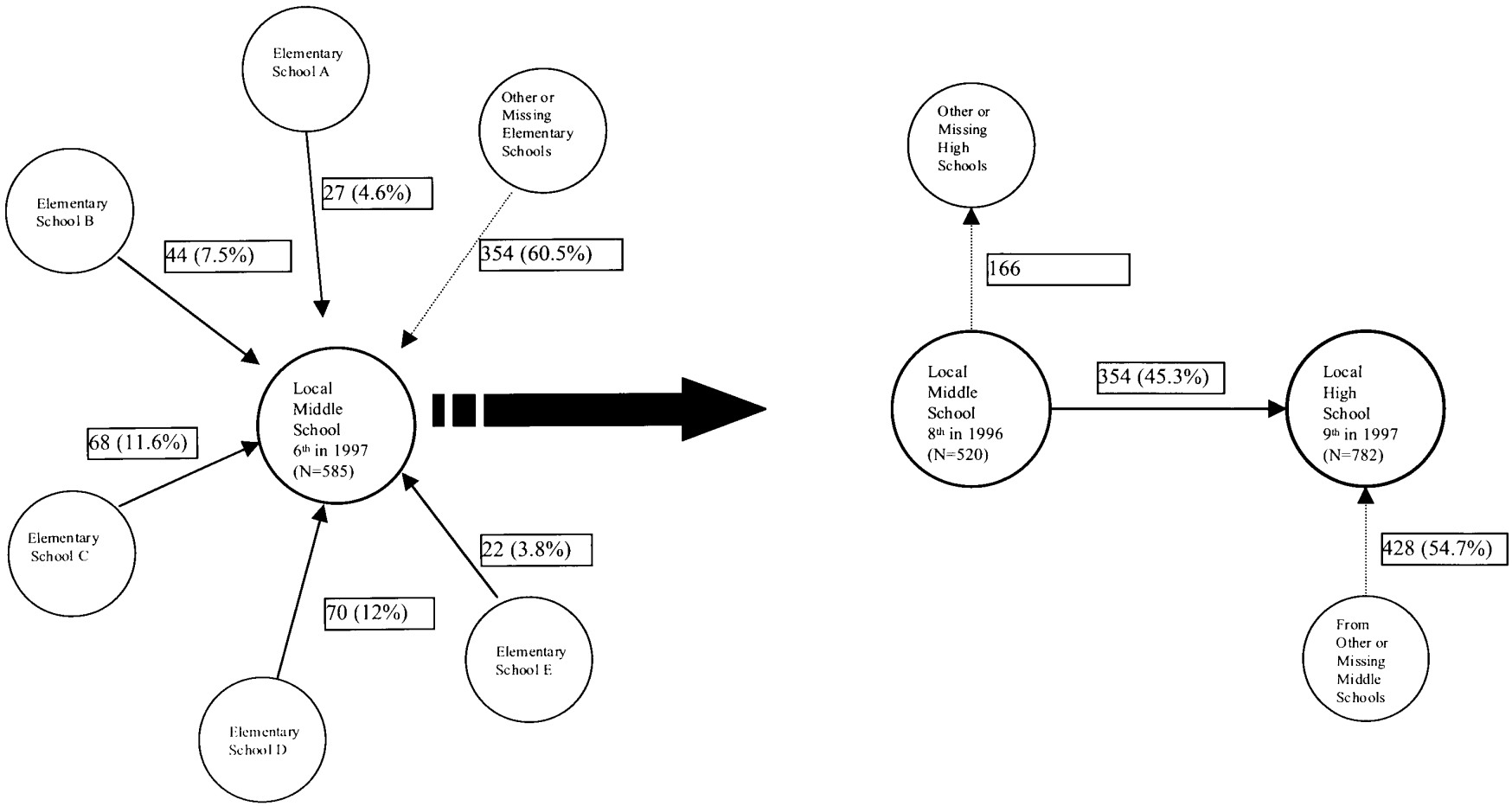
Finally, our work shows the strong and undeniable tension between top-down accountability requirements and authentic, bottom-up school inquiry processes directed at improving student learning. Current policy initiatives put unprecedented pressure on schools and teachers to be accountable and to show progress on what continue to be relatively limited measures of student achievement—in California, for example, the whole accountability system currently rests on the SAT-9, which is not aligned with state standards. Those schools and teachers who are able to show expected progress will reap financial benefits; those who do not may face serious negative consequences (e.g., school takeover and career-affecting "marks" on teachers' and administrators' records). Under pressure to show improvement, improving test scores, rather than improving learning, can easily become the target. For some of the administrators with whom we worked, this indeed was the case. In this context, there is danger that the inquiry process will become reduced to "psyching out" the important elements of the test (and its subscores) and a guide to test preparation instead of an important strategy to improve student learning. Principals and teachers, feeling that they lack realistic strategies to improve both, may resort to other means to guarantee expected improvements—drill and kill curriculum, coursework focusing only on tested content, a meager curriculum relative to what students will really need to succeed in life, or even cheating. Such actions may produce short-term increases in scores, but are likely to have negative long-term consequences.

There is a philosophical tension here, as well. Accountability requirements seek top-down control, while meaningful inquiry processes require bottom-up empowerment. These truly are competing perspectives on how best to support educational improvement. Can we achieve the right balance to move forward productively? Our work with schools shows some of the challenges, but in addition provides examples of the professionalism and serious inquiry in which many schools, administrators, and teachers are engaged.

Appendix A: Slides 1-23

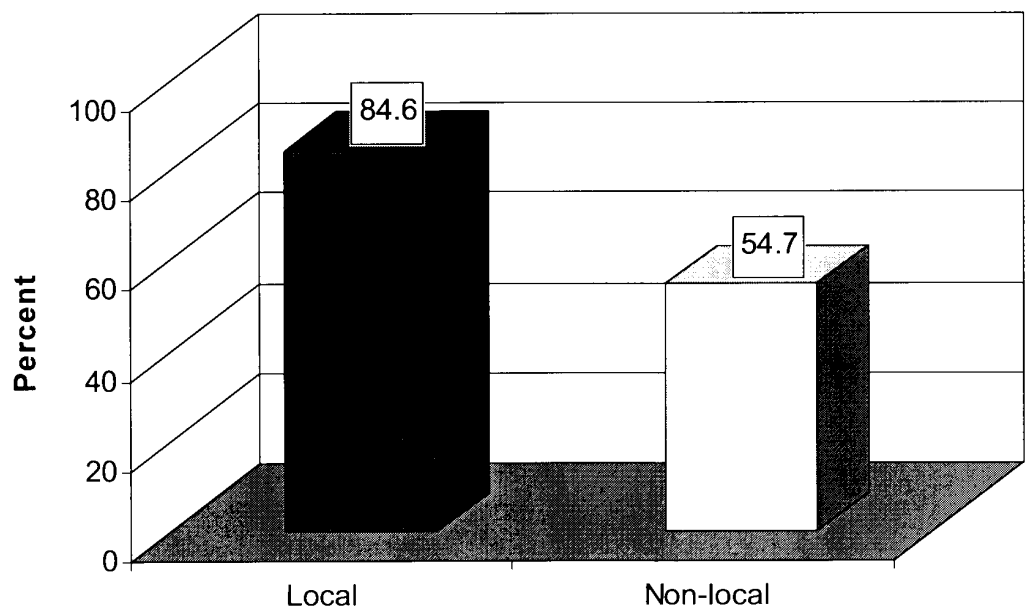
Slide 1 – Student Feeder Pattern
Slides 2-4 – Comparison of Local vs. Non-local Students
Slides 5-8 – Continuation Rates
Slides 9-11 – AP Course Taking Patterns
Slides 12-23 – Math Course Taking Patterns

Student Feeder Pattern
(1997-1998)

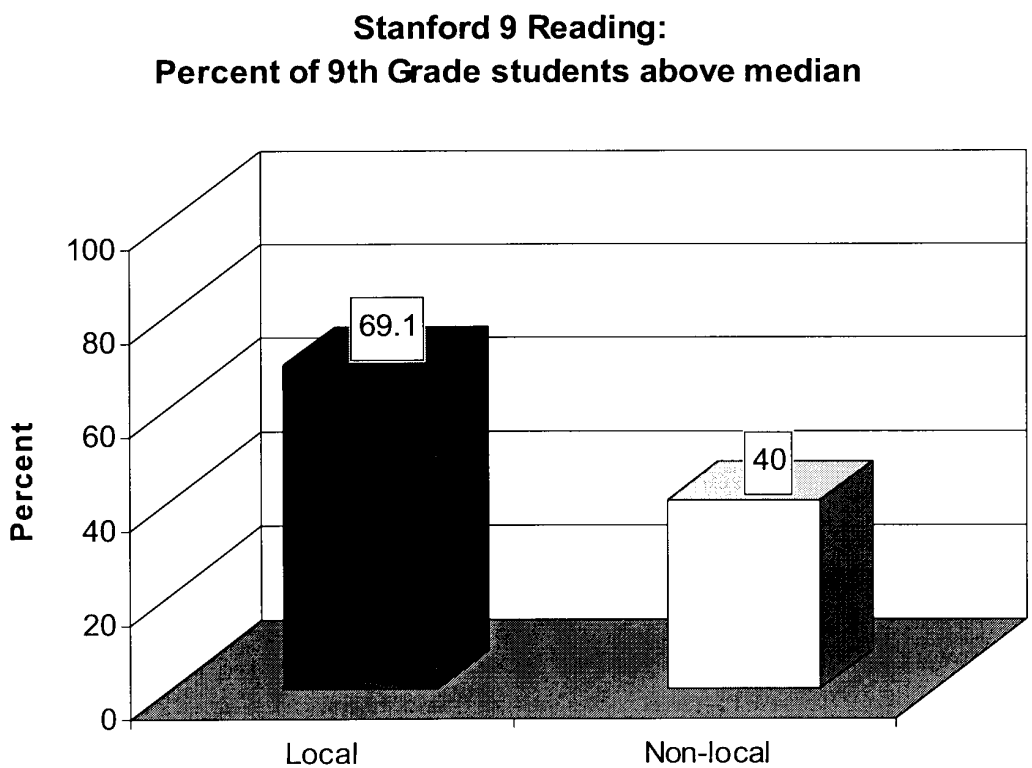


Sample High School:
Comparison of Local vs. Non-local Students

Stanford 9 Language:
Percent of 9th Grade students above median

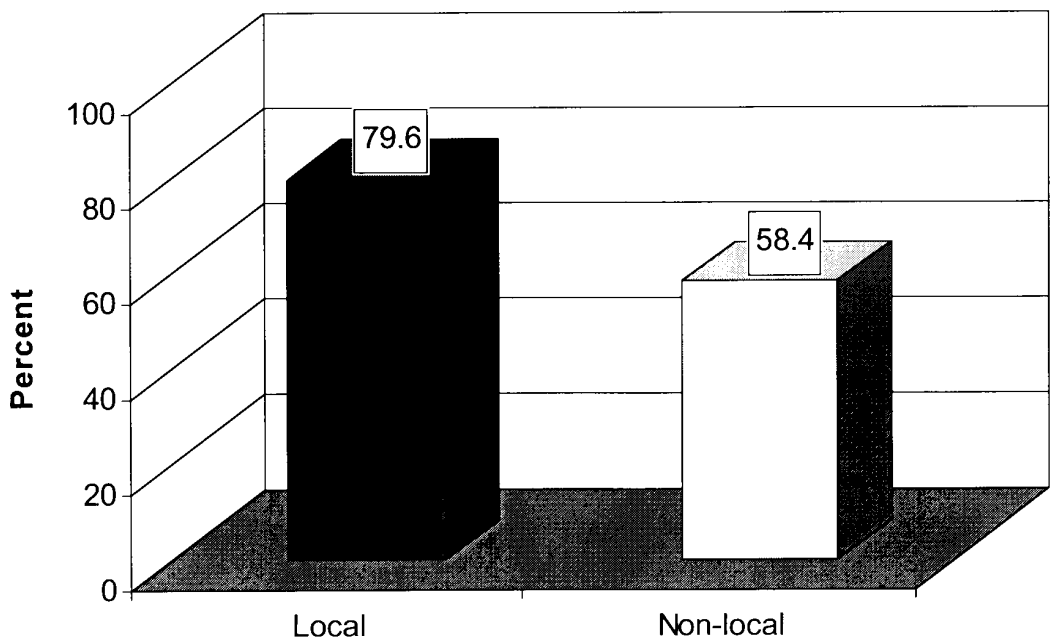


Sample High School:
Comparison of Local vs. Non-local Students



Sample High School:
Comparison of Local vs. Non-local Students

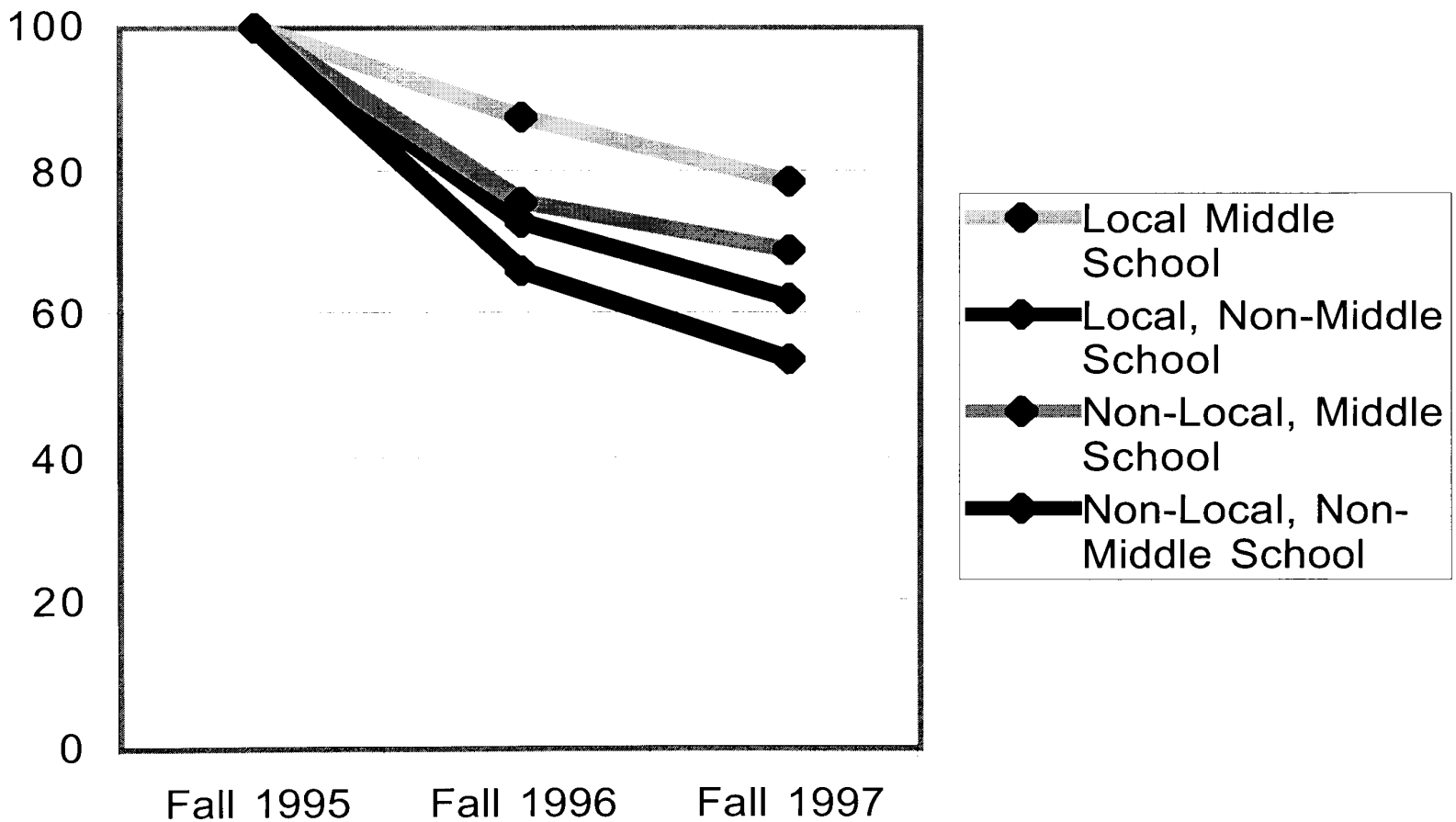
Stanford 9 Mathematics:
Percent of 9th Grade students above median



Continuation Rates

Comparing the Groups

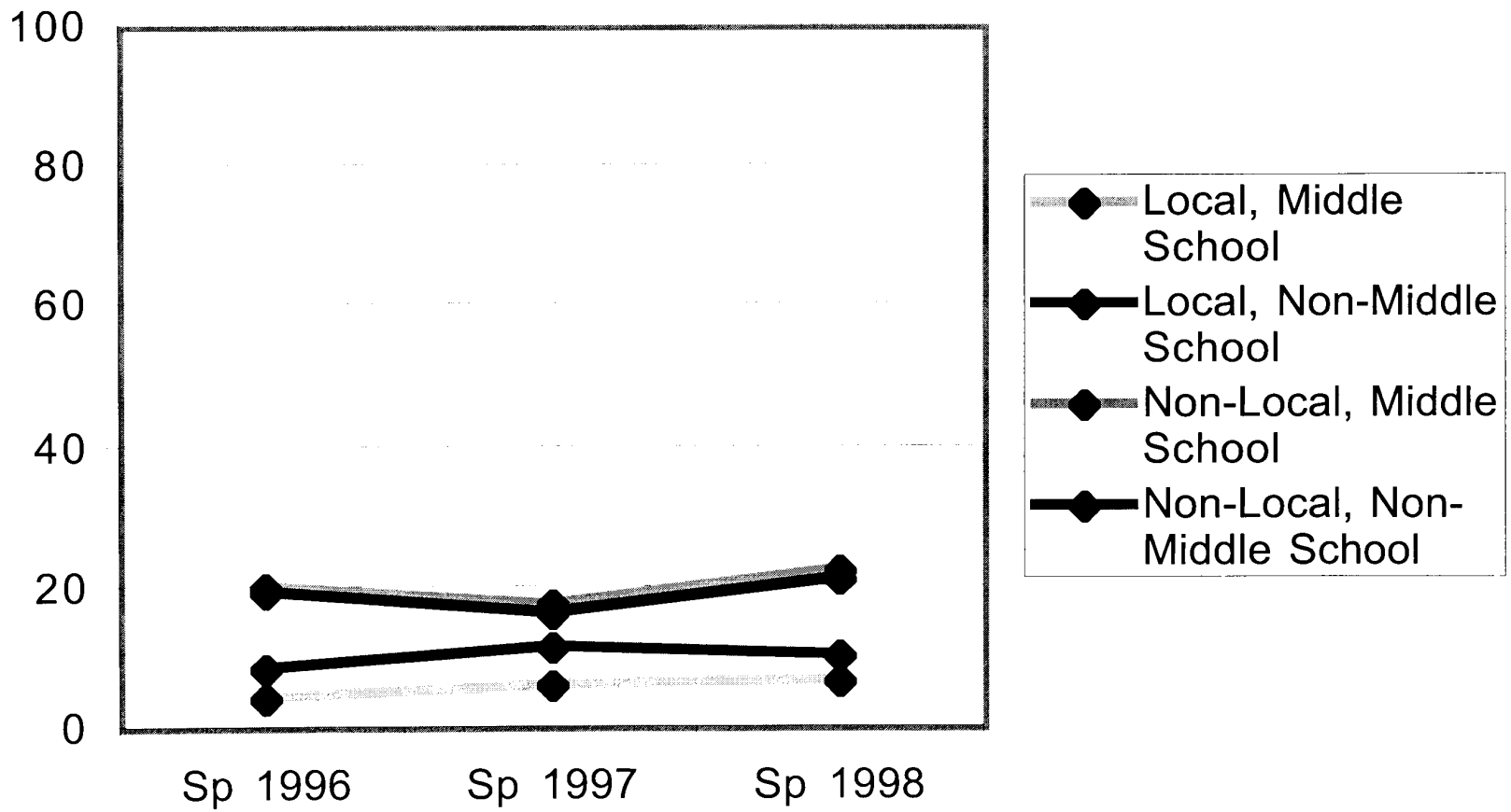
Continuation Rate:
1995 Entering Freshmen Cohort



Absenteeism Rates

Comparing the Groups

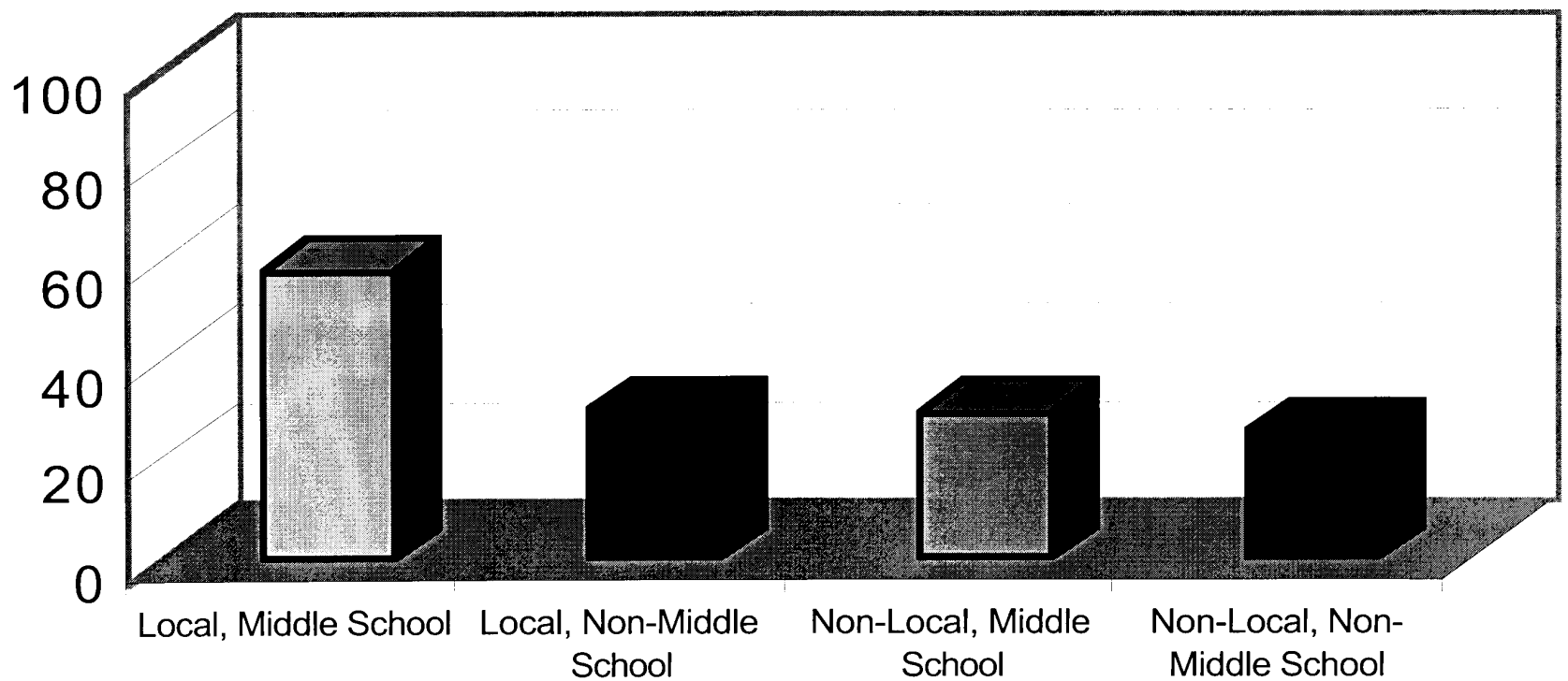
Absenteeism Rates Over All Groups for 1995 Entering Freshmen Cohort



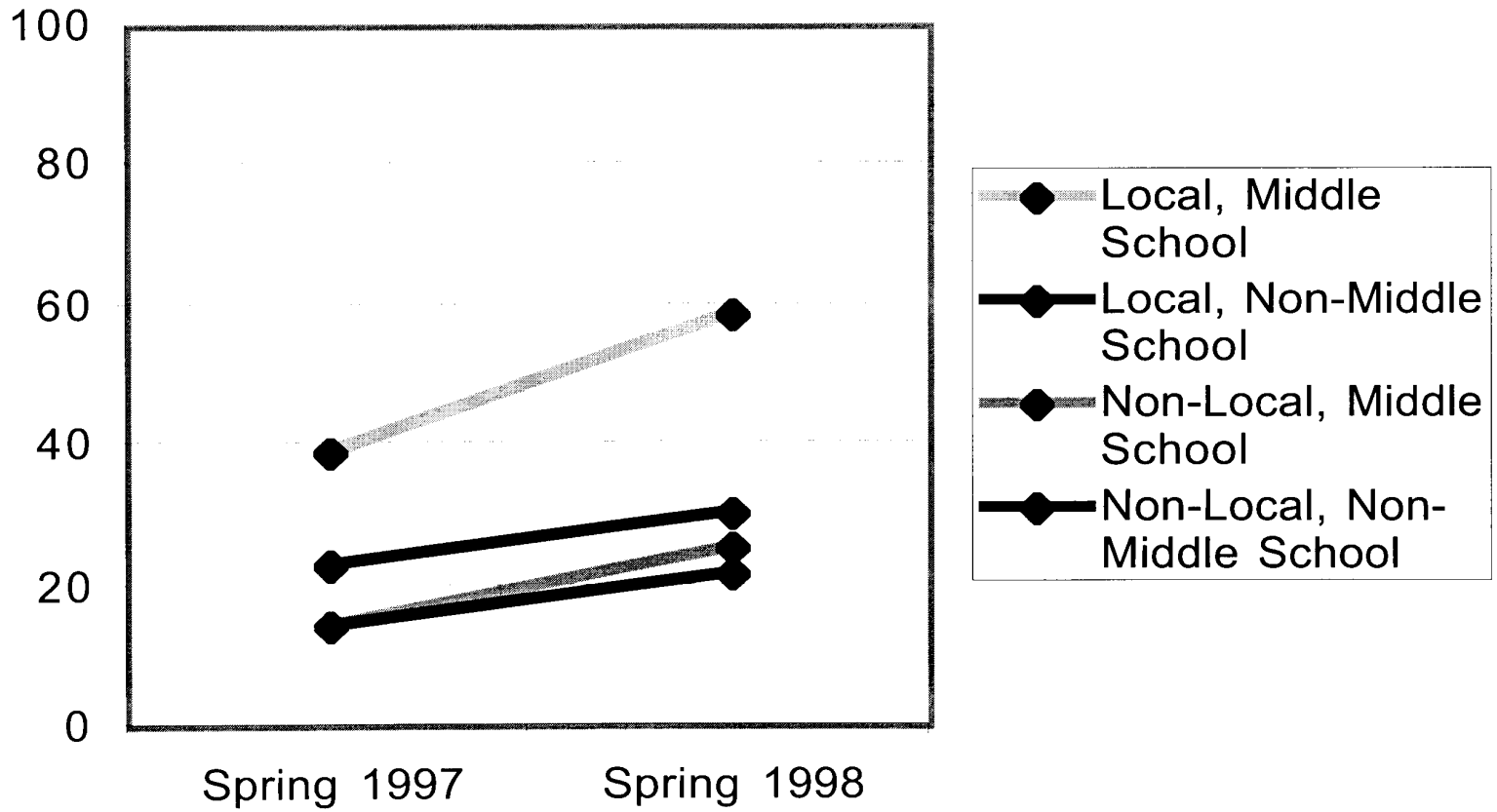
AP COURSE TAKING PATTERNS

Percent enrolled in at least one AP course, comparing groups

Percent of Students Enrolled in at Least One AP Course: 1995 Entering Freshmen Cohort, Spring 1998



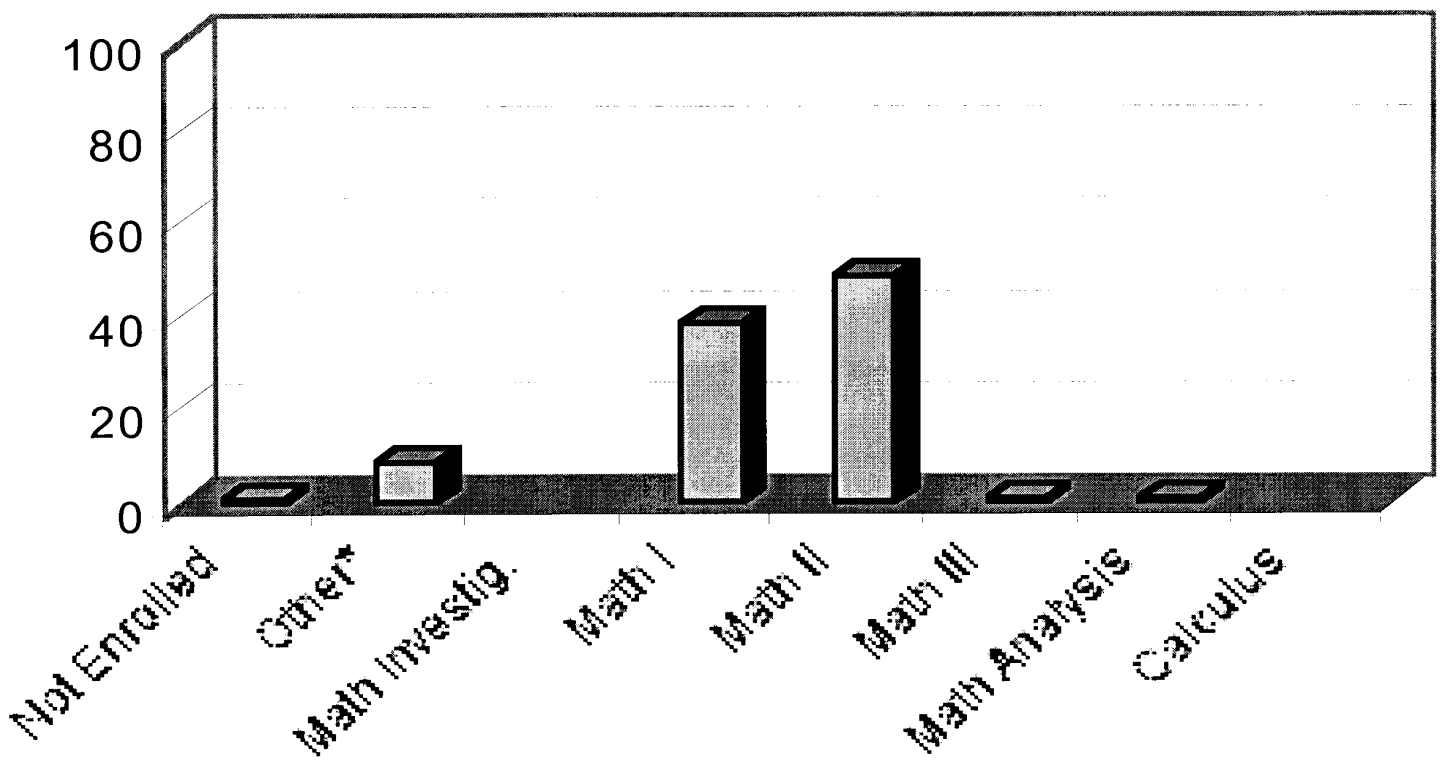
Percent of Students Who Passed at Least One AP Course: 1995 Entering Freshmen Cohort



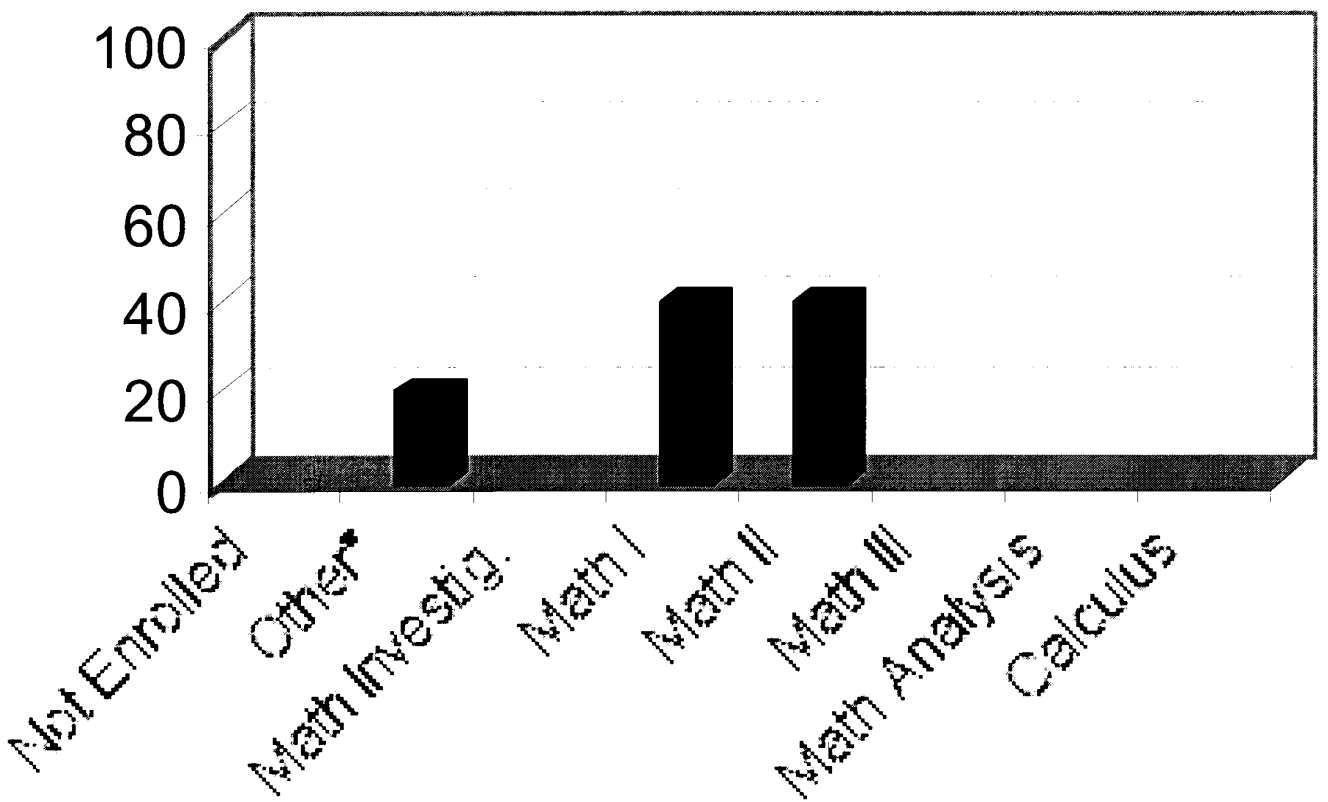
Math Course Taking Patterns

Comparing the Groups

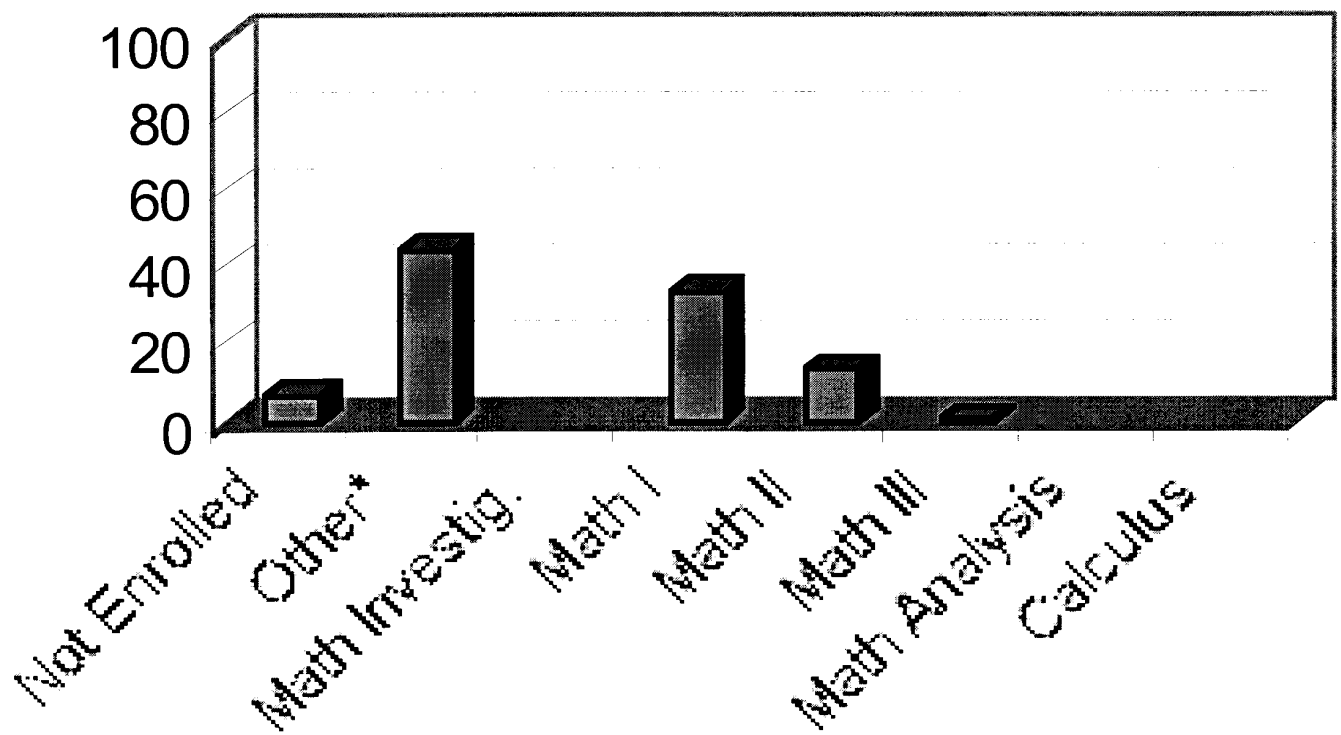
Local, Middle School: Math Course Taking Patterns for 1995 Entering Freshmen Cohort, Fall 1995



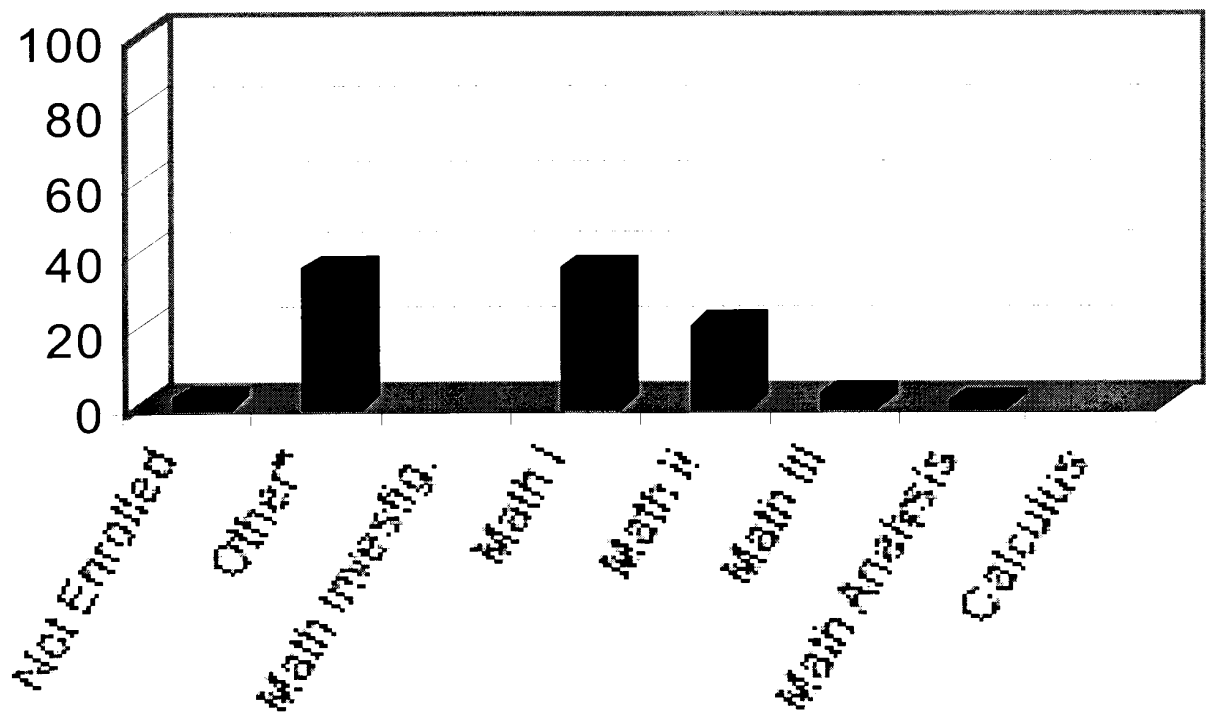
Local, Non-Middle School: Math Course Taking Patterns for 1995 Entering Freshmen Cohort, Fall 1995



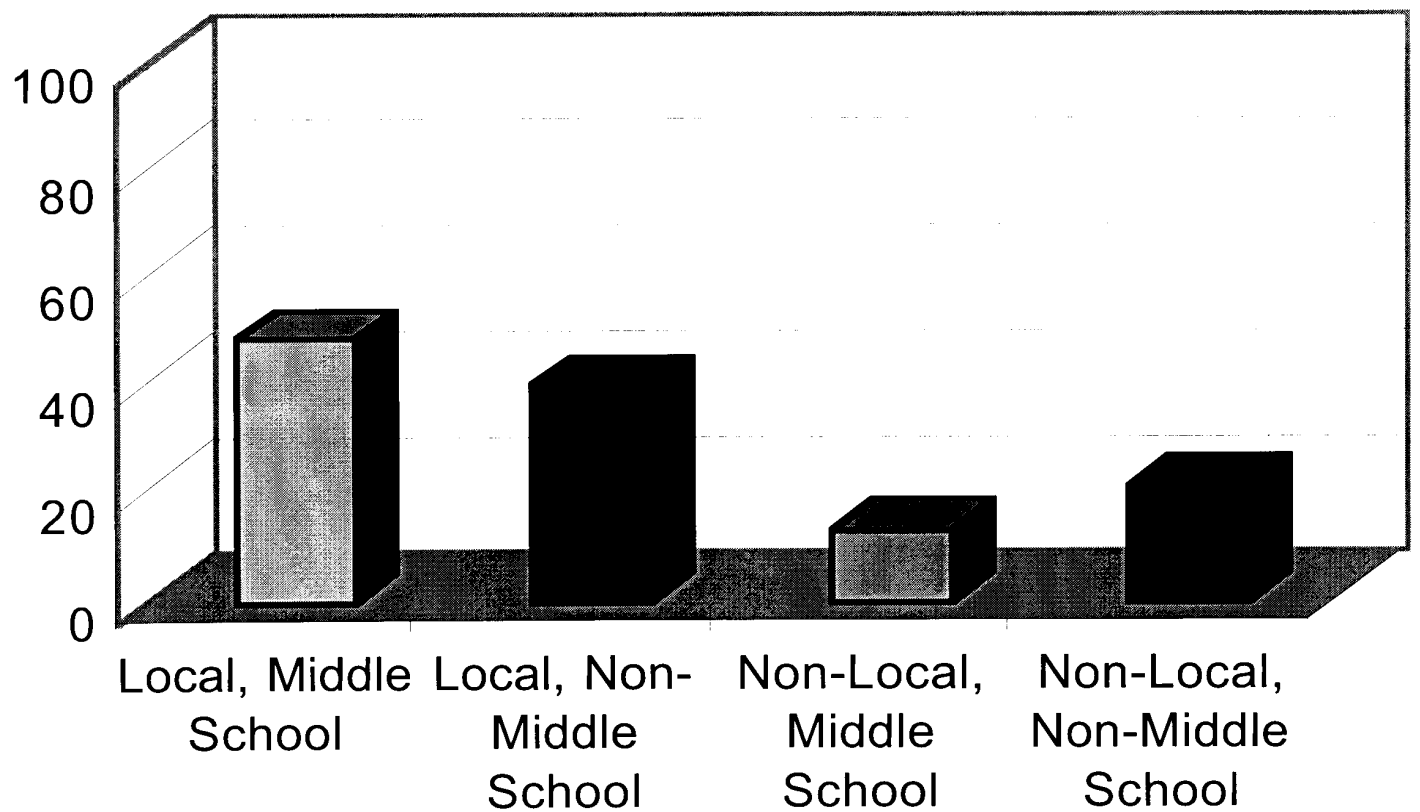
Non-Local, Middle School: Math Course Taking Patterns for 1995 Entering Freshmen Cohort, Fall 1995



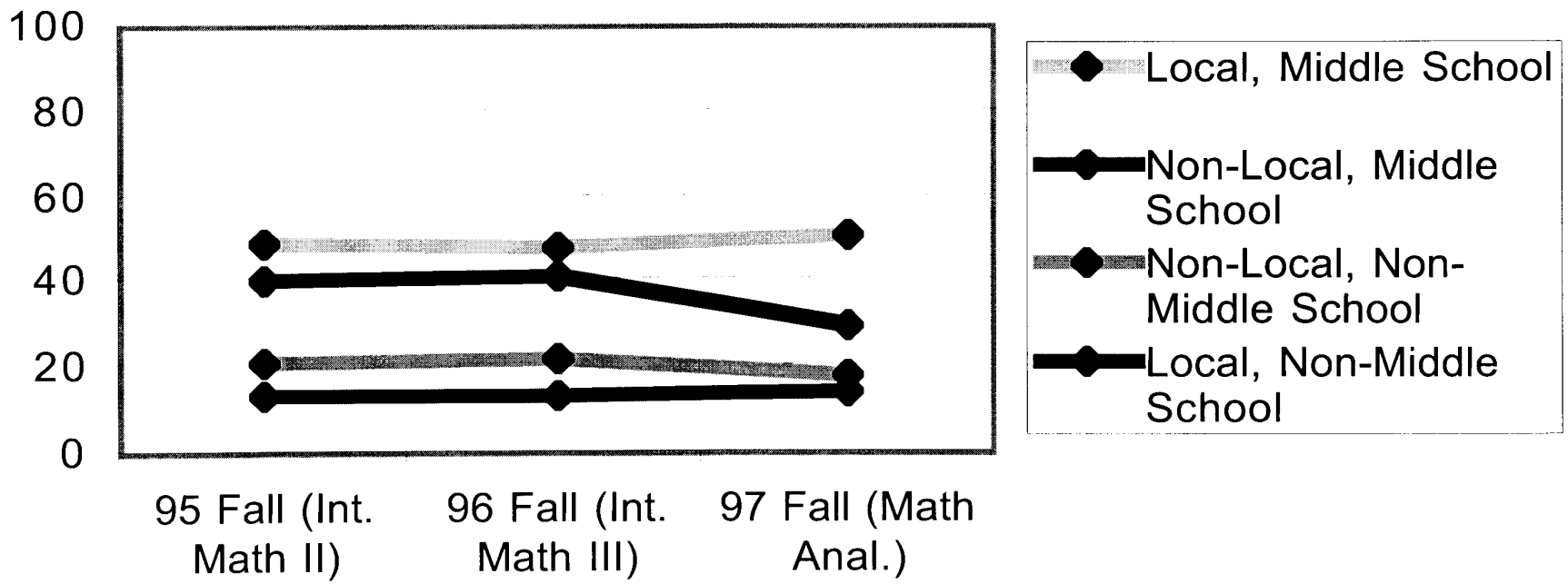
Non-Local, Non-Middle School: Math Course Taking Patterns for 1995 Entering Freshmen Cohort, Fall 1995



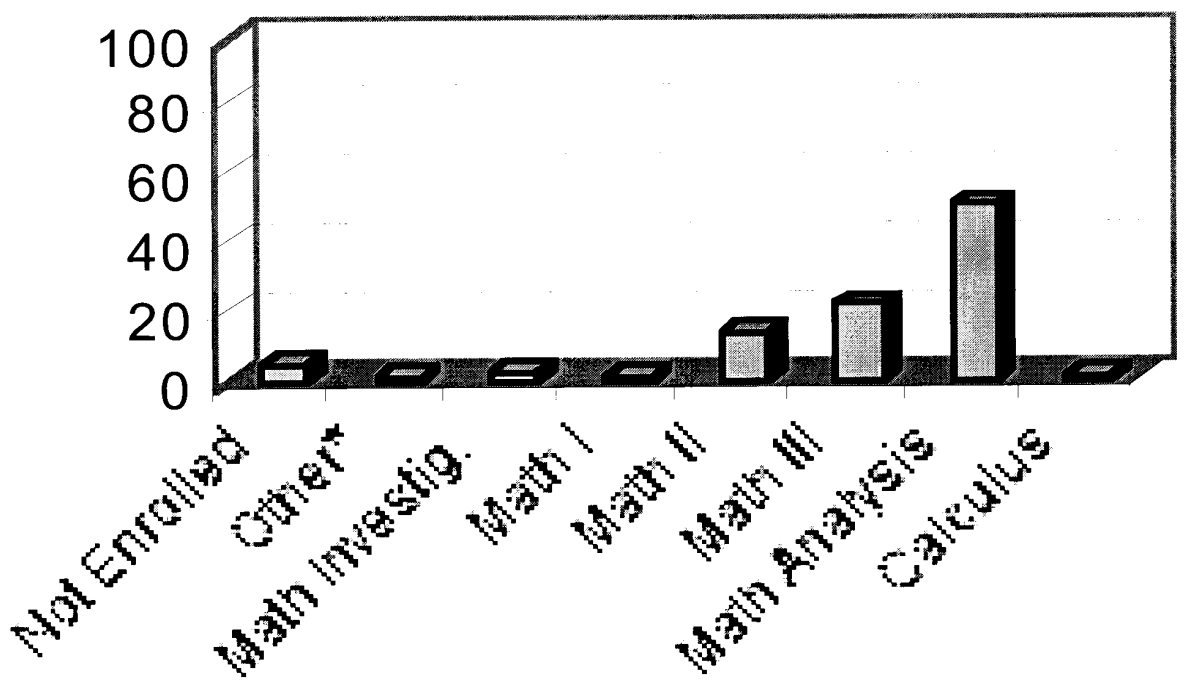
Percent on Track: Freshmen Year Enrollment in Integrated Math II, 1995 Cohort, Fall 1995



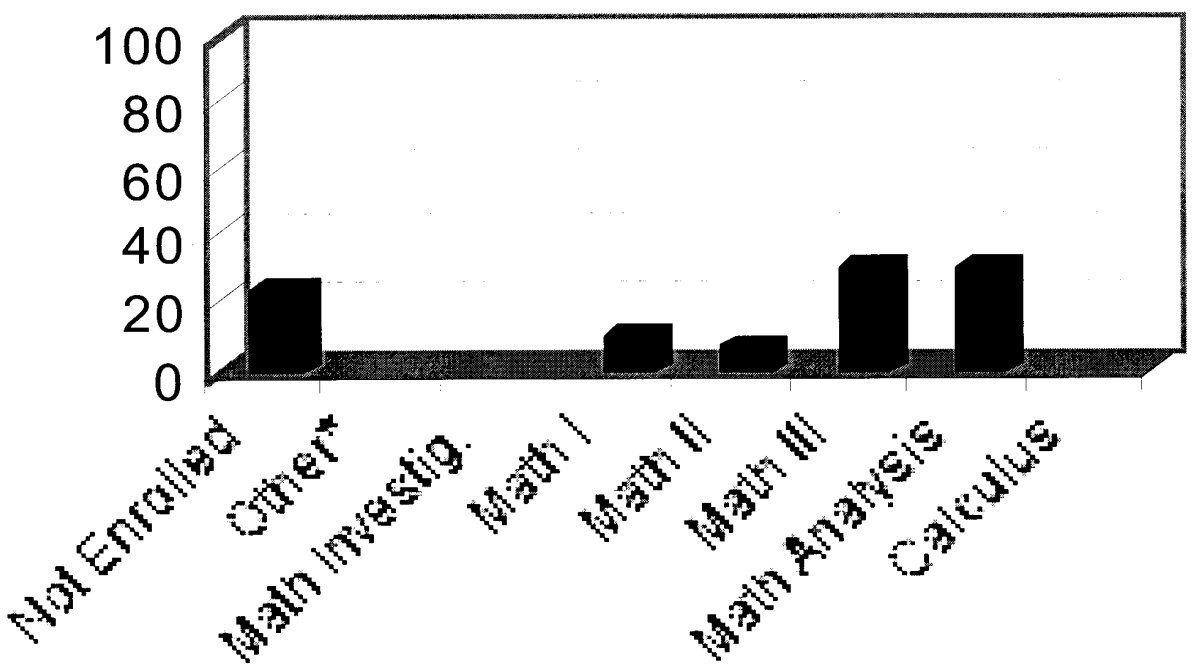
Math Course Taking Patterns: Progress for 1995 Entering Freshmen



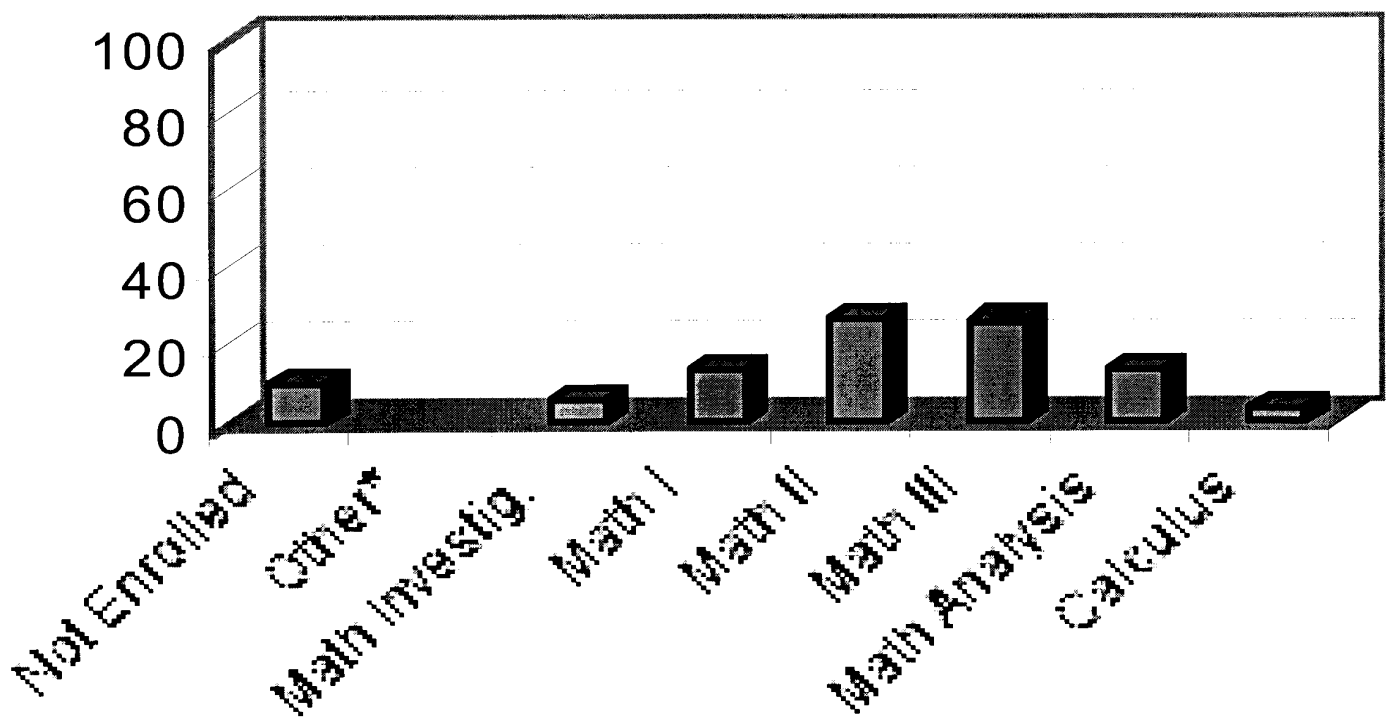
Local, Middle School: Math Course Taking Patterns for 1995 Entering Freshmen Cohort, Fall 1997



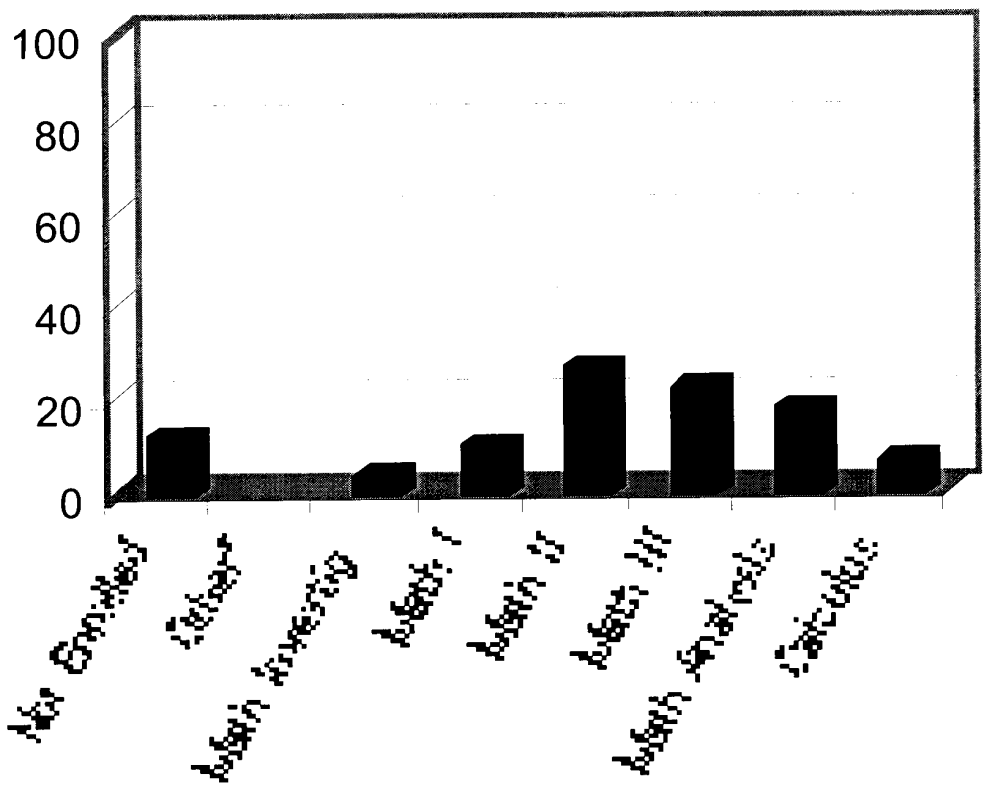
Local, Non-Middle School: Math Course Taking Patterns for 1995 Entering Freshmen Cohort, Fall 1997



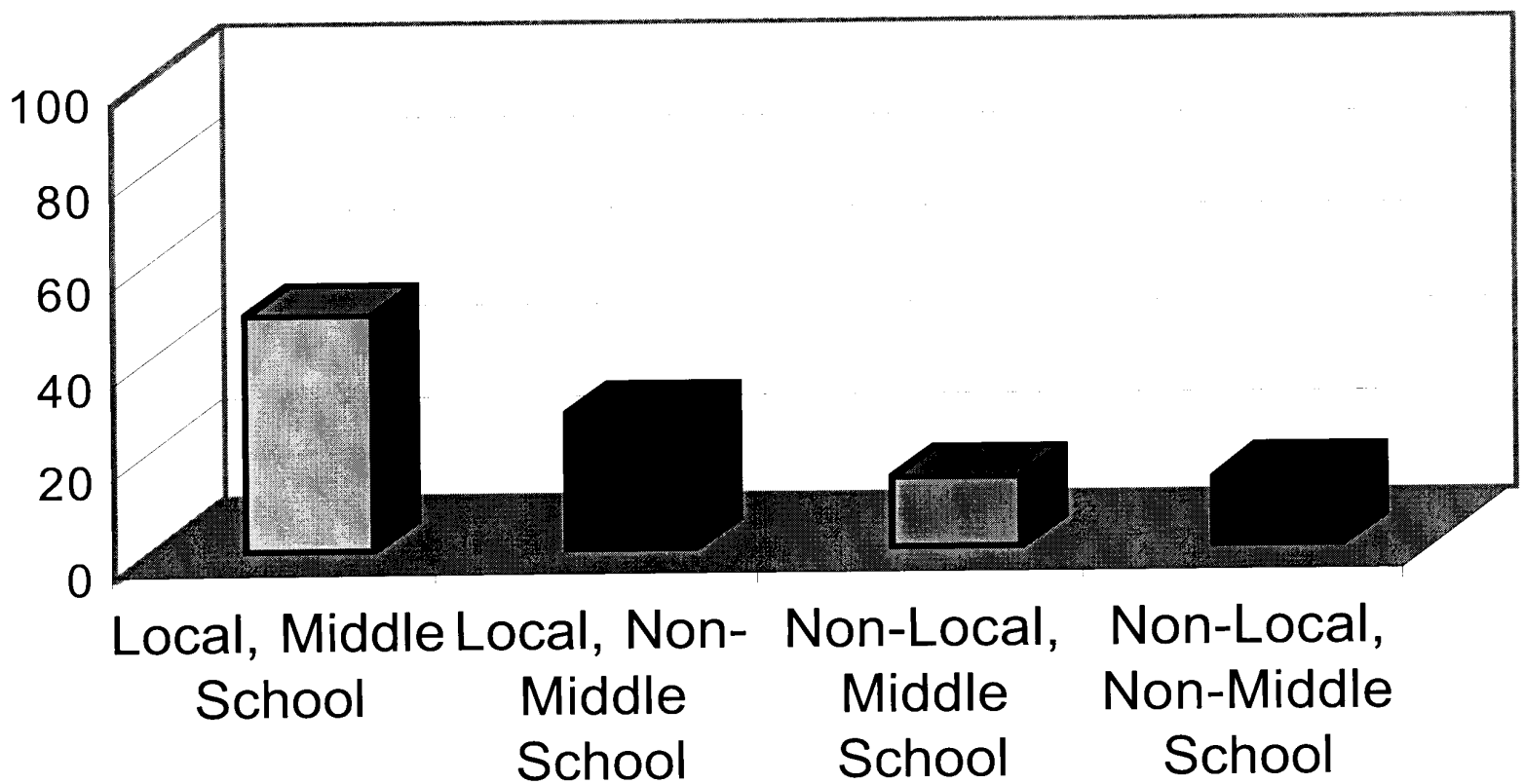
Non-Local, Middle School: Math Course Taking Patterns for 1995 Entering Freshmen Cohort, Fall 1997



Non-Local, Non-Middle School: Math Course Taking Patterns for 1995 Entering Freshmen Cohort, Fall 1997



Percent on Track: Junior Year Enrollment in Math Analysis, 1995 Cohort, Fall 1997



Appendix B: Course Materials

Syllabus

Course Project

Description Course Projects

Homework

Presentation Slides for Teachers and Administrators: Sessions 1-6

University of California, Los Angeles
Graduate School of Education and Information Sciences
Extension Education
Summer 1999

Using Data to Assess and Improve

Eva Baker, Ed.D.
Joan Herman, Ed.D.
Barry Gribbons, Ph.D.
Todd Ullah, Ed.D.

Office Location: **GSEIS #203**
Office Telephone: **310) 794-9167**
Class Location: **Sample High School**
Class Time: **T/TH 6-9pm**

Course Description:

This course is designed to help school staff build the capacity to use longitudinal indicators to stimulate inquiry, drive improvement plans, and inform stakeholders about the status of performance goals. During the course, discussions of critical questions, needs assessment, and goal setting set the stage for using electronic tools to graphically communicate data to school stakeholders. In the context of conducting various analyses, technical information about performance indicators is discussed.

Course Goals

Goal 1: To introduce students to data-based inquiry

Students will be able to:

1. describe framing questions to inquiry focused on schoolwide program development.
2. generate specific questions to guide inquiry in a school setting.
3. describe appropriate and inappropriate uses of data based on technical issues.

Goal 2: Introduce students to curriculum needs assessment

Students will be able to:

1. describe how to frame a review of curriculum, including issues of alignment, performance across measures and content areas, and distributions of performance.
2. describe implications for planning/implementation of curriculum and / or assessment changes based on data for a given school.

Goal 3: Introduce students to equity needs assessment

Students will be able to:

1. describe needs assessment related to equity issues and federal, state, and local mandates.
2. conduct appropriate analyses of traditionally underserved students status and progress.

3. describe the implications for program planning based on disaggregated data.

Goal 4: Introduce students to program evaluation

Students will be able to:

1. identify framing questions to program evaluation.
2. compare and contrast, including strengths and limitations, various approaches to program evaluation.
3. develop an appropriate evaluation design and implementation plan for a specific program in a school.

Goal 5: Introduce students to process of consolidating plans, goals, and indicators

Students will be able to:

1. assess alignment of schoolwide programs, indicators, and school improvement goals
2. develop an appropriate and feasible plan for tracking the progress of school and programs
3. identify strategies for involving key constituents in creating a shared vision
4. demonstrate ability to develop presentation materials using electronic tools.

Goal 6: Introduce students to software designed to support data-driven inquiry

Students will be able to:

1. apply software to analyses and presentation of information designed to support inquiry.

Course requirements:

1. Regular attendance.
2. Active participation in discussions and completion of HW assignments.
3. Satisfactory completion of term project (described separately).

Note:

This course is intended for professionals directly involved with program planning, implementation, and monitoring school's progress. Current assignment in a school is strongly advised to ensure that information from the course has direct utility to the school. However, others may participate provided they have had previous experience in school program development, implementation, and/or evaluation.

Recommended Text:

Herman, J., & Winters, L. (1992). Tracking your School's Success: A guide to sensible evaluation. Newbury Park, CA. Sage.

Grading Procedures:

Evaluation of students' knowledge and skills will be based on weekly assignments (40%) and a term project (60%).

Course Schedule:

Date	Topic	Activity
Aug. 10	Introduction, Framing Inquiry, Generating Questions, Technical Issues	Activity 1
Aug. 12	Curriculum Needs Assessment, Performance Across Content Areas; Implications; Additional Data Needs;	Activity 2
Aug. 17	Equity Needs Assessment; Mandates Related to Equity; Assessing Performance; Implications; Additional Data Needs	Activity 3
Aug. 19	Goal Setting. Characteristics of Good Goals; Process of Setting Goals; Indicators Measuring Progress; Technical Issues in Measuring Progress; Schoolwide Programs	Activity 4
Aug. 24	Program Evaluation; Framing Questions; General Approaches (strengths / limitations); Indicators for Monitoring Programs; Implementing Plans	Activity 5
Aug. 31	Cont. Program Evaluation. Analyses of Current Data	Activity 6 Draft Project Due.
Sept. 2	Consolidating Plans, Alignment; Implications	Activity 7
Sept. 4	Engaging Constituents; Shared Vision; Microsoft Powerpoint	Activity 8
Sept. 9	Review Lessons Learned (catch-up)	Final Project Due

Note: No class Aug. 26.

Using Data to Assess and Improve

Course Project

The purpose of the course project is to apply the content of this course to a specific school. In doing so, to a large extent the course project is a culmination of course assignments. Corresponding to each major topic in the course are the sections of the project: Needs Assessment, Goals, Program Plans, and Evaluation Plan.

The needs assessment section contains four elements. Appropriate data will be presented. A short narrative description accurately describing needs of the school based on these data will be provided. Implications for action are listed. Finally, materials are included for communicating the information and engaging stakeholders. These materials could include presentation slides and narrative descriptions of the process used to disseminate (faculty meetings, web page, etc.).

The goals section contains at least two sample goals for the school that meet the criteria discussed in class. Also included will be two narrative pieces. The first is a short section (paragraph or two) describing the relationship between these goals and the accountability system. Next, describe the process used to have the SBM group and faculty review, revise, and adopt goals. As with the previous section and subsequent section, include presentation materials.

In the program plans section, list existing school programs. Include a description of the alignment to goals and emphasis of resource allocation. From this description, write a short description of the implications for action. Last, include materials and a short description of the process for communicating the review of existing programs and engaging stakeholders, including SBM committee members, faculty, and others involved in each program.

For one school program, develop an Evaluation Plan. This plan should include a short description of the evaluation approach, including indicators to be used for monitoring effectiveness and evidence of success. Also, include a description of how different groups will be engaged in various components of the evaluation (writing, analyses, presentations, etc.). Include a timetable that includes data collection, analyses, reporting results, and revisiting program plans. A short specification of the resources needed to implement the plan should also be included.

Draft Projects are due Aug. 31st. Final projects are due Sept 9th.

Domain	Satisfactory	Excellent
Needs Assessment	• Contains data addressing curriculum and underserved groups of students. • Description is adequate. • Some presentation materials included.	• More than three indicators are used in the needs assessment. • Description of problem areas appropriately reflects relative strengths and weaknesses. • Implications for action follow from data. • Presentation materials are clear, engaging, and appropriately compelling.
Goals	• Two goals are listed. • Both goals meet criteria. • Some connection to Needs Assessment. • Some presentation materials included.	• Same as "Satisfactory" • Clear connection between goals and accountability system. • Description of process for adopting goals is clear and reasonable.
Program Plans	• Programs are listed. • Statement of alignment to goals included. • Indication of resource allocation included. • Some presentation materials included.	• Same as "Satisfactory" • Implications for discrepancies between programs, goals, and resources specified. • Reasonable process described for engaging constituents in review process. • Presentation materials are engaging.
Evaluation Plans	• General description of an appropriate evaluation design included. • Specific indicators for tracking success are identified. • Evidence of success is specified. • Timetable is specified.	• Description of evaluation plan is appropriate for the given program. • Indicators for monitoring progress and evidence of success allow for clear attributions to program effects. • Timetable and resource allocation are sufficient to allow for completion of tasks, and feasible.
Overall style	• Adequately conveys the material.	• Clearly communicates the information. • Details are provided in a concise manner. • Presentation materials are clear and engaging.

What Questions Can and Can Not Be Answered with Data?
With Available Data?

- What is the relationship between tests and curriculum?
- Do incentives / disincentives relate (or have an impact) to test scores?
- How does personal attitude relate to test scores?
- Does teacher credential status effect student performance?
- Has class size reduction increased student performance?
- What is the relationship between test scores and ELD level.
- What is a good way to explain how a student can score a 1 on SAT 9?
- How do the same students compare one year to next?
- To what degree do SAT 9 (& others) test items align with Standards / instruction?
- What is the relationship between SAT 9 scores and Performance Assessments?
- How do redesignated LEPs from Basic class compare to Model A?
- What percent of students that go to college that STAY in college – by ethnicity, gender, SES.
- Is there a positive correlation between performance on teacher / school / family assessments and performance on the SAT9?
- Which kids are leaving? Why?
- Are the Non-LEP students the same students as those who are non-economically disadvantaged?
- What are the other factors contributing to the performance of the LEP students?
- Who are the high school students?
- Where are they from? Feeder Schools?
- What are the graduation rates for students enrolled for entire 4 years vs. new students?
- What are the trends in proficiency pass rates?
- What is the relationship between passing / failing prof. Test and passing academic courses?
- What is the relationship between teacher pass rate, graduation rates, SAT / ACT scores.
- What are practices of effective teachers as evidenced by data over time?
- could one explain EOs under performing LEPs and RFEPs in grades 2-5?
- What has happened to high-achieving LEPs who <u>were</u> in primary language programs and are now in Model A or B?
- Is there a difference in test scores of students in Model A vs. Model B. (elem)
- What is the difference in achievement between transient students, and those who stay at one school for 3+ years.
- How many kids are really at grade level?Does heterogeneous grouping work?

Using Data to Assess and Improve

Homework: Needs Assessment

Due Aug. 12

For the next class session, please bring the following:
1. your school's improvement plan.
2. a list of indicators and types of analyses you are interested in.

Due Aug. 17

4. Briefly summarize and discuss the results. Include any of the following: What do they suggest about the relative strengths and weaknesses across content areas? Are the patterns consistent across grade levels? Across measures? Are we testing what's most important in the subject area? Is the measure aligned with our standards?

5. Are there significant differences across subgroups? Are the differences more pronounced at some grade levels than at others? What reasons might explain the differences?

6. What questions are raised by the data? What additional data or analyses need to be performed?

7. What are the implications for school planning?

8. How would you involve others in this process? Who would you involve?

Due Aug. 19

3. Bring specified data to class for analysis using the Quality School Portfolio. (electronic copies only)

4. Revise Needs Assessment including presentation materials such as a PowerPoint handout.

Using Data to Assess and Improve

Questions

- How have you used data in the past?
- What are some of the types of data used?
- What were the decisions to be made?
- What questions do you have that may be answered by data?

Context

- Title I / AYP
- SB1X / API
- Superintendent's Accountability
- Annenberg
- Other
- Standards-based

Available Data

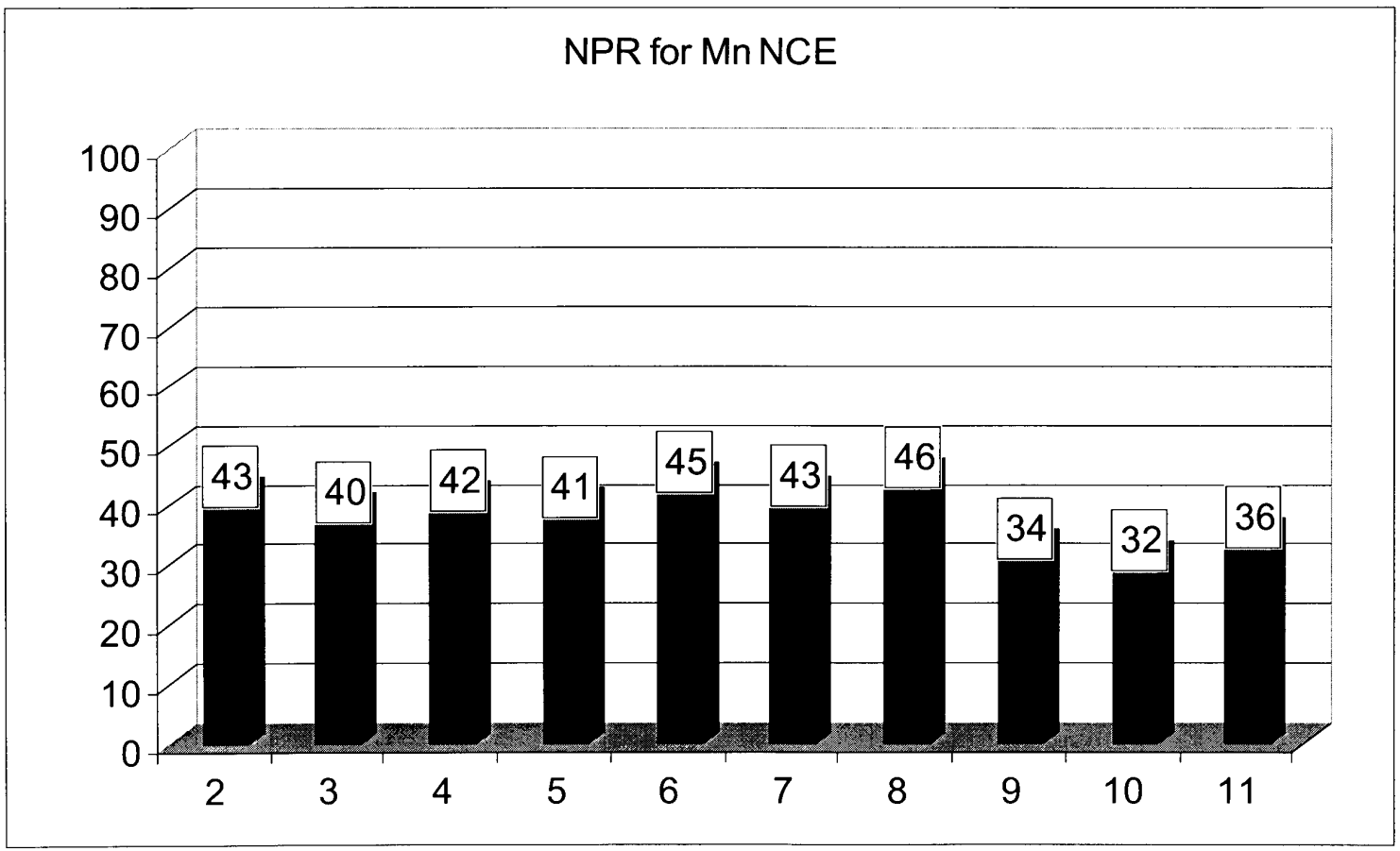
- CDE
- SARC (District)
- Student-level
  - *Longitudinal database*
  - *SIS*

CDE

Types of Data

- SAT-9
  - *Strengths*
  - *Limitations*

SAT-9: Why the differences?



Other Measures

- Attendance
- Graduation rates
- Drop-out rates
- Redesignation rates
- Local assessments

Implications of Accountability Context

- Framing Inquiry
- Defining Measures

HW

- School improvement plans
- List of data and analyses

Session 2: Needs Assessment

Needs Assessment

- Curriculum Areas
- Underserved Groups
- Interaction of Curriculum and Groups
- Other, Inquiry Generated

Approaches to Curriculum Assessment

- Alignment: What of Value Is and Is Not Being Tested?
- Patterns of Performance:
  - *Across Tests*
  - *Across Subscales*

Performance Across Tests

- What Is Being Assessed by Each?
  - *Emphasis? Missing?*
  - *Alignment to Standards*
- Is There Differential Performance Across Tests?
- Implications for Curriculum Planning

Performance Across Subscales

- Do the Subscales Represent Important Foci?
- Is There Differential Performance Across Subscales?
- Caution: Unreliable Subscales

Implications

- Curriculum Planning / Revision
- Assessment Planning / Revision

Examples

- http://star.cde.ca.gov/
- http://goldmine.cde.ca.gov/
- http://www.lausd.k12.ca.us/

Equity

- Serving All Students
- Framing the Problem
- Goal
  - *No Excuses*
  - *Increase Attaining Standards*
  - *Decrease Performance Gap*

How Can we improve effective learning opportunities for underserved students?

- Data raises the question and imperative--answers require reflection and additional assessment
- What are obstacles?
- What are potential strategies for overcoming them?

Who Are Students Most at Risk?

- At Risk of Dropping Out of School?
- At Risk of Dropping Out of the College Pipeline?

Session 3: Needs Assessment: Going Deep

Objectives

- Review ideas for additional data
- Generate set of additional questions
- Identify specific fields
- Review downloading data
- Demonstrate QSP
- Review PowerPoint and Excel

Building on Needs Assessment

- Summary of results: Strengths/weaknesses
  - *Across Groups*
  - *Across Content Areas*
  - *Alignment*
- Additional questions, data, and analyses

Additional Data: Some ideas

- Stuart
- LAAMP
- Technology, Digital High Schools
- Classroom practice
- Focus groups
- Observations
- Other

Considerations

- What is the cause of the problem?
- What are the implications for planning?
- Who needs to be involved in:
  - *identifying problems,*
  - *conducting analyses,*
  - *identifying solution?*
- How to connect to classroom practice?

Session 4: Needs Assessment: and Goal Setting

Session Agenda

- Needs Assessment
  - *Issues*
  - *QSP Tables*
  - *School Summary*
- Goals
- PowerPoint

Needs Assessment and QSP

- Frequency Table
- Cross-tabulation
- Questions Table

Setting Specific Performance Goals on Priority Indicators

- Where do we want to go in terms of specific performance?
- How long should it take us to get there?
- Where should we be in the next year?

  No "right answer"

  Judgmental process, in conjunction with mandates

What Are Reasonable Performance Goals?

- Challenging
- Realistic
- Specific
- Measurable
- Stated completion date

Needs Assessment Cont'd

Objectives

- Finish Four Components of the Needs Assessment prior to goal setting
- Discuss Strategies/Implications for School Planning
- Summaries/Questions/Develop Project Outline
- Use QSP to generate key Tables and Charts
- Goal Setting and Activity
- PowerPoint as a presentation tool

Building on Needs Assessment

- Summary of results: Strengths/weaknesses
  - *Across Groups*
  - *Across Content Areas*
  - *Alignment*
- Additional questions, data, and analyses

Additional Data: Some ideas

- Stuart
- LAAMP
- Technology, Digital High Schools
- Classroom practice
- Focus groups
- Observations
- Other

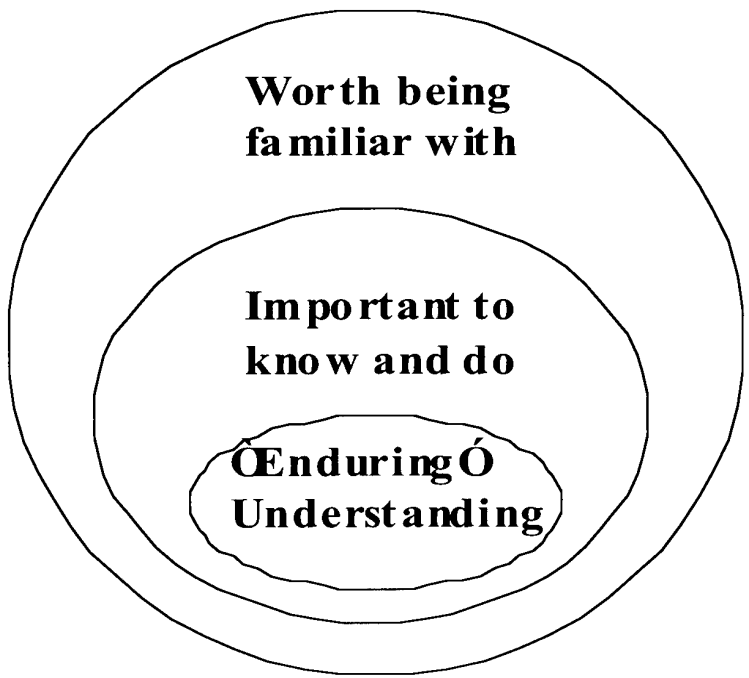
Other Measures

- Attendance
- Graduation rates
- Drop-out rates
- Redesignation rates
- Local assessments

Considerations

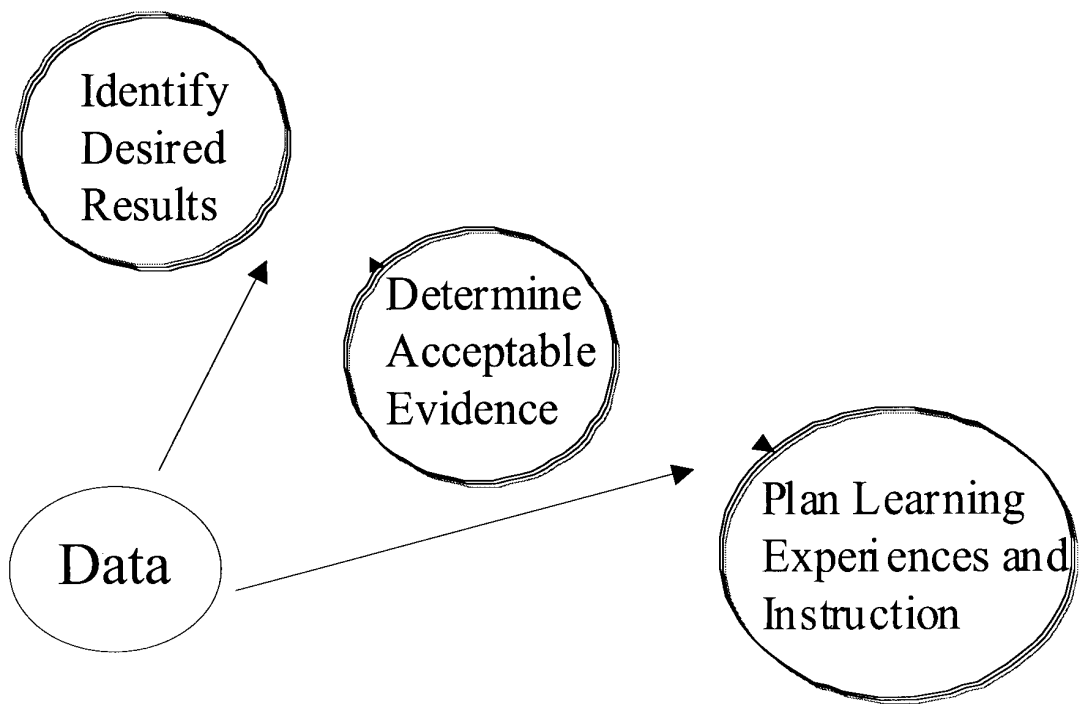
- What is the cause of the problem?
- What are the implications for planning?
- Who needs to be involved in:
  - *identifying problems,*
  - *conducting analyses,*
  - *identifying solution?*
- How to connect to classroom practice?

Establishing Curricular Priorities



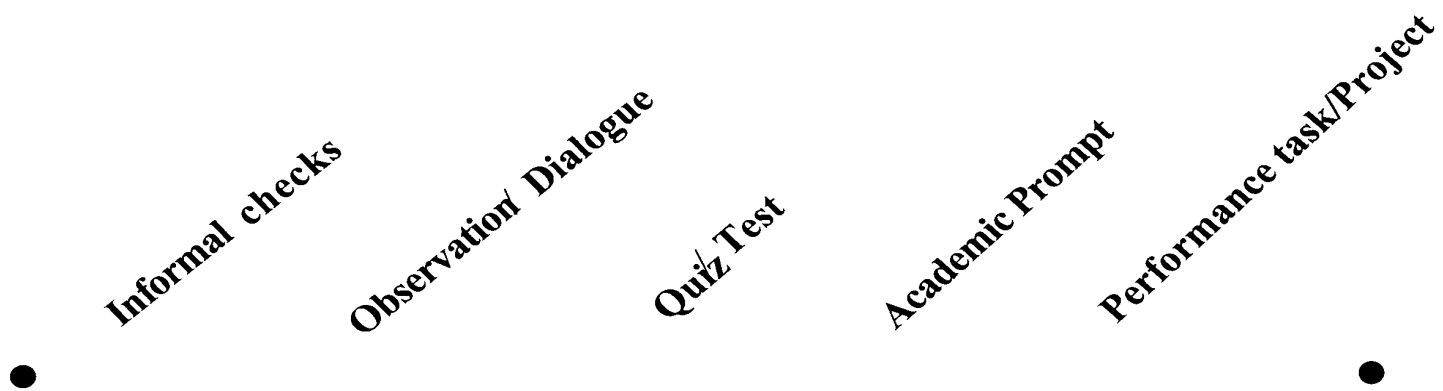
Understanding

STAGES IN THE BACKWARD DESIGN PROCESS



Continuum of Assessment Methods

By
Grant Wiggins and Jay Mc Tighe



CRESST/Palisades

IMPLICATIONS FOR PLANNING

Classroom Practice

▼

School Performance Indicators Drive:

- ¥WASC Accreditation/ ESLRs
- ¥School Improvement Plans
- ¥DHS Evaluation
- ¥LEARN Plans
- ¥Data Driven Resource Allocation
- ¥CHARTER School Plans/Evaluation

▼

High Standard and Assessments

Needs Assessment and QSP

- Frequency Table
- Cross-tabulation
- Questions Table

Setting Specific Performance Goals on Priority Indicators

- Where do we want to go in terms of specific performance?
- How long should it take us to get there?
- Where should we be in the next year?

No "right answer"

Judgmental process, in conjunction with mandates

What Are Reasonable Performance Goals?

- Challenging
- Realistic
- Specific
- Measurable
- Stated completion date

Program Evaluation

Common Elements to Plans

- Mission, Vision, Beliefs
- Needs Assessment
- Improvement Plans
  - *Professional Development*
  - *Curriculum and Instruction*
  - *Parent and Community involvement*
- Evaluation Plan
- Resources

Framing the Evaluation

- Needs Assessment to Interventions
- Implementation
- Impact
- Improving Programs

Connecting Needs Assessment, Goals, and Programs

- Theory of Action (Schon and McDonald)
  - *Espoused*
  - *Design*
- Aligning Needs, Goals, Interventions, Resources

Implementation

- To what degree was the program implemented?
- Did programs get implemented on schedule?
- Are participants actually participating as intended?
- If not implemented as intended, why?
- What constitutes evidence of implementation?

Documenting Approach and Process

- Existing Documents
  - *School Improvement Plans*
  - *Logs*
  - *Attendance records*
  - *Session evaluations*
  - *Other*
- Additional Data Needs

Impact

- To what degree have goals been attained?
- What are other lines of evidence of success?
  - *Designs*
- What unintended positive and negative consequences were realized?

Impact: Evidence of Success

- Attaining Goals
- Comparisons
  - *Before and After*
  - *Participants and Similar Students*
- Multiple Measures
- Multiple Time Points
- Triangulation
  - *Quantitative and Qualitative*

Impact: Data Sources

- Available Data
- Additional Data Needs

Improving Programs

- Were programs implemented as intended?
- To what degree were goals obtained?
- What unintended consequences were realized?
- Why did program have intended effects?
- What midcourse corrections should be made?
  - *Adapting existing programs*
  - *Other options*

Evaluation Plans

- Constituents
- Products
- Tasks
  - *Data collection*
    - *Documents*
    - *Available data*
    - *Additional data*
  - *Analysis*
  - *Reporting*
- Resources
  - *School staff*
  - *District*
  - *Others*
- Management Plans
  - *Responsible*
  - *Timeline*

Engaging Constituents

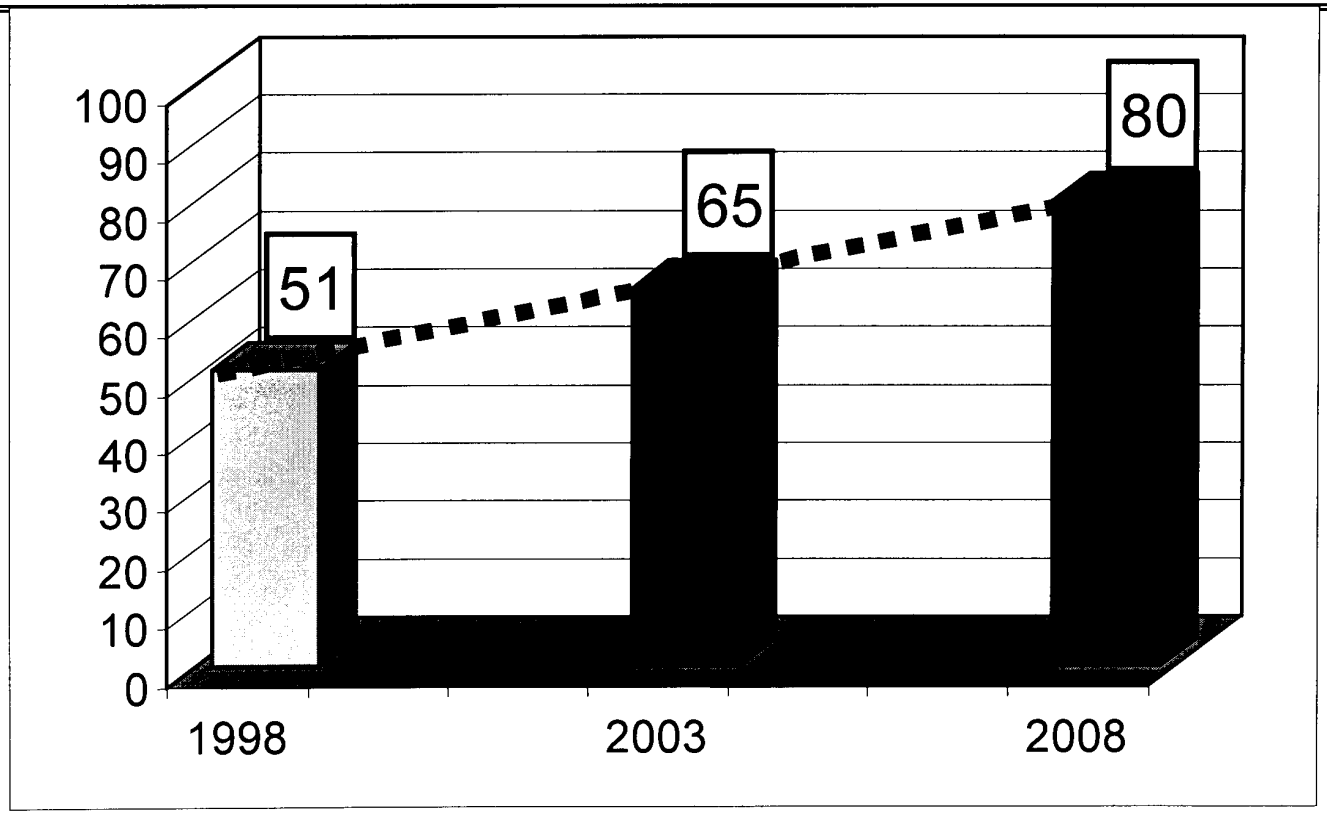
- When?
- How?
- Why?

Review of Projects
Suggestions / Themes

Patterns across Drafts

- Goals
- Longitudinal Designs
- Graphical Representations
  - *Bar*
  - *Stacked Bar*
  - *Labels*
- School Planning/Evaluation Design Tables
- Technical Issues

Sample Goal A:
SAT-9 Reading 11th grade
3% Per Year Increase



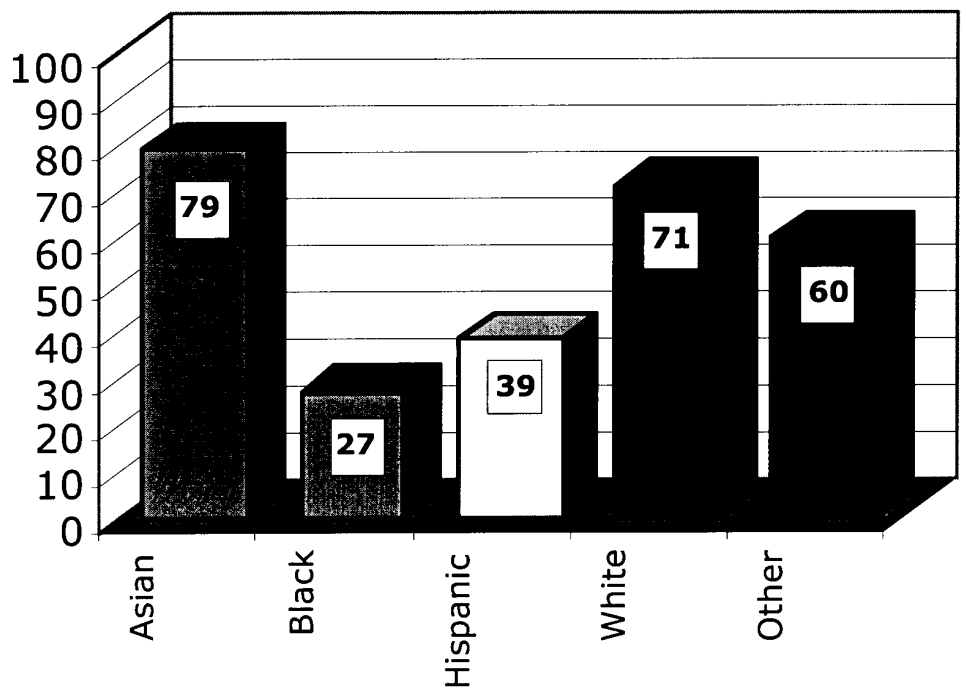
Designs

- Cross-Sectional
  - *E.g. 3rd grade to 3rd grade*
- Quasi-Longitudinal
  - *E.g. 3rd grade to 4th grade*
- True Longitudinal
  - *E.g. Longitudinal 3rd grade to 4th grade*

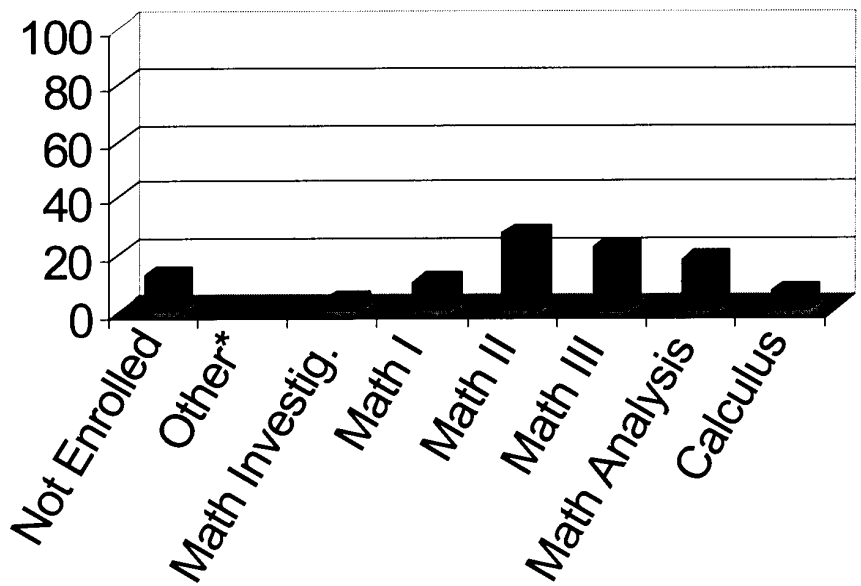
Graphical Representations

- Bar
- Stacked Bar
- Line
- Scatter plot

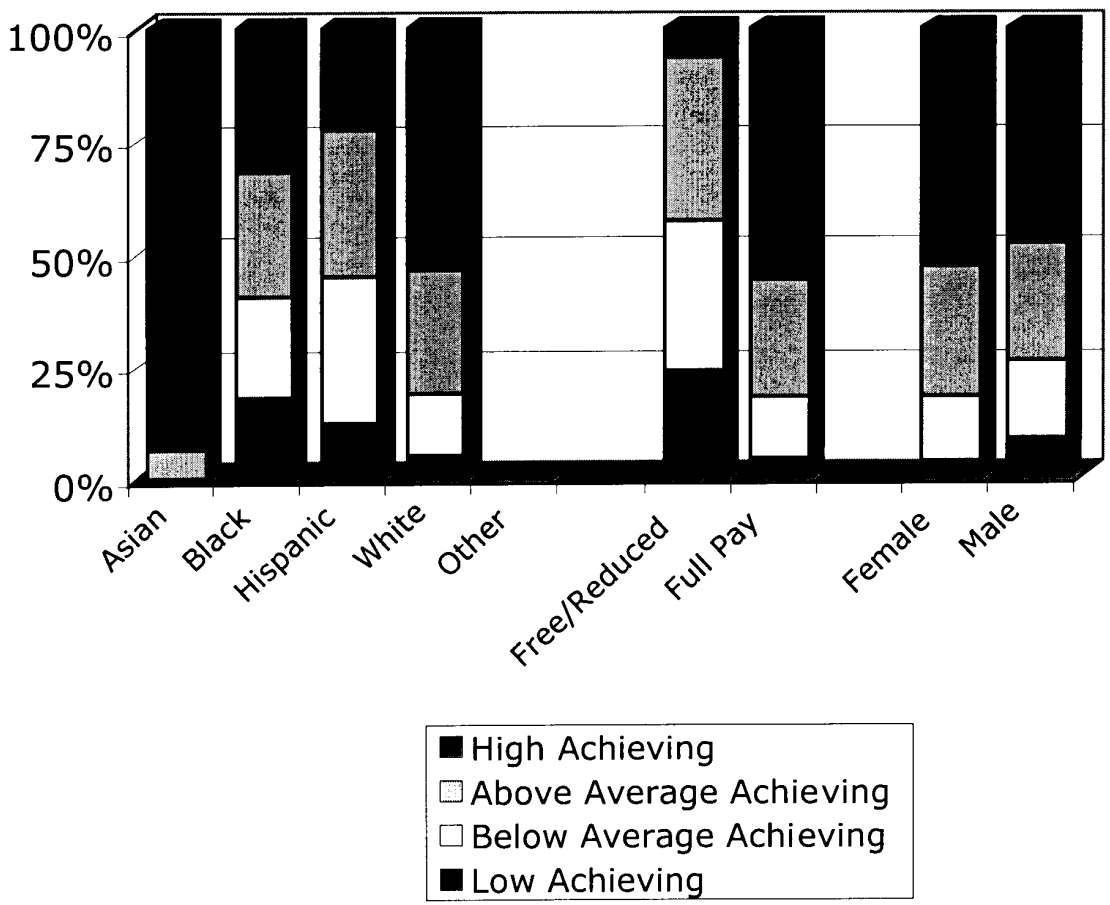
Graduates: Percent Meeting UC Eligibility Requirements by Ethnicity

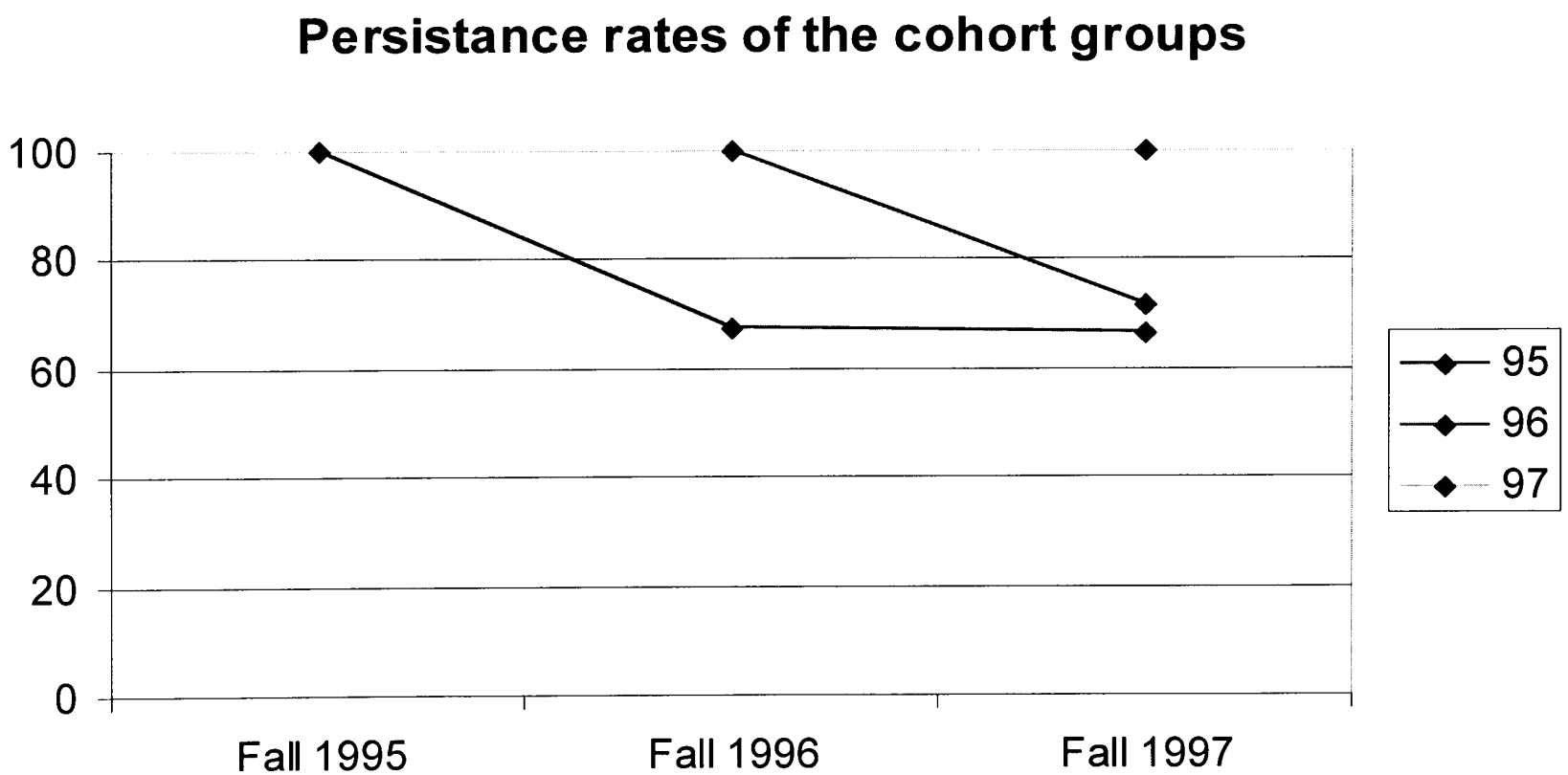


Math Course Taking Patterns
1995 Entering Freshmen Cohort, Fall 1997

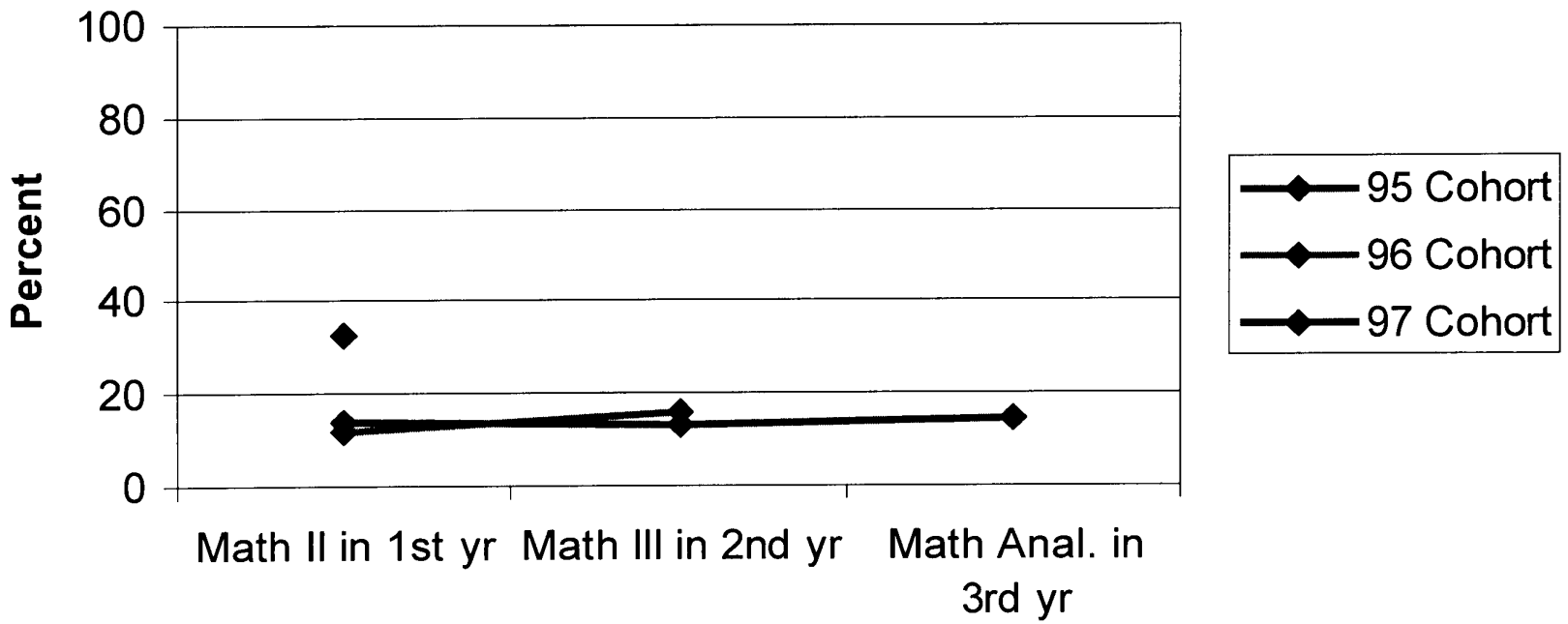


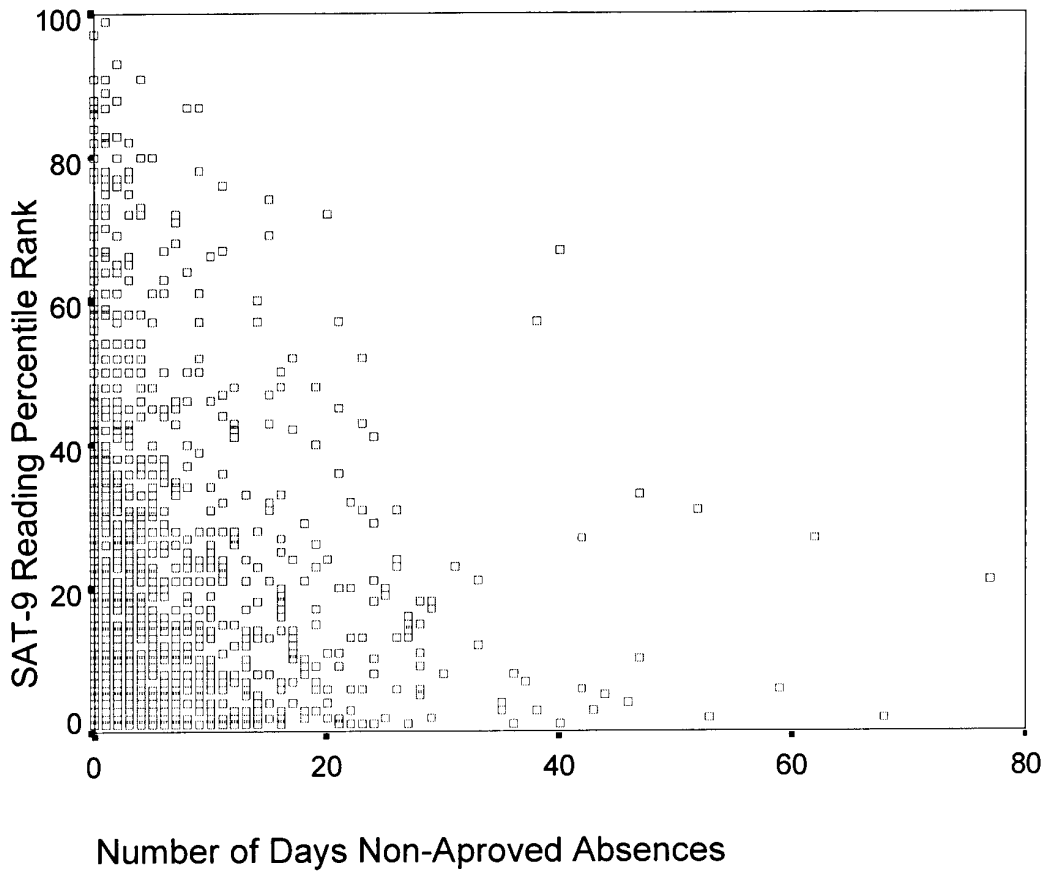
SAT 9 Reading: All Grade 3 Students

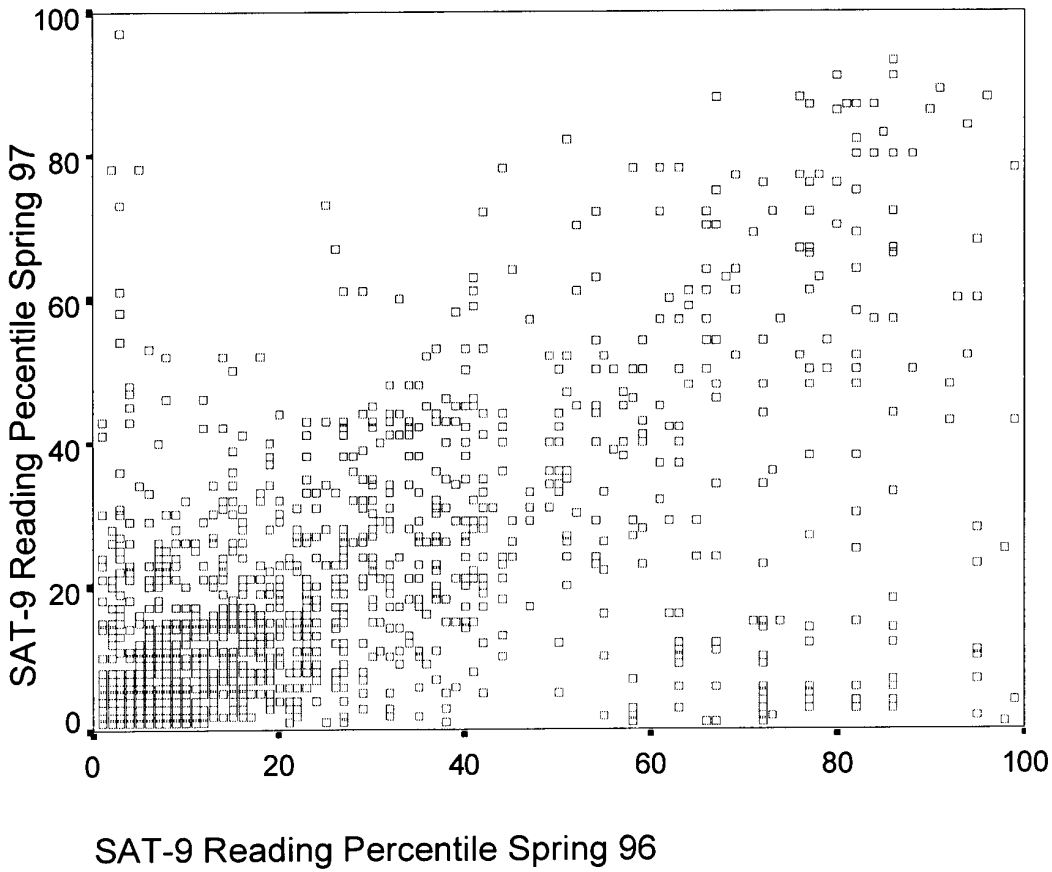



Persistance rates of the cohort groups
100
80
60
40
20
0
Fall 1995
Fall 1996
Fall 1997
95
96
97

Math Course Taking Patterns
Percent "On Track"




100
80
60
40
20
0
SAT-9 Reading Percentile Rank
0
20
40
60
80
Number of Days Non-Aproved Absences


SAT-9 Reading Pecentile Spring 97
100
80
60
40
20
0
0
20
40
60
80
100
SAT-9 Reading Percentile Spring 96

School Planning / Evaluation Design Tables

- LAAMP Learning Plan
- Evaluation Design Chart
- Others

Technical Issues

- NRT's
  - *Changing Tests / Forms*
  - *Changing Norm Group*
    - *Over time*
    - *Across grade levels*
- Sample Size