The Reliability of California's API[1]

Richard Hill
The Center for Assessment[2]

October 2, 2000
Revised February 16, 2001

Background

The Public Schools Accountability Act of 1999 directed Superintendent of Public Instruction to develop an Academic Performance Index (API) that would be used to measure the performance of California's public schools. Working in conjunction with its Technical Design Group (TDG), the Department of Education developed that index.

The performance of any school on any index will vary from year to year, in part because of systematic changes, but also because of random fluctuations. A school that scored higher than another in one year might score lower in another year, not because of any fundamental change in teaching practices or student population, but because the sample of students, and their performance on the given day that testing was done, varied.

Because of the number and importance of the decisions that will be based on schools' APIs, it is important to know, at a minimum, the reliability of the index. That is, if a school's API is likely to vary a substantial amount due to sampling fluctuations, it would be harder to justify, for example, awarding large amounts of reward money for score improvements than if it could be shown that sampling fluctuations played only a small part in school's scores.

For this reason, the Stuart Foundation provided a grant to The National Center for the Improvement of Educational Assessment, Inc. (The Center for Assessment) to calculate the amount of sampling error associated with the API. This paper is a report on the findings of that study.

Calculating the API

Each school's API is calculated using Stanford-9 test data collected from students at grades 2 through 11. Details on the procedure are available from the Department of Education. In short, however, each student's national percentile rank on each of several subscores is converted to an index ranging from 200 to 1000, and a weighted average of those scores becomes the school's API.

[1] This report was commissioned by the Stuart Foundation, 50 California Street, Suite 3350, San Francisco, CA 94111-4735, www.stuartfoundation.org.
[2] Richard Hill, Executive Director, The National Center for the Improvement of Educational Assessment, P.O. Box 4084, Portsmouth, NH 03802-4084, www.nciea.org.

Determining the Standard Error of APIs

The standard error of a school's score for a given year is the square root of the variance of students' scores within school divided by the number of students tested. The standard error of a school's API is an important statistic, since it can be used to determine the probability that a school's observed score is within a certain number of points of the school's true score (the school's average if an infinite number of students were tested). There is a probability of .68, for example, that a school's observed score will be within one standard error of its true score. Since the variance of observed student scores within a school includes measurement error, the major sources of error (measurement and sampling) are jointly accounted for by this approach. Two issues that were explored were the calculation of variance of students within school and whether pooling estimates across schools made sense.

The first issue is a concern for two reasons. The API is not calculated student-by-student. Instead, the scores for each content area are averaged across all students, and the API is the weighted average of those averages. If all students took all tests, there would be no problem, because the average of students would be the average of the tests. But sometimes there are missing data, and the results from the two approaches yield different results.

Ed Haertel, a member of the TDG and a professor at Stanford University, proposed a solution to this problem that worked. Using Dr. Haertel's approach, each school's API is identical whether the results are first averaged across tests (which is how the California Department of Education actually calculates the API) or first averaged across students (which provides us with the necessary statistics to calculate the variance of students within school). Yuan Li, who did the data analyses for this study, has written a summary of this approach, which is provided as Appendix A.

Once that problem was solved, it was straightforward to calculate the variance of students within school for all the schools in the state, and then compute what the average variance of students within school was for different categories of school. Once that was done, it was possible to see whether it made sense to pool the estimates of student variance across schools. This is an important step. Just as the mean for a school in a given year is dependent on the sample of students in that school that year (and therefore can vary from year to year), so is the variance. Often, if it can be shown that there are no systematic differences across schools, it is better to take the average of the variance across several schools and use that one value as the estimate of the variance in each school, rather than compute a separate estimate for each school.

As mentioned above, the first step was to see whether there were subgroups of schools for which the average variance of students within school was substantially different. It seemed as though there were two variables that might define subgroups with different variances—size and API. Size might matter; it would be possible, for example, that small schools were quite homogeneous, while there might be considerably more variance across students in larger schools. API also could matter; if there were floor or ceiling effects, schools with extreme APIs would have significantly smaller variances of students within school than school in the middle of the distribution. Other variables, such as race and School Characteristic Index (SCI—a measure of the socio-economic status of students in the school) are so highly correlated to API that any adjustment made on the basis of API would largely incorporate any differences due to these other variables.

Summaries of the results are provided in Tables 1 and 2. As the results show, the variance of student scores within school was not related to the size of the school; the standard deviations within school size group are much larger than any differences between the groups. However, the variance of students within school was related to the API ranking, and in systematic ways that were consistent across all three levels of schools. As a result, all subsequent analyses of standard errors used pooled estimates for variance of students within school, with different estimates being used for schools within each API rank. Those values are provided in Table 3: the only difference between Table 2 and Table 3 is that the results in Table 2 are the simple average across schools, but in Table 3 they are weighted by the number of students.

Table 1

Mean and Standard Deviation of Variance of Students Within School, Reported by School Type and Size

School Size (Each category contains ten percent of the schools, sorted by number of students tested)	Elementary		Middle		High	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
1	52377	12391	44663	14110	47258	13450
2	51460	12449	49194	11311	46528	11410
3	52530	12137	52160	9951	47035	11252
4	51504	11965	51597	9797	50160	9374
5	51643	13802	50812	10835	49319	8612
6	53113	12501	50974	10825	50504	7336
7	52788	12421	52417	10091	48692	7789
8	52843	12601	53843	10059	50581	8080
9	52195	11508	52023	10892	49177	8581
10	52201	11559	51370	10588	45915	8649

Table 2

Mean and Standard Deviation of Variance of Students Within School, Reported by School Type and API Decile Group

School Decile Group for API	Elementary		Middle		High	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
1	45508	6194	43634	6797	37387	8350
2	54842	5910	51896	5904	46504	6476
3	59917	5730	58105	6043	52052	6230
4	62248	5845	58304	9171	53724	5956
5	62301	5960	59586	5160	54177	5703
6	60951	6575	57759	5694	55666	7024
7	56897	6609	55010	8217	52676	7471
8	50183	5732	50082	6193	50578	5624
9	42295	6523	44417	5736	47063	6455
10	27702	8091	30700	8875	36250	10431

Table 3

Mean and Standard Deviation of Variance of Students Within School, Reported by School Type and API Decile Group, Weighted by Number of Students Tested

School Decile Group for API	Elementary		Middle		High	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
1	45249	5981	43922	5399	37148	5484
2	54879	5261	52808	4565	45274	5354
3	59844	5536	58976	6331	51963	5326
4	62473	5744	59604	4992	53685	4472
5	62873	5715	60609	5257	54786	4968
6	61446	6333	58497	5442	56567	6568
7	57127	6568	56303	6578	53674	4728
8	50283	5642	51005	6247	50807	5115
9	42487	6066	45166	5642	47293	5726
10	27989	7775	31664	8245	37559	8794

The values in the table above are the variances of students within school. The standard error of the mean for a school is found by dividing the appropriate variance by the number of students tested and then taking the square root. Tables 4-6 provide the average standard error of the mean, for various school types, sizes and API decile ranking.

Table 4

Average Standard Error of the Mean for Elementary Schools Reported by API Decile Group, For Schools of Varying Size

Number of Students Tested	API Decile Rank									
	1	2	3	4	5	6	7	8	9	10
100 – 200	17	19	20	20	21	21	20	18	16	14
200 – 400	12	14	14	14	15	15	14	13	12	10
400 – 800	9	10	11	11	11	11	11	10	9	8
800 – 1600	7	8	8	8	8	8	8	8	7	5

Table 5

Average Standard Error of the Mean for Middle Schools
Reported by API Decile Group,
For Schools of Varying Size

Number of Students Tested	API Decile Rank									
	1	2	3	4	5	6	7	8	9	10
100 – 200	18	18	20	n/a	21	20	23	18	18	18
200 – 400	12	13	13	14	14	14	13	14	13	13
400 – 800	9	9	10	10	10	10	10	9	9	9
800 – 1600	6	7	8	8	8	8	7	7	7	7
More than 1600	5	5	6	n/a	6	n/a	6	5	n/a	6

Table 6

Average Standard Error of the Mean for High Schools
Reported by API Decile Group,
For Schools of Varying Size

Number of Students Tested	API Decile Rank									
	1	2	3	4	5	6	7	8	9	10
100 – 200	24	23	23	25	20	22	22	20	20	15
200 – 400	11	14	14	13	13	14	14	13	12	11
400 – 800	8	8	9	9	10	10	9	9	9	8
800 – 1600	6	6	7	7	7	7	7	7	6	6
More than 1600	4	5	5	5	5	5	5	5	5	4

Determining the Probability that a School's Observed API Is within Its Correct API Decile Rank

Schools are classified according to their API decile group. While a school's observed score almost certainly is somewhat different from its true score, that difference is largely immaterial so long as the observed API is in the same API decile group as the true score. Therefore, one statistic of interest is the probability that a school's observed API is within same API decile as its true score.

Each school's observed API and the standard error of the API are known, as well as the cut-points that divide the scores of the state into 10 equal-sized groups. One can use this information to calculate the probability that a school's true score lies within a given decile by using the assumption of normality of errors and computing the area under the normal curve that lies between the cut-points for the given school. Tables 7-11 provide the probability that a school whose observed API lies within a particular decile rank has a true API in each possible API rank. The results are combined for all levels (elementary/middle/high) of schools.

For even the smallest size of school included in the state's accountability system (100-200 students), it is virtual certainty that a school's reported decile rank is within one of its true decile rank. This part of the reporting system is extremely reliable.

Table 7

Probability that a School's True API Is within a Particular API Decile, Given the School's Observed API, For Schools of 100-200 Tested Students

Observed API Decile Rank	True API Decile Rank									
	1	2	3	4	5	6	7	8	9	10
1	.89	.11	.00	.00	.00	.00	.00	.00	.00	.00
2	.18	.69	.14	.00	.00	.00	.00	.00	.00	.00
3	.00	.16	.63	.21	.01	.00	.00	.00	.00	.00
4	.00	.00	.18	.63	.18	.01	.00	.00	.00	.00
5	.00	.00	.01	.20	.60	.19	.01	.00	.00	.00
6	.00	.00	.00	.01	.20	.58	.20	.01	.00	.00
7	.00	.00	.00	.00	.01	.16	.63	.20	.01	.00
8	.00	.00	.00	.00	.00	.00	.15	.69	.15	.00
9	.00	.00	.00	.00	.00	.00	.00	.13	.76	.11
10	.00	.00	.00	.00	.00	.00	.00	.00	.09	.91

Table 8

Probability that a School's True API Is within a Particular API Decile, Given the School's Observed API, For Schools of 200-400 Tested Students

Observed API Decile Rank	True API Decile Rank									
	1	2	3	4	5	6	7	8	9	10
1	.91	.09	.00	.00	.00	.00	.00	.00	.00	.00
2	.11	.75	.14	.00	.00	.00	.00	.00	.00	.00
3	.00	.11	.75	.14	.00	.00	.00	.00	.00	.00
4	.00	.00	.13	.74	.14	.00	.00	.00	.00	.00
5	.00	.00	.00	.13	.72	.14	.00	.00	.00	.00
6	.00	.00	.00	.00	.13	.72	.15	.00	.00	.00
7	.00	.00	.00	.00	.00	.13	.75	.13	.00	.00
8	.00	.00	.00	.00	.00	.00	.11	.77	.11	.00
9	.00	.00	.00	.00	.00	.00	.00	.10	.83	.07
10	.00	.00	.00	.00	.00	.00	.00	.00	.05	.95

Table 9

Probability that a School's True API Is within a Particular API Decile, Given the School's Observed API, For Schools of 400-800 Tested Students

Observed API Decile Rank	True API Decile Rank									
	1	2	3	4	5	6	7	8	9	10
1	.93	.07	.00	.00	.00	.00	.00	.00	.00	.00
2	.08	.82	.10	.00	.00	.00	.00	.00	.00	.00
3	.00	.10	.81	.10	.00	.00	.00	.00	.00	.00
4	.00	.00	.10	.80	.11	.00	.00	.00	.00	.00
5	.00	.00	.00	.11	.79	.10	.00	.00	.00	.00
6	.00	.00	.00	.00	.11	.77	.12	.00	.00	.00
7	.00	.00	.00	.00	.00	.10	.81	.10	.00	.00
8	.00	.00	.00	.00	.00	.00	.09	.80	.11	.00
9	.00	.00	.00	.00	.00	.00	.00	.07	.86	.06
10	.00	.00	.00	.00	.00	.00	.00	.00	.04	.96

Table 10

Probability that a School's True API Is within a Particular API Decile, Given the School's Observed API, For Schools of 800-1600 Tested Students

Observed API Decile Rank	True API Decile Rank									
	1	2	3	4	5	6	7	8	9	10
1	.98	.02	.00	.00	.00	.00	.00	.00	.00	.00
2	.05	.89	.07	.00	.00	.00	.00	.00	.00	.00
3	.00	.08	.85	.07	.00	.00	.00	.00	.00	.00
4	.00	.00	.06	.85	.09	.00	.00	.00	.00	.00
5	.00	.00	.00	.09	.84	.07	.00	.00	.00	.00
6	.00	.00	.00	.00	.09	.80	.12	.00	.00	.00
7	.00	.00	.00	.00	.00	.10	.81	.09	.00	.00
8	.00	.00	.00	.00	.00	.00	.07	.86	.07	.00
9	.00	.00	.00	.00	.00	.00	.00	.05	.88	.08
10	.00	.00	.00	.00	.00	.00	.00	.00	.05	.95

Table 11

Probability that a School's True API Is within a Particular API Decile, Given the School's Observed API, For Schools with Over 1600 Tested Students

Observed API Decile Rank	True API Decile Rank									
	1	2	3	4	5	6	7	8	9	10
1	.95	.05	.00	.00	.00	.00	.00	.00	.00	.00
2	.03	.93	.04	.00	.00	.00	.00	.00	.00	.00
3	.00	.07	.91	.03	.00	.00	.00	.00	.00	.00
4	.00	.00	.07	.81	.12	.00	.00	.00	.00	.00
5	.00	.00	.00	.10	.84	.06	.00	.00	.00	.00
6	.00	.00	.00	.00	.11	.86	.03	.00	.00	.00
7	.00	.00	.00	.00	.00	.02	.94	.04	.00	.00
8	.00	.00	.00	.00	.00	.00	.09	.89	.02	.00
9	.00	.00	.00	.00	.00	.00	.00	.04	.96	.00
10	.00	.00	.00	.00	.00	.00	.00	.00	.08	.92

Determining the Probability that a School's Observed API Is within Its Correct Similar Schools Decile Rank

Schools also are classified according to how they are doing relative to similar schools. Each school receives a School Characteristics Index (SCI). Each school then is compared to the 100 schools in the state whose SCI is closest to the given school (50 on each side). Those schools are divided into 10 groups, and the given school is assigned a Similar Schools Rank depending on which decile its API falls into against this comparison group.

Given a school's API, the observed scores of its 100 similar schools and the standard deviation of students within schools, one can calculate the probability that a school's true score lies within a given Similar Schools decile, using a method similar to that used to calculate the tables above. However, one major difference between these calculations and the previous ones is that the observed scores of the 100 similar schools also have uncertainty associated with them. Upon resampling, each of the 100 similar schools' means might change, thereby possibly changing the cut points used to create the Similar Schools deciles. Therefore, this study was done using the following method:

1. The API for the given school and the APIs for its 100 similar schools were determined.
2. For each of the 100 similar schools, a possible API for that school was created by selecting a random normal deviate, multiplying that value by the standard error of the mean for the school, and adding the product to the observed API for the school. This simulated what the set of observed scores might be for the 100 similar schools if another sample of students had been tested for each of these schools.
3. This set of possible means was used to create cut scores for the Similar Schools Decile Ranks.
4. We then calculated the probability that the given school would have a true score within each of the bands created by those decile ranks by computing the area under the normal curve that lies between the cut-points for the given school.

5. This process was repeated 1,000 times for each school.

Thus, for each school, we had 1,000 sets of Similar School Decile Ranks for each school, and for each set, the probabilities that the true score of the given school would lie within the ranges defined by those decile ranks. We then calculated the average probability within each rank over the 1,000 sets. That provided the probability that any given school's true API would fall within each possible Similar Schools Decile Rank. Tables 12-16 provide the averages of those probabilities for schools of different size.

Not surprisingly, the probability that a school's observed score will fall within one of its true Similar School Rank is lower than the probability that it will fall within one of its true API rank. There is a high correlation between SCI and API, and therefore the distance between API scores within SCI deciles is considerably less than the width of the API deciles. As a result, even though the standard error of school mean API scores is the same between the two sets of analyses, the likelihood that a school's score will change is considerably higher in the second analysis.

As expected, the probability that a school's true Similar Schools Rank will be within one of its observed rank increases as schools become larger, and as schools' ranks become more extreme. For the smallest schools in the middle of the distribution, the probability that its true Similar Schools Rank will be within one of its observed rank is as low as .69; for the largest schools in the tails of the distribution, as high as 1.00. Given the consequences associated with the possible misplacement of a school, however, it is fair to say that the system is sufficiently reliable even in the former case, and quite reliable in the latter.

Table 12

Probability that a School's True API Is within a Particular Similar Schools Rank, Given the School's Observed Similar Schools, For Schools of 100-200 Tested Students

Observed Similar Schools Rank	True Similar Schools Rank									
	1	2	3	4	5	6	7	8	9	10
1	.82	.14	.03	.01	.00	.00	.00	.00	.00	.00
2	.27	.42	.21	.07	.02	.01	.00	.00	.00	.00
3	.06	.24	.32	.22	.11	.04	.01	.00	.00	.00
4	.02	.10	.23	.27	.21	.11	.05	.01	.00	.00
5	.00	.04	.11	.22	.26	.21	.11	.04	.01	.00
6	.00	.01	.04	.11	.22	.27	.21	.11	.03	.00
7	.00	.00	.02	.05	.11	.20	.28	.23	.10	.01
8	.00	.00	.00	.01	.04	.10	.21	.32	.27	.06
9	.00	.00	.00	.00	.01	.02	.07	.22	.45	.23
10	.00	.00	.00	.00	.00	.00	.00	.02	.15	.83

Table 13

Probability that a School's True API Is within a Particular Similar Schools Rank, Given the School's Observed Similar Schools, For Schools of 200-400 Tested Students

Observed Similar Schools Rank	True Similar Schools Rank									
	1	2	3	4	5	6	7	8	9	10
1	.81	.17	.02	.00	.00	.00	.00	.00	.00	.00
2	.20	.52	.23	.05	.01	.00	.00	.00	.00	.00
3	.02	.23	.42	.24	.07	.01	.00	.00	.00	.00
4	.00	.06	.24	.37	.24	.08	.02	.00	.00	.00
5	.00	.01	.07	.24	.35	.23	.08	.01	.00	.00
6	.00	.00	.01	.08	.24	.35	.24	.07	.01	.00
7	.00	.00	.00	.02	.08	.24	.38	.24	.05	.00
8	.00	.00	.00	.00	.01	.07	.24	.44	.22	.01
9	.00	.00	.00	.00	.00	.01	.04	.22	.55	.18
10	.00	.00	.00	.00	.00	.00	.00	.01	.15	.84

Table 14

Probability that a School's True API Is within a Particular Similar Schools Rank, Given the School's Observed Similar Schools, For Schools of 400-800 Tested Students

Observed Similar Schools Rank	True Similar Schools Rank									
	1	2	3	4	5	6	7	8	9	10
1	.81	.18	.01	.00	.00	.00	.00	.00	.00	.00
2	.13	.59	.24	.04	.00	.00	.00	.00	.00	.00
3	.01	.21	.48	.25	.05	.00	.00	.00	.00	.00
4	.00	.03	.23	.45	.24	.05	.00	.00	.00	.00
5	.00	.00	.05	.24	.42	.24	.05	.01	.00	.00
6	.00	.00	.00	.05	.24	.42	.24	.05	.00	.00
7	.00	.00	.00	.01	.06	.25	.44	.22	.02	.00
8	.00	.00	.00	.00	.01	.04	.23	.50	.22	.01
9	.00	.00	.00	.00	.00	.00	.02	.21	.61	.16
10	.00	.00	.00	.00	.00	.00	.00	.01	.15	.84

Table 15

**Probability that a School's True API Is within a Particular Similar Schools Rank,
Given the School's Observed Similar Schools,
For Schools of 800-1600 Tested Students**

Observed Similar Schools Rank	True Similar Schools Rank									
	1	2	3	4	5	6	7	8	9	10
1	.81	.19	.01	.00	.00	.00	.00	.00	.00	.00
2	.10	.63	.24	.03	.00	.00	.00	.00	.00	.00
3	.01	.20	.53	.22	.03	.00	.00	.00	.00	.00
4	.00	.02	.22	.48	.23	.04	.00	.00	.00	.00
5	.00	.00	.03	.23	.47	.23	.03	.00	.00	.00
6	.00	.00	.00	.03	.23	.47	.24	.03	.00	.00
7	.00	.00	.00	.00	.03	.23	.49	.22	.02	.00
8	.00	.00	.00	.00	.00	.04	.24	.56	.16	.00
9	.00	.00	.00	.00	.00	.00	.02	.21	.67	.10
10	.00	.00	.00	.00	.00	.00	.00	.00	.13	.87

Table 16

**Probability that a School's True API Is within a Particular Similar Schools Rank,
Given the School's Observed Similar Schools,
For Schools with over 1600 Tested Students**

Observed Similar Schools Rank	True Similar Schools Rank									
	1	2	3	4	5	6	7	8	9	10
1	.75	.24	.00	.00	.00	.00	.00	.00	.00	.00
2	.08	.73	.19	.00	.00	.00	.00	.00	.00	.00
3	.00	.16	.61	.21	.02	.00	.00	.00	.00	.00
4	.00	.00	.15	.61	.22	.02	.00	.00	.00	.00
5	.00	.00	.01	.21	.57	.20	.01	.00	.00	.00
6	.00	.00	.00	.01	.17	.57	.23	.02	.00	.00
7	.00	.00	.00	.00	.01	.25	.59	.14	.00	.00
8	.00	.00	.00	.00	.00	.02	.24	.61	.13	.00
9	.00	.00	.00	.00	.00	.00	.01	.16	.66	.17
10	.00	.00	.00	.00	.00	.00	.00	.00	.30	.69

The Probability that a School Will Meet Its Growth Target

All schools are assigned a Growth Target. The Growth Target is computed by subtracting the school's current API from 800 and dividing by 20. Several variables influence the probability that a school will meet its Growth Target, not the least of which is the growth the school actually makes. However, other important factors include the size of the Growth Target (which in itself a function of the API for the school), the size of the school and the percentage of students tested in one year who are tested in a second year.

Note that the probability that a school will meet its Growth Target is not the same as the probability that it will get a reward. To get a reward, a school's gain in API must not only equal or exceed its Growth Target, but also all "numerically significant ethnic and socioeconomically disadvantaged subgroups" must make at least 80 percent of the gain established by the Growth Target. Thus, a school might meet its Growth Target, but fail to receive a reward because the observed gain for one or more of its subgroups is insufficient. In this section of the paper, we look only at the probability that a school's observed growth, across all students, will equal or exceed its Growth Target. In the next section, we will provide the probabilities that a school will receive a reward.

The amount of true growth that a school makes in one year can never be determined. We can only calculate changes in observed scores. Those changes are influenced by true change, but also by the sampling error in each of the two years. If all the students tested in one year were not present the next year, the standard error of the difference would be the square root of the sum of the squared standard errors for each of the two years. When some students return, the samples for the two years are not independent, and the sampling error is reduced somewhat. For the purposes of this study, we assumed that the correlation, within school, of student performance across two years was equal to 0.7.

If a school's true change from one year to the next is small, most of the observed change will be the result of sampling error. If the true change is large relative to sampling error, however, then it is likely that a school's observed scores will increase as well.

Schools that have lower APIs have larger Growth Targets, so it seemed as though a good approach to this analysis would be to divide schools into groups based on their starting-year API, and then to posit different amounts of true growth. We grouped schools into four categories: API less than 400, 400-599, 600-799, and 800 or above. For each group except the last one, we posited five possible values of true growth: none, approximately half the Growth Target, equal to the Growth Target, approximately equal to 1.5 times the Growth Target, and twice the Growth Target.

Tables 17-20 provide the results. Each table provides the results for one of the four groups of schools, based on starting API. For a typical school in the first group—schools with an API below 400, and therefore a typical school in that group would have a starting API of 300—the average Growth Target was 25 points ({800 – 300} / 20 = 25). Therefore, the results in Table 17 provide the average probability that a school in that group would reach its Growth Target (i.e., have its observed API increase by as much as its Growth Target), if its true growth were 0, 12.5, 25, 37.5, or 50 points.

In all four tables, the results are reported depending on the percentage of students tested in the second year who also were present for testing in that school the previous year. The larger the percentage of students retested, the smaller the standard error of the gain scores, which means the greater likelihood that a school with small true growth will not reach its Growth Target and that one with large true growth will. The percentage of such students is dependent on the mobility rate for the school and the grade configuration. For example, students at grade 2, unless retained in grade, could not have been tested the previous year. So, a school containing grades K-5 would have a maximum of three-fourths of its students tested in consecutive years, and would reach that maximum only if all the first year's second, third and fourth graders stayed in the school through the end of the third, fourth and fifth grades, respectively.

Results also are reported by school size. As would be expected, the number of students tested in the school largely determines the standard error of a school's mean score. Although California reports results only for schools with at least 100 students tested, this analysis looked at all schools, regardless of size. The analyses shown in previous tables provided results that indicated highly reliable results, even for smaller schools. As a result, we extended this analysis to smaller schools.

It might be worthwhile at this point to provide a couple of examples from the tables to ensure that their meaning is clear. The first cell in Table 17 tells us that, for schools with an API less than 400, fewer than 50 students, and having 25 percent of its students return from one year to the next, 37.7 percent of them will meet their Growth Target even if they do not truly grow from one year to the next. On the other hand, for schools with an effective N of 100 to 200 and 75 percent of its students returning from one year to the next, there is a 96 percent chance that their observed scores will increase as much as their Growth Target if their true growth is 50 points.

The tables tell us that when schools' true growth is nearly equal to their Growth Target, they have a 50-50 chance of having their observed scores increase by as much as their Growth Target; that is not new or surprising information. It is interesting that mobility does not play much of a role; while having a greater percentage of students present for two consecutive years increases the probability of correct classification, it does not increase the probability by all that much. Also, the probability that a higher-scoring school will be correctly classified, given that its Growth Target is smaller, is smaller than it is for lower-scoring schools. As ever, however, size is the most important factor in determining the probability of accurate classification of a school.

As noted earlier, the previous analyses (the probability that a school's true API was within one of its reported Decile Rank, and the probability that its reported Similar Schools Decile Rank was correct within one) showed that results were quite reliable, even for the smallest schools included in the reporting system (those with 100 to 200 students). Consequently, schools with fewer than 100 students were included in these analyses. The results, however, show that the probability that a school will be incorrectly classified is quite high when it is small. In fact, some of the results show that the probability of misclassification is quite high even for schools of substantial size. Table 18, for example, provides results for schools with a starting API between 400 and 599. Such schools typically have a Growth Target of about 15 points (A school with a starting API of 500—the midpoint of this range—would have a Growth Target of exactly 15 points). If a school in this group has between 100 and 200 students and made 30 points worth of gain (twice its Growth Target and a very substantial amount of improvement for one year), there is better than a one in five chance that its observed API would improve less than its Growth Target. If the school has 50 to 100 students, the odds change to greater than one in four that its observed API would improve less than its Growth Target, and if the school is in the smallest category, one in three. Thus, a substantial portion of small schools would fail to meet their Growth Target even if they made twice the amount of real improvement expected of them.

Table 17

Average Probability of Meeting Growth Target, by the Percentage of Students Returning to School, the Number of Students Tested and the Amount of True Growth in API, for Schools with Starting API less than 400

School's Starting API less than 400		Amount of True Growth in API				
Percentage of Students Returning to School	Number of Students Tested	0	12.5	25	50	75
.25	Less than 50	0.377	0.440	0.505	0.631	0.737
	50-100^{-}	0.252	0.391	0.546	0.814	0.949
	100-200^{-}	0.158	0.337	0.566	0.909	0.994
	200-400^{-}	0.085	0.285	0.595	0.967	1.000
	400-800^{-}	0.034	0.223	0.626	0.993	1.000
	800-1600^{-}	0.007	0.152	0.678	0.999	1.000
	1600 or more	0.000	0.076	0.731	1.000	1.000
.50	Less than 50	0.363	0.432	0.505	0.646	0.759
	50-100^{-}	0.227	0.377	0.551	0.842	0.967
	100-200^{-}	0.129	0.318	0.575	0.934	0.997
	200-400^{-}	0.062	0.261	0.607	0.980	1.000
	400-800^{-}	0.020	0.196	0.641	0.997	1.000
	800-1600^{-}	0.003	0.124	0.698	1.000	1.000
	1600 or more	0.000	0.054	0.755	1.000	1.000
.75	Less than 50	0.341	0.421	0.506	0.668	0.790
	50-100^{-}	0.191	0.358	0.560	0.878	0.983
	100-200^{-}	0.093	0.290	0.587	0.960	0.999
	200-400^{-}	0.036	0.227	0.624	0.992	1.000
	400-800^{-}	0.009	0.159	0.663	0.999	1.000
	800-1600^{-}	0.001	0.090	0.727	1.000	1.000
	1600 or more	0.000	0.032	0.788	1.000	1.000

Table 18

Average Probability of Meeting Growth Target, by the Percentage of Students Returning to School, the Number of Students Tested and the Amount of True Growth in API, for Schools with Starting API between 400 and 599

School's Starting API between 400 and 599		API Amount of True Growth				
Percentage of Students Returning to School	Number of Students Tested	0	7.5	15	30	45
.25	Less than 50	0.407	0.455	0.504	0.601	0.690
	50-100$^-$	0.351	0.432	0.517	0.680	0.814
	100-200$^-$	0.282	0.392	0.511	0.739	0.893
	200-400$^-$	0.211	0.350	0.514	0.812	0.958
	400-800$^-$	0.137	0.291	0.505	0.875	0.988
	800-1600$^-$	0.068	0.238	0.534	0.956	0.999
	1600 or more	0.016	0.127	0.477	0.989	1.000
.50	Less than 50	0.396	0.450	0.505	0.613	0.712
	50-100$^-$	0.334	0.424	0.519	0.701	0.842
	100-200$^-$	0.259	0.379	0.513	0.764	0.919
	200-400$^-$	0.184	0.333	0.516	0.840	0.974
	400-800$^-$	0.111	0.270	0.505	0.901	0.994
	800-1600$^-$	0.049	0.214	0.538	0.971	1.000
	1600 or more	0.010	0.107	0.474	0.994	1.000
.75	Less than 50	0.379	0.441	0.506	0.632	0.742
	50-100$^-$	0.308	0.412	0.522	0.731	0.879
	100-200$^-$	0.226	0.359	0.515	0.799	0.948
	200-400$^-$	0.149	0.308	0.519	0.877	0.988
	400-800$^-$	0.080	0.240	0.506	0.932	0.998
	800-1600$^-$	0.030	0.183	0.544	0.985	1.000
	1600 or more	0.004	0.083	0.471	0.998	1.000

Table 19

Average Probability of Meeting Growth Target, by the Percentage of Students Returning to School, the Number of Students Tested and the Amount of True Growth in API, for Schools with Starting API between 600 and 799

School's Starting API between 600 and 799		API Amount of True Growth				
Percentage of Students Returning to School	Number of Students Tested	0	2.5	5	10	15
.25	Less than 50	0.469	0.486	0.502	0.535	0.567
	50-100^{-}	0.438	0.466	0.495	0.553	0.609
	100-200^{-}	0.416	0.457	0.498	0.579	0.657
	200-400^{-}	0.385	0.441	0.497	0.609	0.711
	400-800^{-}	0.344	0.416	0.491	0.638	0.765
	800-1600^{-}	0.275	0.371	0.475	0.679	0.837
	1600 or more	0.210	0.326	0.458	0.717	0.892
.50	Less than 50	0.465	0.484	0.502	0.539	0.576
	50-100^{-}	0.430	0.462	0.494	0.559	0.622
	100-200^{-}	0.406	0.452	0.498	0.589	0.675
	200-400^{-}	0.372	0.433	0.497	0.622	0.734
	400-800^{-}	0.327	0.406	0.490	0.653	0.790
	800-1600^{-}	0.253	0.356	0.472	0.697	0.862
	1600 or more	0.187	0.309	0.454	0.736	0.913
.75	Less than 50	0.460	0.481	0.503	0.546	0.588
	50-100^{-}	0.418	0.456	0.494	0.569	0.642
	100-200^{-}	0.391	0.444	0.497	0.603	0.701
	200-400^{-}	0.351	0.423	0.496	0.640	0.765
	400-800^{-}	0.302	0.392	0.488	0.675	0.823
	800-1600^{-}	0.222	0.336	0.467	0.722	0.894
	1600 or more	0.157	0.287	0.447	0.759	0.938

Table 20

Average Probability of Meeting Growth Target, by the Percentage of Students Returning to School, the Number of Students Tested and the Amount of True Growth in API, for Schools with Starting API 800 or Greater

School's Starting API 800 or Greater		API Amount of True Growth			
Percentages of Students Returning School	Number of Students Tested	0	1	5	10
.25	Less than 50	0.493	0.500	0.529	0.566
	50-100^{-}	0.485	0.500	0.561	0.634
	100-200^{-}	0.478	0.500	0.586	0.687
	200-400^{-}	0.469	0.500	0.622	0.757
	400-800^{-}	0.462	0.500	0.648	0.803
	800-1600^{-}	0.452	0.500	0.686	0.860
	1600 or more	0.437	0.500	0.736	0.920
.50	Less than 50	0.492	0.500	0.533	0.574
	50-100^{-}	0.483	0.500	0.568	0.651
	100-200^{-}	0.476	0.500	0.596	0.708
	200-400^{-}	0.465	0.500	0.637	0.784
	400-800^{-}	0.457	0.500	0.666	0.831
	800-1600^{-}	0.446	0.500	0.707	0.888
	1600 or more	0.429	0.500	0.761	0.943
.75	Less than 50	0.490	0.500	0.539	0.586
	50-100^{-}	0.480	0.500	0.580	0.675
	100-200^{-}	0.472	0.500	0.612	0.739
	200-400^{-}	0.459	0.500	0.659	0.820
	400-800^{-}	0.450	0.500	0.692	0.868
	800-1600^{-}	0.436	0.500	0.738	0.921
	1600 or more	0.417	0.500	0.797	0.967

The Probability that a School Will Get a Reward

As noted in the previous section, Senate Bill 1X provided for a Governor's Performance Award Program. Under that program, awards would be given to schools that "meet or exceed API performance targets" and "demonstrate comparable improvement in academic achievement by all numerically significant ethnic and socioeconomically disadvantaged subgroups within schools."

The design team carefully considered interpretation and implementation of this aspect of the legislation. If implemented poorly, it could seriously affect the reliability of the accountability system. If a school had no numerically significant subgroups, it would receive a reward if it met its Growth Target—the probability that it would get a reward would be equal to the values shown in Tables 17-20. But if another school had numerically significant subgroups, if might not receive a reward even though it had identical progress to the more homogeneous school. In order to minimize the impact of this legislatively-mandated requirement on the accountability system, the design team decided to establish minimum requirements for the size of these groups, and to require that each group gain only 80 percent of the school's Growth Target in order to be eligible for rewards.

The "ethnic or socioeconomically disadvantaged subgroups" are as follows:

- African American
- American Indian or Alaskan Native
- Asian or Asian-American
- Filipino or Filipino-American
- Hispanic or Latino
- Pacific Islander
- Economically disadvantaged (neither parent has a high school diploma OR students participates in the free or reduced price lunch program)

A subgroup is "numerically significant" if there are at least 100 students in the subgroup in the school who provide test scores, or if there are at least 30 such students and the subgroup comprises at least 15 percent of the students in the school who provide test scores.

For example, suppose a school has 200 students and two numerically significant subgroups: 100 African American students (assume that the remaining students are mostly white, with fewer than 30 students in any of the other ethnic subgroups) and 150 economically disadvantaged students (some of whom are African American and some who are not). Suppose further that the API for the school is 500. The school's Growth Target is 15 ({800 – 500} / 20). Eighty percent of the school's Growth Target is 12. Therefore, this school would receive a reward only if the API for the school increased by at least 15 points, the API for the African American students increased by at least 12 points, and the average for the economically disadvantaged students also increased by at least 12 points. This school receives a reward only if it meets all three criteria.

As noted in earlier sections of this report, the number of students in a subgroup plays a substantial role in determining the reliability of a score. A school with some relatively small subgroups might make a real gain with each of its subgroups, but have one or more of the subgroups show inadequate gain because of random error. As groups get smaller, the probability of correct placement goes down, and as the number of such subgroups increases, the probabilities can drop dramatically.

Therefore, it was important to test these scenarios against real data. If schools had few numerically significant subgroups of small size, or if those subgroups were likely to make at least 80 percent of their targeted gain when their true scores increased by the full amount of their targeted gain, then the probability that a school would receive a reward would almost equal the probability that it met its Growth Target. On the other hand, it was quite possible that the inclusion of this additional requirement would add significant error to the system.

This analysis incorporated many of the concepts of the previous ones, but was somewhat more complex. As was done for the study of whether a school would meet its Growth Target, several possible amounts of true growth were posited. In addition, however, these simulations had to be conducted for each of the subgroups within the school if they were numerically significant. Thus, the first step in the study was to place students into of 16 cells (seven ethnic groups plus “other”, crossed two socioeconomic groups). Then, the API and number tested were computed for each of the marginals. Next, we determined which subgroups were numerically significant.

To this point, we had determined the performance of the school and its subgroups for the base year. We then had to determine the probability that a school would receive a reward if the true gain for all the subgroups equaled the true gain posited for the entire school. Note that this assumption—that the true gain for all subgroups was the same as the true gain for the school as a whole—is just one of many possible assumptions that could have been made. We could have posited, for example, that the true gain for some of the subgroups was just 80 percent of the true gain for the school as a whole. Each assumption would have required a set of data analyses parallel to the ones done in this study; that is, each additional assumption would have required as much data analysis as the one assumption we actually pursued. To keep the scope of the study within reasonable bounds, we conducted analyses on just this one assumption; but clearly others could have been chosen and would have provided interesting information.

Once we had posited the amount of true gain for the school, we drew a random normal number for each of 16 cells, and, knowing that random number, the observed API for the cells, the number of students tested, the variance of students within school for that cell and the posited gain, we calculated an observed score for that cell for the second year of testing. Note that the variance of students within cell within school was a pooled estimate; a different estimate was established for each of 480 cells (three school types {elementary, middle, high} by ten deciles by eight ethnic subgroups by two socioeconomic levels) by calculating the variance of students within school for each of the 16 student-level cells for schools in each of the 30 school-level cells.

Once we had a possible second-year score for each cell, we could calculate a second-year score for each subgroup by calculating the API of the margins. Note that the random variable chosen for each of the cells was independent, which meant that the gains made for each of the ethnic groups were independent of each other, but not independent of the gain made by the low socioeconomic group. Knowing the observed change in API for each subgroup, we could determine whether that school would be eligible to receive a reward.

This process of selecting 16 random normal values for each cell in a school and then determining whether the school would receive a reward was done 1,000 times for each school under each of the conditions posited (three possible percentages of students returning by four or five possible true score gains). Tables 21-24 provide the results.

Table 21

The Average Probability that a School Will Be Eligible to Receive a Reward, Reported by the Percentage of Students Returning School, the Number of Students Tested and the API Amount of True Growth, For Schools with a Starting API Less Than 400 (Replications = 1000)

School's Starting API less than 400		Amount of True Growth in API				
Percentage of Students Returning to School	Number of Students Tested	0	12.5	25	50	75
.25	Less than 50	0.358	0.422	0.477	0.632	0.743
	50-100⁻	0.151	0.239	0.349	0.563	0.706
	100-200⁻	0.104	0.240	0.432	0.798	0.949
	200-400⁻	0.054	0.198	0.454	0.874	0.981
	400-800⁻	0.022	0.170	0.525	0.957	0.996
	800-1600⁻	0.004	0.106	0.551	0.972	0.998
	1600 or more	0.000	0.058	0.626	0.991	1.000
.50	Less than 50	0.353	0.418	0.492	0.640	0.749
	50-100⁻	0.141	0.232	0.342	0.585	0.728
	100-200⁻	0.093	0.232	0.442	0.828	0.967
	200-400⁻	0.038	0.183	0.467	0.907	0.988
	400-800⁻	0.014	0.155	0.547	0.971	0.998
	800-1600⁻	0.002	0.087	0.579	0.982	0.999
	1600 or more	0.000	0.040	0.650	0.996	1.000
.75	Less than 50	0.312	0.406	0.499	0.656	0.781
	50-100⁻	0.116	0.216	0.353	0.611	0.743
	100-200⁻	0.061	0.205	0.462	0.880	0.982
	200-400⁻	0.023	0.154	0.487	0.940	0.995
	400-800⁻	0.006	0.122	0.575	0.983	0.999
	800-1600⁻	0.000	0.058	0.616	0.989	1.000
	1600 or more	0.000	0.023	0.680	0.999	1.000

Table 22

The Average Probability that a School Will Be Eligible to Receive a Reward, Reported by the Percentage of Students Returning School, the Number of Students Tested and the API Amount of True Growth, For Schools with a Starting API between 400 and 599 (Replications = 1000)

School's Starting API between 400 and 599		Amount of True Growth in API				
Percentage of Students Returning to School	Number of Students Tested	0	7.5	15	30	45
.25	Less than 50	0.383	0.428	0.479	0.576	0.668
	50-100⁻	0.276	0.347	0.428	0.583	0.728
	100-200⁻	0.145	0.217	0.309	0.510	0.703
	200-400⁻	0.098	0.182	0.300	0.578	0.795
	400-800⁻	0.057	0.146	0.295	0.662	0.877
	800-1600⁻	0.016	0.079	0.246	0.704	0.911
	1600 or more	0.002	0.034	0.230	0.785	0.930
.50	Less than 50	0.374	0.423	0.478	0.587	0.691
	50-100⁻	0.265	0.341	0.427	0.607	0.764
	100-200⁻	0.133	0.210	0.310	0.540	0.745
	200-400⁻	0.084	0.173	0.304	0.616	0.834
	400-800⁻	0.045	0.134	0.300	0.705	0.908
	800-1600⁻	0.011	0.071	0.253	0.754	0.933
	1600 or more	0.001	0.027	0.231	0.819	0.950
.75	Less than 50	0.354	0.417	0.475	0.607	0.719
	50-100⁻	0.238	0.331	0.434	0.639	0.808
	100-200⁻	0.114	0.198	0.316	0.584	0.802
	200-400⁻	0.065	0.159	0.311	0.669	0.882
	400-800⁻	0.031	0.118	0.306	0.762	0.941
	800-1600⁻	0.006	0.059	0.267	0.813	0.956
	1600 or more	0.000	0.018	0.234	0.863	0.967

Table 23

The Average Probability that a School Will Be Eligible to Receive a Reward, Reported by the Percentage of Students Returning School, the Number of Students Tested and the API Amount of True Growth, For Schools with a Starting API between 600 and 799 (Replications = 1000)

School's Starting between 600 and 799		Amount of True Growth in API				
Percentages of Students Returning to School	Number of Students Tested	0	2.5	5	10	15
.25	Less than 50	0.451	0.466	0.483	0.519	0.551
	50-100$^-$	0.348	0.373	0.404	0.460	0.515
	100-200$^-$	0.227	0.256	0.287	0.350	0.418
	200-400$^-$	0.173	0.207	0.243	0.323	0.408
	400-800$^-$	0.149	0.191	0.239	0.347	0.459
	800-1600$^-$	0.081	0.121	0.171	0.295	0.436
	1600 or more	0.033	0.061	0.101	0.223	0.382
.50	Less than 50	0.444	0.465	0.483	0.522	0.558
	50-100$^-$	0.338	0.368	0.403	0.466	0.527
	100-200$^-$	0.221	0.254	0.287	0.358	0.435
	200-400$^-$	0.166	0.203	0.244	0.335	0.430
	400-800$^-$	0.141	0.188	0.240	0.362	0.489
	800-1600$^-$	0.075	0.117	0.171	0.312	0.471
	1600 or more	0.029	0.057	0.104	0.243	0.427
.75	Less than 50	0.439	0.461	0.482	0.524	0.571
	50-100$^-$	0.328	0.364	0.400	0.477	0.549
	100-200$^-$	0.211	0.247	0.287	0.372	0.459
	200-400$^-$	0.156	0.198	0.245	0.352	0.464
	400-800$^-$	0.130	0.181	0.243	0.385	0.531
	800-1600$^-$	0.065	0.111	0.175	0.340	0.523
	1600 or more	0.024	0.054	0.106	0.274	0.490

Table 24

The Average Probability that a School Will Be Eligible to Receive a Reward, Reported by the Percentage of Students Returning School, the Number of Students Tested and the API Amount of True Growth, For Schools with a Starting API Greater than 800 (Replications = 1000)

School's Starting API Greater than 800		Amount of True Growth in API			
Percentage of Students Returning to School	Number of Students Tested	0	1	5	10
.25	Less than 50	0.471	0.488	0.510	0.553
	50-100^{-}	0.419	0.437	0.496	0.571
	100-200^{-}	0.344	0.356	0.439	0.537
	200-400^{-}	0.296	0.322	0.427	0.559
	400-800^{-}	0.300	0.333	0.475	0.640
	800-1600^{-}	0.206	0.239	0.403	0.595
	1600 or more	0.154	0.187	0.356	0.551
.50	Less than 50	0.474	0.484	0.510	0.556
	50-100^{-}	0.421	0.432	0.502	0.598
	100-200^{-}	0.341	0.361	0.446	0.558
	200-400^{-}	0.293	0.321	0.440	0.589
	400-800^{-}	0.296	0.334	0.492	0.674
	800-1600^{-}	0.203	0.239	0.421	0.634
	1600 or more	0.138	0.192	0.365	0.590
.75	Less than 50	0.470	0.482	0.520	0.566
	50-100^{-}	0.413	0.438	0.515	0.619
	100-200^{-}	0.333	0.362	0.463	0.592
	200-400^{-}	0.287	0.321	0.461	0.629
	400-800^{-}	0.290	0.333	0.519	0.722
	800-1600^{-}	0.194	0.240	0.451	0.682
	1600 or more	0.136	0.186	0.404	0.630

The results in Tables 21 through 24 are interesting in their own right, but of much more interest when compared to Tables 17-20—that is, seeing the impact of the additional requirement that means for all numerically significant subgroups also increases by at least 80 percent of the Growth Target for the school as a whole. Tables 17-20 provide the probability that a school will meet its Growth Target; Tables 21-24 provide the probability that a school will meet its Growth Target AND have all numerically significant subgroups grow at least 80 percent of the school's Growth Target.

Table 25 summarizes the results for the probabilities of success when a school's true growth is approximately twice its Growth Target (50 points for the school in the lowest API group, 30 points for the next group of schools, 10 points for the next, and 5 points for the highest) and half the students present for testing one year return for the next. Note that Table 25 contains no new information; it simply extracts information for Tables 17-24.

Table 25

Average Probabilities for Schools Whose True Gain Is Approximately Twice Their Growth Target, Summarized from Tables 17-24

Starting API (and Amount of True Gain)	Number of Students Tested	Probability of Meeting Growth Target	Probability of Being Eligible for Reward
Less than 400 (50 points)	Less than 50	0.646	0.640
	50-100^{-}	0.842	0.585
	100-200^{-}	0.934	0.828
	200-400^{-}	0.980	0.907
	400-800^{-}	0.997	0.971
	800-1600^{-}	1.000	0.982
	1600 or more	1.000	0.996
400-599 (30 points)	Less than 50	0.613	0.587
	50-100^{-}	0.701	0.607
	100-200^{-}	0.764	0.540
	200-400^{-}	0.840	0.616
	400-800^{-}	0.901	0.705
	800-1600^{-}	0.971	0.754
	1600 or more	0.994	0.819
600-799 (10 points)	Less than 50	0.539	0.522
	50-100^{-}	0.559	0.466
	100-200^{-}	0.589	0.358
	200-400^{-}	0.622	0.335
	400-800^{-}	0.653	0.362
	800-1600^{-}	0.697	0.312
	1600 or more	0.736	0.243
800 or more (5 points)	Less than 50	0.533	0.510
	50-100^{-}	0.568	0.502
	100-200^{-}	0.596	0.446
	200-400^{-}	0.637	0.440
	400-800^{-}	0.666	0.492
	800-1600^{-}	0.707	0.421
	1600 or more	0.761	0.365

The interesting part of Table 25 is not just the fact that the subgroup requirement lowers the probability that a school will get a reward; that was obvious without doing any data analysis. What is interesting is the differential impact that it has on larger schools. Remember that Table 25 refers to schools whose true growth far exceeds their Growth Target. Thus, as schools get larger, the probability that they will meet their Growth Target increases; greater numbers of students tested increases the likelihood that a school will be correctly classified. But for schools with starting API of 600 or above, the likelihood that they will receive a reward actually *decreases* as they get larger. This seemingly incongruous result happens because larger schools are more likely to have numerically significant subgroups (and some subgroups that just barely make it over the limit of 30 students), which means that their eligibility for reward no longer is based on large numbers of students, but on several small groups, each of which may or may not have an observed gain equal to at least 80 percent of the school's Growth Target.

One obvious question is what would happen if we changed the rule requiring that each numerically significant subgroup make a gain equal to at least 80 percent of the school's Growth Target. Tables 26-29 and Tables 30-33 are comparable to Tables 21-24, but look at two possible different rules for subgroup growth. In Tables 26-29, schools would be eligible for a reward only if every numerically significant subgroup made observed gains equal to 100 percent of the school's Growth Target; for Tables 30-33, the standard was 60 percent.

Table 26

The Average Probability that a School Would Be Eligible to Receive a Reward, If Requirement Were that All Numerically Significant Subgroups Gain 100 Percent of a School's Growth Target, Reported by the Percentage of Students Returning School, the Number of Students Tested and the API Amount of True Growth, For Schools with a Starting API Less Than 400 (Replications = 1000)

School's Starting API less than 400		Amount of True Growth in API				
Percentage of Students Returning to School	Number of Students Tested	0	12.5	25	50	75
.25	Less than 50	0.352	0.421	0.485	0.623	0.738
	50-100$^-$	0.146	0.227	0.333	0.547	0.705
	100-200$^-$	0.091	0.209	0.393	0.771	0.938
	200-400$^-$	0.042	0.162	0.397	0.846	0.972
	400-800$^-$	0.016	0.133	0.459	0.939	0.994
	800-1600$^-$	0.003	0.074	0.464	0.958	0.997
	1600 or more	0.000	0.037	0.492	0.985	1.000
.50	Less than 50	0.342	0.410	0.474	0.630	0.762
	50-100$^-$	0.130	0.221	0.334	0.564	0.721
	100-200$^-$	0.068	0.194	0.400	0.794	0.959
	200-400$^-$	0.029	0.146	0.410	0.877	0.984
	400-800$^-$	0.009	0.114	0.474	0.954	0.997
	800-1600$^-$	0.001	0.061	0.478	0.970	0.998
	1600 or more	0.000	0.028	0.527	0.988	1.000
.75	Less than 50	0.329	0.401	0.477	0.655	0.791
	50-100$^-$	0.102	0.204	0.338	0.586	0.739
	100-200$^-$	0.050	0.178	0.413	0.851	0.974
	200-400$^-$	0.016	0.124	0.423	0.915	0.992
	400-800$^-$	0.004	0.092	0.492	0.973	0.999
	800-1600$^-$	0.000	0.041	0.514	0.983	1.000
	1600 or more	0.000	0.015	0.561	0.995	1.000

Table 27

The Average Probability that a School Would Be Eligible to Receive a Reward, If Requirement Were that All Numerically Significant Subgroups Gain 100 Percent of a School's Growth Target, Reported by the Percentage of Students Returning School, the Number of Students Tested and the API Amount of True Growth, For Schools with a Starting API between 400 and 599 (Replications = 1000)

School's Starting API between 400 and 599		Amount of True Growth in API				
Percentage of Students Returning to School	Number of Students Tested	0	7.5	15	30	45
.25	Less than 50	0.381	0.426	0.476	0.569	0.665
	50-100-	0.269	0.337	0.410	0.571	0.716
	100-200-	0.132	0.199	0.283	0.483	0.674
	200-400-	0.083	0.158	0.266	0.537	0.765
	400-800-	0.045	0.119	0.252	0.611	0.853
	800-1600-	0.011	0.057	0.190	0.641	0.888
	1600 or more	0.001	0.021	0.166	0.730	0.914
.50	Less than 50	0.369	0.421	0.477	0.585	0.689
	50-100-	0.251	0.329	0.413	0.592	0.747
	100-200-	0.117	0.189	0.283	0.508	0.716
	200-400-	0.069	0.148	0.267	0.572	0.808
	400-800-	0.034	0.106	0.251	0.652	0.887
	800-1600-	0.007	0.049	0.190	0.690	0.917
	1600 or more	0.000	0.014	0.161	0.770	0.935
.75	Less than 50	0.353	0.414	0.476	0.602	0.718
	50-100-	0.229	0.314	0.415	0.621	0.794
	100-200-	0.098	0.177	0.282	0.547	0.772
	200-400-	0.052	0.132	0.268	0.622	0.858
	400-800-	0.022	0.090	0.251	0.708	0.924
	800-1600-	0.004	0.038	0.196	0.753	0.945
	1600 or more	0.000	0.009	0.158	0.817	0.959

Table 28

The Average Probability that a School Would Be Eligible to Receive a Reward, If Requirement Were that All Numerically Significant Subgroups Gain 100 Percent of a School's Growth Target, Reported by the Percentage of Students Returning School, the Number of Students Tested and the API Amount of True Growth, For Schools with a Starting API between 600 and 799 (Replications = 1000)

School's Starting between 600 and 799		Amount of True Growth in API				
Percentages of Students Returning to School	Number of Students Tested	0	2.5	5	10	15
.25	Less than 50	0.448	0.466	0.482	0.516	0.554
	50-100^{-}	0.345	0.372	0.396	0.456	0.511
	100-200^{-}	0.220	0.249	0.278	0.340	0.406
	200-400^{-}	0.164	0.197	0.232	0.310	0.394
	400-800^{-}	0.140	0.181	0.225	0.328	0.438
	800-1600^{-}	0.073	0.110	0.155	0.269	0.407
	1600 or more	0.027	0.051	0.086	0.195	0.347
.50	Less than 50	0.447	0.460	0.480	0.519	0.561
	50-100^{-}	0.336	0.366	0.393	0.459	0.523
	100-200^{-}	0.214	0.245	0.277	0.349	0.424
	200-400^{-}	0.157	0.192	0.232	0.320	0.415
	400-800^{-}	0.131	0.175	0.225	0.340	0.465
	800-1600^{-}	0.067	0.105	0.155	0.285	0.438
	1600 or more	0.024	0.047	0.088	0.209	0.379
.75	Less than 50	0.439	0.458	0.480	0.524	0.570
	50-100^{-}	0.322	0.359	0.392	0.468	0.541
	100-200^{-}	0.204	0.238	0.276	0.359	0.446
	200-400^{-}	0.146	0.187	0.233	0.335	0.445
	400-800^{-}	0.120	0.168	0.226	0.359	0.503
	800-1600^{-}	0.058	0.099	0.156	0.307	0.482
	1600 or more	0.020	0.045	0.090	0.234	0.436

Table 29

The Average Probability that a School Would Be Eligible to Receive a Reward, If Requirement Were that All Numerically Significant Subgroups Gain 100 Percent of a School's Growth Target, Reported by the Percentage of Students Returning School, the Number of Students Tested and the API Amount of True Growth, For Schools with a Starting API Greater than 800 (Replications = 1000)

School's Starting API Greater than 800		Amount of True Growth in API			
Percentage of Students Returning to School	Number of Students Tested	0	1	5	10
.25	Less than 50	0.476	0.478	0.512	0.542
	50-100^{-}	0.421	0.434	0.498	0.574
	100-200^{-}	0.340	0.359	0.437	0.536
	200-400^{-}	0.293	0.318	0.424	0.556
	400-800^{-}	0.298	0.329	0.470	0.638
	800-1600^{-}	0.199	0.232	0.393	0.590
	1600 or more	0.147	0.182	0.344	0.548
.50	Less than 50	0.468	0.482	0.517	0.556
	50-100^{-}	0.417	0.435	0.502	0.594
	100-200^{-}	0.335	0.359	0.446	0.556
	200-400^{-}	0.290	0.318	0.434	0.586
	400-800^{-}	0.291	0.329	0.488	0.669
	800-1600^{-}	0.195	0.234	0.412	0.623
	1600 or more	0.141	0.172	0.360	0.574
.75	Less than 50	0.474	0.478	0.517	0.571
	50-100^{-}	0.422	0.434	0.515	0.616
	100-200^{-}	0.335	0.357	0.458	0.588
	200-400^{-}	0.284	0.317	0.456	0.625
	400-800^{-}	0.285	0.328	0.512	0.718
	800-1600^{-}	0.190	0.233	0.441	0.678
	1600 or more	0.136	0.180	0.394	0.612

Table 30

The Average Probability that a School Would Be Eligible to Receive a Reward, If Requirement Were that All Numerically Significant Subgroups Gain 60 Percent of a School's Growth Target, Reported by the Percentage of Students Returning School, the Number of Students Tested and the API Amount of True Growth, For Schools with a Starting API Less Than 400 (Replications = 1000)

School's Starting API less than 400		Amount of True Growth in API				
Percentage of Students Returning to School	Number of Students Tested	0	12.5	25	50	75
.25	Less than 50	0.367	0.428	0.493	0.621	0.749
	50-100^{-}	0.164	0.256	0.355	0.572	0.719
	100-200^{-}	0.117	0.264	0.461	0.814	0.959
	200-400^{-}	0.061	0.215	0.490	0.901	0.984
	400-800^{-}	0.026	0.187	0.561	0.968	0.997
	800-1600^{-}	0.004	0.120	0.597	0.982	0.999
	1600 or more	0.000	0.058	0.655	0.996	1.000
.50	Less than 50	0.355	0.418	0.497	0.641	0.761
	50-100^{-}	0.153	0.242	0.360	0.586	0.733
	100-200^{-}	0.098	0.248	0.467	0.855	0.973
	200-400^{-}	0.043	0.201	0.505	0.928	0.992
	400-800^{-}	0.015	0.163	0.582	0.980	0.999
	800-1600^{-}	0.002	0.099	0.619	0.988	0.999
	1600 or more	0.000	0.045	0.690	0.997	1.000
.75	Less than 50	0.331	0.400	0.499	0.662	0.790
	50-100^{-}	0.124	0.235	0.369	0.626	0.748
	100-200^{-}	0.068	0.223	0.494	0.900	0.988
	200-400^{-}	0.024	0.178	0.526	0.956	0.997
	400-800^{-}	0.007	0.133	0.611	0.990	0.999
	800-1600^{-}	0.001	0.070	0.660	0.993	1.000
	1600 or more	0.000	0.024	0.724	0.999	1.000

Table 31

The Average Probability that a School Would Be Eligible to Receive a Reward, If Requirement Were that All Numerically Significant Subgroups Gain 60 Percent of a School's Growth Target, Reported by the Percentage of Students Returning School, the Number of Students Tested and the API Amount of True Growth, For Schools with a Starting API between 400 and 599 (Replications = 1000)

School's Starting API between 400 and 599		Amount of True Growth in API				
Percentage of Students Returning to School	Number of Students Tested	0	7.5	15	30	45
.25	Less than 50	0.385	0.433	0.479	0.578	0.667
	50-100$^-$	0.284	0.356	0.434	0.594	0.739
	100-200$^-$	0.158	0.238	0.331	0.539	0.725
	200-400$^-$	0.110	0.203	0.329	0.610	0.817
	400-800$^-$	0.067	0.168	0.330	0.701	0.897
	800-1600$^-$	0.022	0.102	0.293	0.754	0.927
	1600 or more	0.003	0.046	0.269	0.822	0.945
.50	Less than 50	0.374	0.427	0.482	0.589	0.690
	50-100$^-$	0.269	0.352	0.437	0.619	0.772
	100-200$^-$	0.147	0.228	0.335	0.566	0.769
	200-400$^-$	0.096	0.194	0.336	0.651	0.857
	400-800$^-$	0.054	0.156	0.337	0.745	0.925
	800-1600$^-$	0.015	0.091	0.303	0.801	0.946
	1600 or more	0.002	0.037	0.274	0.858	0.959
.75	Less than 50	0.356	0.414	0.479	0.602	0.722
	50-100$^-$	0.249	0.340	0.445	0.653	0.824
	100-200$^-$	0.126	0.221	0.343	0.618	0.823
	200-400$^-$	0.076	0.182	0.346	0.708	0.902
	400-800$^-$	0.038	0.139	0.347	0.801	0.953
	800-1600$^-$	0.009	0.078	0.324	0.857	0.964
	1600 or more	0.001	0.027	0.282	0.897	0.974

Table 32

The Average Probability that a School Would Be Eligible to Receive a Reward, If Requirement Were that All Numerically Significant Subgroups Gain 60 Percent of a School's Growth Target, Reported by the Percentage of Students Returning School, the Number of Students Tested and the API Amount of True Growth, For Schools with a Starting API between 600 and 799 (Replications = 1000)

School's Starting between 600 and 799		Amount of True Growth in API				
Percentages of Students Returning to School	Number of Students Tested	0	2.5	5	10	15
.25	Less than 50	0.450	0.466	0.483	0.518	0.551
	50-100^{-}	0.354	0.379	0.409	0.462	0.519
	100-200^{-}	0.235	0.264	0.294	0.358	0.427
	200-400^{-}	0.181	0.217	0.253	0.336	0.423
	400-800^{-}	0.158	0.202	0.253	0.364	0.480
	800-1600^{-}	0.090	0.134	0.187	0.320	0.465
	1600 or more	0.038	0.071	0.117	0.252	0.422
.50	Less than 50	0.444	0.464	0.487	0.522	0.561
	50-100^{-}	0.342	0.376	0.408	0.471	0.536
	100-200^{-}	0.228	0.262	0.298	0.369	0.445
	200-400^{-}	0.175	0.213	0.256	0.349	0.447
	400-800^{-}	0.151	0.200	0.256	0.382	0.511
	800-1600^{-}	0.083	0.129	0.191	0.341	0.504
	1600 or more	0.034	0.069	0.120	0.277	0.468
.75	Less than 50	0.440	0.461	0.482	0.529	0.572
	50-100^{-}	0.334	0.368	0.405	0.482	0.556
	100-200^{-}	0.219	0.256	0.296	0.383	0.473
	200-400^{-}	0.165	0.209	0.259	0.368	0.484
	400-800^{-}	0.140	0.195	0.259	0.408	0.557
	800-1600^{-}	0.074	0.126	0.195	0.372	0.561
	1600 or more	0.030	0.064	0.128	0.313	0.540

Table 33

The Average Probability that a School Would Be Eligible to Receive a Reward, If Requirement Were that All Numerically Significant Subgroups Gain 60 Percent of a School's Growth Target, Reported by the Percentage of Students Returning School, the Number of Students Tested and the API Amount of True Growth, For Schools with a Starting API Greater than 800 (Replications = 1000)

School's Starting API Greater than 800		Amount of True Growth in API			
Percentage of Students Returning to School	Number of Students Tested	0	1	5	10
.25	Less than 50	0.471	0.477	0.510	0.546
	50-100⁻	0.427	0.441	0.497	0.579
	100-200⁻	0.347	0.362	0.440	0.541
	200-400⁻	0.300	0.324	0.429	0.563
	400-800⁻	0.304	0.338	0.478	0.646
	800-1600⁻	0.206	0.243	0.404	0.601
	1600 or more	0.152	0.192	0.349	0.550
.50	Less than 50	0.472	0.481	0.517	0.553
	50-100⁻	0.418	0.434	0.506	0.593
	100-200⁻	0.342	0.364	0.451	0.558
	200-400⁻	0.297	0.325	0.445	0.590
	400-800⁻	0.300	0.336	0.496	0.680
	800-1600⁻	0.206	0.247	0.427	0.643
	1600 or more	0.152	0.191	0.389	0.594
.75	Less than 50	0.473	0.486	0.517	0.571
	50-100⁻	0.421	0.433	0.521	0.618
	100-200⁻	0.340	0.360	0.467	0.592
	200-400⁻	0.291	0.324	0.467	0.632
	400-800⁻	0.294	0.338	0.524	0.726
	800-1600⁻	0.199	0.245	0.455	0.687
	1600 or more	0.148	0.199	0.415	0.642

Table 34 is similar to Table 25, but adds two columns from the data in Tables 26-33, so that the reader can get some sense of the impact of changing the requirement for the amount of gain that numerically significant subgroups must show.

Table 34

Average Probabilities for Schools Whose True Gain Is Approximately Twice Their Growth Target, Summarized from Tables 17-33

Starting API (and Amount of True Gain)	Number of Students Tested	Probability of Meeting Growth Target	Probability of Being Eligible for Reward if 80 Percent	Probability of Being Eligible for Reward if 100 Percent	Probability of Being Eligible for Reward if 60 Percent
Less than 400 (50 points)	Less than 50	0.646	0.640	0.630	0.641
	50-100^{-}	0.842	0.585	0.564	0.586
	100-200^{-}	0.934	0.828	0.794	0.855
	200-400^{-}	0.980	0.907	0.877	0.928
	400-800^{-}	0.997	0.971	0.954	0.980
	800-1600^{-}	1.000	0.982	0.970	0.988
	1600 or more	1.000	0.996	0.988	0.997
400-599 (30 points)	Less than 50	0.613	0.587	0.585	0.589
	50-100^{-}	0.701	0.607	0.592	0.619
	100-200^{-}	0.764	0.540	0.508	0.566
	200-400^{-}	0.840	0.616	0.572	0.651
	400-800^{-}	0.901	0.705	0.652	0.745
	800-1600^{-}	0.971	0.754	0.690	0.801
	1600 or more	0.994	0.819	0.770	0.858
600-799 (10 points)	Less than 50	0.539	0.522	0.519	0.522
	50-100^{-}	0.559	0.466	0.459	0.471
	100-200^{-}	0.589	0.358	0.349	0.369
	200-400^{-}	0.622	0.335	0.320	0.349
	400-800^{-}	0.653	0.362	0.340	0.382
	800-1600^{-}	0.697	0.312	0.285	0.341
	1600 or more	0.736	0.243	0.209	0.277
800 or more (5 points)	Less than 50	0.533	0.510	0.517	0.517
	50-100^{-}	0.568	0.502	0.502	0.506
	100-200^{-}	0.596	0.446	0.446	0.451
	200-400^{-}	0.637	0.440	0.434	0.445
	400-800^{-}	0.666	0.492	0.488	0.496
	800-1600^{-}	0.707	0.421	0.412	0.427
	1600 or more	0.761	0.365	0.360	0.389

Obviously, requiring that gains for each numerically significant subgroup be only 60 percent of the Growth Target increases the probability that a school will be eligible for a reward, and increasing the requirement to 100 percent decreases the probability. What is interesting in the table (which admittedly is a selected subset of all the data available) is that the impact of changing the requirement is rather minimal—at least when compared to not having any such requirement at all. Part of the reason for this for schools with high starting APIs is that their Growth Target is so small that 60 percent of the Growth Target isn't a much smaller number than 100 percent of their Growth Target. But even for the lower scoring schools, where Growth Targets are much larger, the difference between requiring 60 percent and 80 percent is quite small. Thus, these results suggest that having a requirement that all subgroups meet a minimum standard of improvement has a substantial impact on the likelihood that a school that truly deserves a reward will get one—but if there is going to be such a rule, requiring that each of the subgroups makes 80 percent of the Growth Target does not wrongfully deprive many more schools than a requirement of 60 percent. Changing the requirement from 80 to 100 percent has more of an impact than changing from 60 to 80 percent.

Finally, note that all the analyses in this section assume that there are no validity issues to be concerned about, such as having the portion of students participating in the assessment changing dramatically from year to year. In California, parents have the option of excusing their children from the assessment. If a greater proportion of students are tested in one year than another, it is quite possible that the two years' worth of data will not be comparable. These analyses have assumed that such comparability exists and that fluctuations from one year to the next are solely a function of true change in the school combined with random error, but that no systematic changes (other than increased effectiveness in teaching) have occurred.

Appendix A

Calculating the Standard Error of the Academic Performance Index of a School

I. Principles for Calculating Standard Error of School Mean

Edward H Haertel (2000) has proposed detailed procedures to account for missing data while calculating the student-level API (Academic Performance Index) and school-level API, as well as the variance, standard deviation and standard error for the school's API. All the methodology used in this project came from his method. This document was revised from his original e-mail document (2000).

These procedures are necessary because a school's API will be equal to the average of its students' API only if all students have taken all tests. A school's API is not normally computed by taking the average of the student scores; it is computed by taking the average of scores for each test and then taking the weighted average of those values. To compute the standard error of the school mean, however, one must calculate and use student-level scores, since it is the variance of students (and the count of them) that determines the standard error of the mean for the school. Dr. Haertel's contribution was discovering a way to calculate student APIs so that their average would equal a school's API.

A. Calculating Student-level APIs

This section will describe two methods for calculating student-level APIs. The first method can be applied when there are no missing data; the second method gives the same results as the first method and is more complex, but can be used when missing data exist.

Without Missing Data

The normal way to calculate a student's API is as follows:

1. A student's percentile rank scores in each content area are converted by their ranges to the 200-500-700-875-1000 system of scale scores (also called "weighting factors;" see Table 1 and refer to the explanatory notes for the 1999 academic performance index report, California Department of Education),

Table 1: Conversion from National Percentile Rank to Weighting Factor

National Percentile Rank	Weighting Factor
$80\text{-}99^{th}$ NPR	1000
$60\text{-}79^{th}$ NPR	875
$40\text{-}59^{th}$ NPR	700
$20\text{-}39^{th}$ NPR	500
$1\text{-}19^{th}$ NPR	200

2. Take the average score, using the weights assigned to content areas. For example, a high school student with percentile ranks in reading, language, math, science, and social studies of, say, 25, 45, 50, 47, and 30 (see the second column in Table 2) would have a student 's API of .2*500+.2*700+.2*700+.2*700+.2*500=620 (see the last column in Table 2) with the 0.2 weighting value applied to each content area. The total weight across all content areas should be equal to 1 (see the last row on the fourth column in Table 2).

Table 2: An Example to Illustrate the Method for Computing the Student's API

Content Area	NPR	Mapping Score (S)	Weight (W)	S x W
Reading	25	500	.2	500x.2=100
Language	45	700	.2	700x.2=140
Mathematics	50	700	.2	700x.2=140
Science	47	700	.2	700x.2=140
Social Study	30	500	.2	500x.2=100
Sum			1.0	620

<u>With Missing Data</u>

The above calculations for student-level APIs assume all students tested had all 5 scores for Grades 9-11 (all 4 scores for Grades 2-8). On the other hand, when some students lack some content-area scores, the procedures for calculating student-level API's with missing data are more complicated. They are described below.

Assume that the numbers of valid scores in a school for the 5 content areas are N_{1s}, N_{2s}, N_{3s}, N_{4s}, and N_{5s}, where S represents a school (e.g., N_{1s}=the number of students tested in reading within School S.). And the W_1, W_2, W_3, W_4, and W_5 represent the content area weights (for high schools, $W_1 = W_2 = W_3 = W_4 = W_5 = .2$). Elementary and middle schools are handled the same way. Each score for a student in content area C is to be weighted by:

$$\frac{W_c}{N_{cs}} \qquad (1)$$

where W_c represents a weight for a content area, and
N_{cs} represents a valid sample size in a content area within a school.

Each student's weight in each of content areas within a school, as calculated in Equation 1, seems numerically small, but that doesn't affect the calculation of the school mean. (If you want sum of weights to come out closer to N of students tested in a school, you could multiply through by N.) A student's API is the weighted average of all his/her valid scores, with the weights calculated by equation (1). In addition, the total weight for the student is the sum of the weights for just those scores the student has.

For example, suppose there are 2 subject areas, with 3 valid scores for the first subject area and 4 for the second among five students in the school (see Table 3). Let's suppose for the sake of the calculation that the subject area weights were .6 and .4, respectively. The student-API method calculations would be as follows:

N_{1s}=3, N_{2s}=4,
W_1=.6, W_2=.4.

Therefore,

W_1/N_{1s}=.6/3= .2 (see the fourth column in Table 3),
W_2/N_{2s}=.4/4= .1 (see the fifth column in Table 3)

Student 1's API is (.2*200+.1*500)/(.2+.1)=300 (see the second row on the sixth column in Table3).

The remaining students' APIs, calculated similarly, are 500, 400, 700, and 500, respectively. The procedures illustrated in this section for computing the student's API have been used in this project.

B. Calculating School-level APIs

This section will describe two methods for the calculation of school-level APIs. The first is done without using student-level APIs, which is the method normally used by the California Department of Education to compute school's APIs. The second method, which uses student-level APIs, is used in this study only so that student-level variance within school, and therefore the standard error of school mean scores, can be calculated. What is important about the second method is that it produces *exactly* the same APIs as the first method.

Without Using Student-level APIs

The data in Table 3 will be used to demonstrate how the school-level API is calculated. First, school scores are computed by calculating the averages of all valid scores for areas 1 and 2, and then taking the weighted average of those, as follows:

Area 1: (200 + 500 + 500)/3 = 400 (see the last row on column 2 in Table 3),
Area 2: (500 + 200 + 700 + 500)/4 = 475 (see the last row on column 3 in Table 3),

School API = .6*400 + .4*475 = 430 (see the last row on column 4 in Table 3).

Again, note that these were the procedures used for calculating school-level APIs by the CDE (California Department of Education).

Using Student-level APIs

Using the same data provided in Table 3, another method to calculate the school-level APIs is accomplished by taking the weighted average of student-level APIs, using student total weights (sum of the weights across all content areas). In the example, these are .3, .2, .3, .1, and .1

respectively (see the sixth column in Table 3). Based on this method, the school-level API is calculated by:

(.3*300 + .2*500 + .3*400 + .1*700 + .1*500)/(.3+.2+.3+.1+.1) = 430.

Note that the API index derived by both methods is identical. It is noted that the summation across all students' total weights within a school should equal 1 except in a case where all students' scores in one of content areas are missing. This type of case is rare when school sample size is not too small (e.g., >50).

Table 3

An Example to Illustrate the Method for Calculating the Student-level API and School-level API, as well as the Effective Sample Size (EN) for a School's API

Pupil	Area 1 Score	Area 2 Score	W_{i1}	W_{i2}	Total Weight (TW_i)	API for Pupil (X_i)	Weight Score	TWX^2	TW^2
1	200	500	.2	.1	.1+.2=.3	(.2x200+.1x500)/.3=300	.3x300=90	27000	.09
2	500	M	.2	0	.2+0=.2	(.2x500+ 0)/.2=500	.2x500=100	50000	.04
3	500	200	.2	.1	.2+.1=.3	(.2x500+.1x200)/.3=400	.3x400=120	48000	.01
4	M	700	0	.1	0+.1=.1	(0+.1x700)/.1=700	.1x700=70	49000	.01
5	M	500	0	.1	0+.1=.1	(0+.1x500)/.1=500	.1x500=50	25000	.01
Sum	1200	1900	.6	.4	1	N/A	API= 430	199000	.24
Note	Mean= 400	Mean= 475	API=(400x.6)+(475x.4)=430						EN= 1/.24= 4.16

Note: "M" represents a missing value

C. Calculating Effective Sample Size

When the student-level APIs are calculated by the weighting procedure, a design effect adjustment for the unequal weighting can be applied in order to obtain an "effective sample size." If any students have missing data, the effective sample size is less than the actual sample size. The effective sample size is the square of the sum of the student weights divided by the sum of the squares of the student weights. For the hypothetical case (see the last column in Table 3), this would be:

$$EN_s = \frac{1}{\sqrt{(0.3)^2 + (0.2)^2 + (0.1)^2 + (0.1)^2}} = \frac{1}{.24} = 4.167$$

Note that this effective sample size is less than the actual sample size of 5. The design effect is 5/4.166667 = 1.2 .

D. Calculating the Standard Deviation of Student-level APIs within a School

Calculating the variance of the students' APIs is done by weighting them in the same way as was used in calculating the mean. More specifically, the formula for computing the variance for the APIs within a school is:

$$V(API_s) = \sigma^2 API_s = \frac{[\sum_i^{Ns} X_{is}^2 \cdot TW_{is}]}{\sum_i^{Ns} TW_{is}} - [(\sum_i^{Ns} TW_{is}) \cdot \mu_s^2)] \qquad (2)$$

where X_{is} represents the i[th] student 's API at school s (refer to the 7[th] column in Table 3),

TW_{is} represents the i[th] student 's total weight across all content areas at School S (refer to the 6[th] column in Table 3). As indicated, the summation of this total weight across all students within a school, in general, equals 1.

μ_s represents the weighted mean of API across all students' APIs within a school.

Ns is the number of students who are classified valid cases within a school.

The standard deviation is the square root of the variance, calculated from Equation 2.

E. Calculating the Standard Error of the Mean of the API for a School

The standard error (SE) of the Academic Performance Index (API) for a school is calculated by:

$$SE_{API_s} = \frac{\sqrt{\sigma^2_{API_s}}}{\sqrt{EN_s}} = \frac{\sigma_{API_s}}{\sqrt{EN_s}} \qquad (3)$$

where EN_s is the effective sample size for the school.

The calculation of the standard error of the mean for the previous hypothetical example (see Table 3) is shown below. The weighted mean, already obtained, is 430. The weighted sum of squares is:

$.3*300^2 + .2*500^2 + .3*400^2 + .1*700^2 + .1*500^2 = 199000$ (see the second from the last column in Table 3).

The variance (calculated with "N" not "N-1" for convenience) is:

199000/1 - (1*430*430) = 14100.

The standard deviation (SD) is:

$\sqrt{14100} = 118.743$.

The standard error of the mean (SEM) is found by dividing the SD by the square root of the effective sample size:

$$\frac{118.743}{\sqrt{4.167}} = 58.172.$$

One assumption made along the way was that students' APIs were identically distributed. In fact, a student's API based on 5 scores is more accurate than one based on fewer scores. However, in order to say *how much* more accurate is a student API based on more scores, it would be necessary to disentangle the sampling error vs. the measurement error. A student's transformed score in a single content area is composed of a student component plus a student by content area interaction, plus some combination of other miscellaneous error sources. If we average 5 scores versus 4 or 3, we reduce the variability associated with the subject area and the interaction terms, but do not reduce the variance associated with the student component. Because that component is dominant, however, the assumption that all student APIs are equally precise is adequate, and, in fact, slightly conservative. That is, the result of ignoring this effect will be to slightly overstate the standard error, and make it look as though APIs are slightly less accurate than they actually are.

II. Notes for Calculating the Standard Error of a School Mean

All the technical issues associated with how to compute the SE of API for a school were described in the previous section. An initial step in this study was to apply those principles to the student-level data file, calculate a mean API for each school, and compare those means to those published by the California Department of Education. As noted in the previous section, the Department of Education computed the mean API for each school by first computing the mean for each content area across students; we needed to compute the mean by first computing the mean score for each student. Those two methods should produce identical values for every school when the proper procedures are implemented. An initial set of calculations provided identical results for many schools, but not all; there were many data-handling issues that were not apparent to us until this comparison was made. This section describes our understanding of these rules for unusual circumstances. Application of these rules permitted us to duplicate the Department of Education's published APIs for every school

A. Rules Used for Identifying the Valid Content Scores for Each Pupil

The following rules, summarized by Brian Gong (2000), were used for identifying the valid cases for calculating the APIs.

1. District mobility represents the grade when the student first enrolled in the district. To be included, a student must have been in the district at least one year. This is operationally defined as: a student is not included in API calculations if the district mobility equals the grade the student is in. This means the district mobility can be lower or HIGHER than the grade (e.g., you will find some 7th graders who claim to have entered the district in 11th grade. These students are included). Students with blanks and multiple marks ("+") are also acceptable and included for API calculations.

2. In 7-12 and 9-12 districts, all students in grades 7 and 9, respectively, are not included in API calculations because they have not been in the district at least one year.

3. For a score to be included, the respective FLAG must be set to Y(es). Thus, if a student's Reading Flag is Y, and the Math Flag is N(o), then the student's reading score is included for API calculations, and the Math score is not.

B. Valid Samples (N_{cs}) for Content Area Score for Each School

Based on the above three rules for screening valid content area scores, the valid number (N_{cs}) of students taking a content area can be calculated by each school (see Equation 1). It is obvious that the number of N_{cs} for the content areas within a school could vary. The N_{cs} is used in Equation 1.

C. Total Number (N_s) of Students for Each School

A student is counted as one of the total number (N_s, see Equation 2) of students for the school if the student has at least one valid content score among all test score areas. These Ns are used in Equation 2.

D. SE of API for the bridge schools

The bridge schools have students at both primary and secondary grades. For this type of schools, a variance of API is calculated for the students in 2-8 and a variance of API is calculated for 9-11. These variances were pooled by the following formula to produce a single variance of API for the school.

$$\frac{(EN_{s2\_8}-1)\sigma^2_{s2\_8}+(EN_{s9\_11}-1)\sigma^2_{s9\_11}}{(EN_{s2\_8}-1)+(EN_{s9\_11}-1)} \qquad (4)$$

where EN_{s2_8} is the effective sample size from Grade 2 through Grade 8 students for school *s*,
EN_{s9_11} is the effective sample size from Grade 9 through Grade 11 students for school *s*,
$\sigma^2_{s2\_8}$ is the variance of API from Grade 2 through Grade 8 students for school *s*, and
$\sigma^2_{s9\_11}$ is the variance of API from Grade 9 through Grade 11 students for school *s*.

When the pooling variance is obtained, the SE of API for the bridge schools will be calculated using Equation 3. The effective size for this case is the denominator in Equation 4.[3]

[3] It would have been more proper to compute the pooled variance without subtracting 1 from each of the terms, since each of the variance terms was computed without subtracting 1. However, the difference that this change made was trivial and the effort it would have required to redo the study was considerable. Therefore, we choose to complete the study with these estimates.

III. Calculating the Probability that a School's Mean API Will Fall within a Given Decile

All the schools in the state were ranked according to their mean API and divided into 10 groups, with 10 percent of the schools in each group. Nine decile cut-scores were computed by averaging the mean APIs for the highest scoring school in the lower group and the lowest scoring school in the higher group.

The calculations above provided a mean and standard error of the mean for each school in the state. Each of the nine decile boundary points was transformed to a z-score, using the school's mean and standard error. The proportions of the school's normal distribution in each of the ten-decile ranges, which extend to negative and positive infinity at either end, are the probabilities that the school's true mean API falls in each of the given deciles.

The above procedures were carried by a mathematical function, CDF.normal(Cut_d, API,SE), supported by SPSS. The CDF.normal is used to calculate the cumulative probability that a normal random variance falls bellow Cut_d, where Cut_d represent the cutoff scores for the ten deciles, API is the school mean and SE is the corresponding SE of a school's API. Once the cumulative probabilities for the locations of the ten cutoff scores were obtained, the probability of the school's normal distribution in each of the ten decile ranges can computed as:

P(decile 1)= CDF.normal(Cut_1, API,SE),
P(decile 2)= [CDF.normal(Cut_2, API,SE)] – [CDF.normal(Cut_1, API,SE)],
P(decile 3)= [CDF.normal(Cut_3, API,SE)] – [CDF.normal(Cut_2, API,SE)],
…… with similar procedures for the rest of deciles.

The above procedures were carried out separately for the elementary, middle and high schools.

For the computing the probabilities that a school would be classified within each of the ten comparable schools deciles, the same procedures were used except that the cutoff scores for the ten deciles for each school are different and were computed separately using each school's 100 comparable schools.