Monitoring Student Performance and Change with Modified California Academic Performance Index

Joseph Adwere-Boamah
Alliant International University

1310 Portland Ave.
Albany, California 94706

joeadwb@pacbell.net

Abstract

California's Public School Accountability Act of 1999 (PSAA) established, among other goals, improved academic achievement of all students in the state. It requires California public school students in grades 2 through 11 to take a single standardized achievement test, the Stanford Achievement Test Form T (SAT 9), each spring to measure achievement in basic skills. The current strategy for monitoring a district's academic progress is through a comparison of the performance of successive groups of students at selected grade levels, the successive cohort approach. This method compares the percent of students (e.g., second graders in a district) above the 25^{th}, at or above the 50^{th} or above the 75^{th} percentile in a particular content area in one year to the percent of students at the same grade level who performed at or above the selected percentile points in the previous years.

The successive cohort approach for monitoring student progress has intuitive appeal. However with the current emphasis on improving the achievement of low performing students in California, the three percentile points selected for reporting district data do not allow for greater precision in observing the changes in performance of student subgroups in the district. Specifically, the use of the three percentile points for monitoring student progress does not enable the district to determine just where progress is occurring or not occurring within the full achievement range.

This article demonstrates the utility of a modified API, which enables an administrator to monitor student performance and change throughout the score scale rather than at three percentile points.

Monitoring Student Performance and Change with Modified California Academic Performance Index

Introduction

California has a long history in using statewide assessment as a policy tool for program improvement. For three decades, California students took a statewide test known as the California Assessment Program (CAP). The California Learning Assessment System (CLAS) replaced this program in the early 90s when authentic assessment was in great demand. However unlike CAP, CLAS did not last long and was discontinued in 1995.

The most recent form of State testing in California public schools came from the Standardized Testing and Reporting (STAR) Program authorized through Senate Bill 376 in 1997. It requires California public school students in grades 2 through 11 to take a single standardized achievement test, the Stanford Achievement Test Form T (SAT 9), each spring to measure achievement in basic skills. Students in grades 2 through 8 are tested in reading, mathematics, language and spelling. Grades 9 through 11 are tested in reading, mathematics, language, science and history-social science (Educational Data Partnership 2002).

California's Public School Accountability Act of 1999 (PSAA) established, among other goals, improved academic achievement of all students in the state. The main instrument for assessing student achievement, SAT 9 was first administered in spring 1998. For accountability purposes, the State Department of Education publishes SAT 9 results as Academic Performance Index (API) scores on a scale using integers 200 through 800.

The API is a weighted index based on the distribution of student scores in five performance bands or levels. Specifically, the weighted index is equal to 200 times the percent of students (e.g., grades 2-5 in a typical elementary school) in performance level 1 (1-19th percentile), plus 500 times the percent in performance level 2 (20-39th percentile), plus 700 times the percent in level 3 (40- 59th percentile), plus 875 times the percent in level 4 (60- 79th percentile), plus 1000 times the percent in level 5 (80-99th percentile). The total weighted score for each content area (e.g., reading, language, spelling and mathematics for grades 2-8) is multiplied by its weight and summed across the content areas to obtain the API for a school[1] (California Department of Education 1999).

[1] The content weights for grades 2-8 are: Reading =30%, Language =15%, Spelling =15% and Mathematics =40%. The weights for grades 9-11 are: Reading =20%, Language =20%, Mathematics =20%, History-Social Science –20% and Science =20%.

API is used to rank schools, compare schools with similar characteristics, establish growth targets for schools and to monitor schools' progress toward meeting their targets. The California Department of Education also reports the SAT 9 results as percent above percentiles. This reporting method shows the percent of students (at site, state or district level) scoring above the 25^{th}, at or above the 50^{th} or above the 75^{th} percentile.

API has solved one of the persistent problems that faced many states on reporting statewide assessment results that are comprehensible, informative and useful to both the public and state decision makers. For example, API is used for selecting schools for the Immediate Intervention/ Under Performing School Program (II/USP) and the Governor's Performance Award Program (GAPAP). While API is an appropriate tool at the state level for monitoring schools' progress, it is not an effective monitoring instrument for superintendents or site administrators who need meaningful information about specific programs or student subgroups for making decisions. School leaders must assess and monitor specific instructional programs, i.e.: reading, math, language and science. API combines student achievement data across multiple grades and content areas. As such it does not provide information on specific programs designed to improve student achievement. It masks the relative strengths and weaknesses of students in specific content areas and at specific grade levels. Another drawback of API is that it does not provide a school district with a means or an index for directly tracking its progress.

The current strategy for monitoring a district's academic progress is through a comparison of the performance of successive groups of students at selected grade levels, the successive cohort approach. This method compares the percent of students (e.g., second graders in a district) above the 25^{th}, at or above the 50^{th} or above the 75^{th} percentile in a particular content area in one year to the percent of students at the same grade level who performed at or above the selected percentile points in the previous years.

The successive cohort approach for monitoring student progress has intuitive appeal. However with the current emphasis on improving the achievement of low performing students in California, the three percentile points selected for reporting district data do not allow for greater precision in observing the changes in performance of student subgroups in the district. Specifically, the use of the three percentile points for monitoring student progress does not enable the district to determine just where progress is occurring or not occurring within the full achievement range.

The purpose of this article is to demonstrate the utility of a modified API, which enables an administrator to monitor student performance and change throughout the score scale rather than at three percentile points.

Performance Level Data from API

In order to compute API from student achievement scores, the SAT 9 national percentile scale is first partitioned into five achievement levels that cover the full score continuum (1st through 99th percentile). The achievement scores of students on the test are defined by percentile groupings as follows:

API Performance Levels

Figure 1

Percentile Rank	Performance Level	Label
80-99	5	Advanced
60-79	4	Proficient
40-59	3	Basic
20-39	2	Below Basic
1-19	1	Far Below Basic

The distribution of student achievement scores in the performance levels provides the means for calculating schools' API's as described above. But most importantly, it can provide site administrators a tool for monitoring trends in student achievement and for establishing school improvement goals. Unlike the "Percent At or Above the National Average" approach, which contains very little relevant information for program monitoring, performance level scores contain the essential information for monitoring student performance and change. In particular they reveal 1) how low the low achieving students are performing, 2) how high the top students are achieving and 3) concentrations of low and or high performing students in the group.

The labels of the performance levels (see Figure 1) provide administrators with a criterion-referenced interpretation for describing and reporting school/student achievement (Education Data Partnership 2002). The five categories of performance levels as opposed to the usual "Below, At/Above National Average" provide better descriptions of school/student performance. For example, a school can use the distribution of its spring 1998 student scores in the five performance levels as a baseline or benchmark for comparing student achievement in subsequent years. It can also use them to establish growth targets as standards. Such comparisons allow site administrators to determine the progress of students at each performance level, i.e., whether student groups in successive years are gaining, remaining steady, or losing ground relative to the spring 1998 baseline performance. The baseline performance can also be used along with long-term goals to establish annual improvement targets. The observed year-to-year changes in student performance can be interpreted in light of the selected targets for reaching long-term goals. It is

precisely this type of meaningful information needed by school leaders that neither API nor "Percent At/Above National Average" reporting provide.

Monitoring Year-to-Year School Progress with Performance Level (PL) Data

The remainder of this article focuses on the use of performance level (PL) data for monitoring instructional programs. The data for this study come from a large urban school district in northern California. The district has more than 50,000 students in 92 schools. More than 3,500 students at each grade level in the elementary schools participate in the testing program each year. The racial/ ethnic composition of the student body is: African-American 48.4%, Hispanic 26.2, Asian 16.3, White 5.6, Other 3.1.

At present, California schools have four years of SAT 9 achievement test data (spring 1998-2001). The four years of accumulated achievement test scores of the district were used to compute the proportion of students in each performance level (PL data). Successive cohorts of two benchmark grade levels, 2 and 5, were selected for this study. Specifically, 1998 reading and mathematics PL data of grades 3 and 5 students were compared to reading and mathematics PL data of successive 1999, 2000 and 2001, 3rd and 5th graders. We have presented reading PL data in Table 1 to illustrate how the PL approach, instead of the API, can be used to obtain essential information which is simultaneously comprehensible and useful to site administrators who must monitor school/program or student subgroup improvement.

The PL data in Table 1 show how successive grades 3 and 5 students distributed themselves along the reading performance continuum. For example, of the 4,435 grade 3 students who took the test in 1998, approximately 48% were in Level 1, 23% were in Level 2, the percentages for Levels 3, 4 and 5 were 12%, 9% and 8% respectively.

For 2001 grade 3 students the percentages were: Level 1, 35%, Level 2, 27%, Level 3, 15%, Level 4, 12% and Level 5, 11%. The distributions of reading achievement of the successive cohort groups provide a means of posing several related questions relevant to monitoring and reporting school progress. For example, how much has student achievement in reading changed in the past four years; are there differences in the nature of changes within the groups at different performance levels? Are most of the performance values clumped together at one end of PL with a few very high values?

The reading achievement PL data show some improvement in the achievement of the successive grades 3 and 5 students. In particular, the percentage of students who are far below basic (PL 1), decreased from 48% to 35% (13 percentage points) at grade 3, and the corresponding decrease at grade 5 is 8%. Student achievement at PL 2, 3, 4 and 5 did not show appreciable improvement within the past four years. The small

changes in the distribution of reading achievement in the levels suggest that student achievement at the two grade levels has remained almost steady in the past four years.

Table 1

Successive Cohort Group Percentages of Students Scoring at Each Performance Level on SAT 9 in 1998, 1999, 2000, 2001

Reading

Year	Grade	PL 1	PL 2	PL 3	PL 4	PL 5	Total
1998	3	48.4	22.9	12.0	9.2	7.5	4435
	5	48.0	21.4	13.1	9.9	7.7	4042
1999	3	31.7	23.9	17.4	14.5	12.5	3283
	5	41.3	22.0	15.9	11.4	9.4	3681
2000	3	38.4	25.2	15.8	10.7	9.9	4188
	5	43.8	21.4	15.0	10.4	9.5	3887
2001	3	35.4	26.8	15.1	12.1	10.6	4574
	5	40.3	23.0	17.4	10.3	9.1	4402

Table 2 shows the distribution of mathematics achievement of the same successive cohort groups, grades 3 and 5. The groups registered greater improvement in mathematics than in reading. At performance Level 1, the baseline percentages of the two grades were reduced by approximately 16 percentage points. Performance Level 5 also shows modest percentage increases, 8% and 6% for grades 3 and 5 respectively. The pattern of changes in reading and mathematics achievement in the two tables are very similar. With exception of grade 3 students who are at advanced level (L 5) in mathematics, both distributions show no appreciable changes in student achievement at Levels 2, 3, 4 and 5. The PL data in the two tables indicate quite clearly that in both reading and mathematics there was and still is a heavy concentration of students in Far Below Basic level. The annual percentage change in reading or mathematics achievement is so small that without substantial improvement in the quality of instruction, the achievement distribution of 2008 grades 3 and 5 cohorts will remain about the same as it is now. The reductions in percentage of students at Far Below Basic are not accompanied by increases in the proportion of those at proficient and advanced levels.

Table 2

Successive Cohort Group Percentages of Students Scoring at each Performance Level on SAT 9 in 1998, 1999, 2000, 2001

Mathematics

Year	Grade	PL 1	PL 2	PL 3	PL 4	PL 5	Total
1998	3	45.3	22.3	11.9	11.3	9.3	4571
	5	48.4	20.4	12.2	9.7	9.3	4179
1999	3	32.5	21.8	14.9	15.3	15.5	3349
	5	39.1	22.4	13.5	12.1	12.8	3721
2000	3	30.7	22.5	14.7	15.6	16.5	4222
	5	36.6	22.9	15.1	11.4	14.0	3970
2001	3	29.4	22.5	15.2	15.8	17.0	4613
	5	32.3	23.9	15.5	13.0	15.3	4466

The successive cohort grade 3 mathematics achievement data have been disaggregated by ethnicity to illustrate the use of PL data to study differential impact of programs on student subgroups. As shown in Table 3, all four ethnic groups (African-American, Hispanic, Asian and White), registered impressive reductions in the percentage of students at Far Below Basic level. The approximate Level 1 percentage reductions for the groups were: African-American, 18%, Hispanic, 20%, Asian, 14% and White, 9%. At the Advanced level, African-American and Hispanic students showed modest percentage increases (5% and 4% respectively), as contrasted to the improvements made by Asian (27%) and White students (17%). The percentage changes at PL 2, 3 and 4 are relatively small in magnitude for all the groups. However, they reveal an important pattern of improvement in mathematics achievement within the successive cohorts of the ethnic groups. For example, both African-American and Hispanic students show percentage increases at PL 2 and 3. Asian and White students on the other hand registered percentage decreases at the same performance levels. At the Advanced level, only White students showed percentage decrease. Negative changes (decreases) in the percentage of students in Levels 1, 2, and 3 accompanied by positive changes (increases) in the percentage of students at proficient and advanced levels indicate improvement in student performance across all the five performance levels. The data presented in Table 3 show that such overall improvements were made by Asian and especially by Whites, although the size of the percentage changes in Levels 2 and 3 were small compared to the changes in Levels 1 and 5.

Table 3

Successive Grade 3 Cohort Groups
Percentages of Students Scoring at each PL on SAT 9 by Ethnicity

Mathematics

Ethnicity	Year	PL 1	PL 2	PL 3	PL 4	PL 5	Total
African-American	1998	51.2	22.4	11.3	9.2	5.9	2255
	1999	40.1	25.5	14.2	12.3	7.9	2290
	2000	36.4	26.3	13.7	14.6	9.1	2064
	2001	33.7	24.1	15.8	15.2	11.2	2043
Hispanic	1998	57.7	24.1	9.1	6.5	2.7	914
	1999	30.6	21.7	19.0	17.4	11.2	258
	2000	38.6	26.8	16.9	11.6	6.0	1093
	2001	37.8	26.5	16.1	13.4	6.2	1458
Asian	1998	21.6	23.5	16.2	20.4	18.4	741
	1999	4.7	9.0	16.3	27.1	42.9	343
	2000	9.2	12.8	14.6	22.6	40.8	650
	2001	9.0	14.3	14.0	21.2	45.5	664
White	1988	14.8	12.4	10.0	24.1	38.8	291
	1999	8.2	10.4	14.2	22.1	45.1	317
	2000	8.0	4.2	11.2	20.6	55.9	286
	2001	5.4	8.0	8.0	22.5	56.2	276

The 1999 successive cohort group of Hispanic and Asian students made great improvements in mathematics. The Hispanic group reduced its 1998 baseline Level 1 percentage from 57.7 to 30.6 (27%) in 1999. The corresponding reduction by Asians was 17%. At the Advanced level, both groups showed substantial percentage increases. Hispanics and Asians made 9% and 25% increases in the proportion of students in Level 5 respectively. The PL data also reveal that fewer Hispanic and Asian students participated in the 1999 testing program (see Total column in Table 3) compared to 1998, 2000 and 2001 Hispanic and Asian students who took the tests. The high mathematics achievement registered by the two ethnic groups in 1999 may be attributed to the relatively small number of students who participated in the testing program. Both Hispanic and Asian students in the district have large numbers of English Language Learners. The participation of this subgroup of students in testing programs is affected by many factors including state and district policies as

well as parental consent. It is not unusual to find year-to-year fluctuations in the rate of English Language Learners who participate in the standardized testing program.

Summary

The school district performance data presented here provide benchmark or baseline data for comparing student achievement to subsequent years. The significance of such comparisons is that they enable school administrators to determine whether students in different performance levels are gaining, remaining steady or losing ground in successive years. PL data, unlike API, enable administrators to see the variability in achievement within the student population. The distribution of achievement allows managers and district planning teams to detect instructional impact or lack of impact on subgroups of students. The California Department of Education expects educational leaders, especially those of Underperforming Schools, to become familiar with and use SAT 9 results, API and other existing district /school data to make sound instructional decisions. We believe the performance level approach with successive cohorts of students is a simple and practical method for data analysis and presentation, which provides educational leaders with meaningful information for decision-making.

In summary, the use of PL data for program management is a much better administrative tool than those currently available to site administrators from SAT 9 results.

References

Adwere-Boamah, J. (2002, November). School Improvement: What Principals Really Need From API. Presentation at School Principals Meeting in Oakland, California.

California Department of Education. (1999). *The Public Schools Accountability Act and the 1999 Base Year Academic Performance Index*. Sacramento, CDE: Author.

Education Data Partnership. (2002,July). *Understanding the API.* http://www.ed-data.k12.ca.us

Linn, R.L., and Haug, C. (2002) Stability of School-Building Accountability Scores and Gains. *Educational Evaluation and Policy Analysis*, 24, 29-36.

Frisbie, D. A., Waltman, K, K., and Forsyth, R. A. (1999 March). Achievement Level Reporting With Norm-Referenced Tests. Presentation at the Annual Meeting of NCME. Chicago.