
Attachment O
Issue Paper:
The chances of false negatives:
Failing AYP because of random processes

Altogether, there are many pairs of rates that must be evaluated each year in the process of determining California's schools' achievement of the goals mandated by the No Child Left Behind (NCLB) Act of 2001. Some of the rates pertain to proficiency measured school-wide in the state's 8,000-plus schools, but the majority pertain to the 10 subgroups of children within these schools who must be monitored. While the law requires that progress be determined for all significant subgroups of children, it also requires that the statistics used to describe this progress be reliable. Reliability is related to the possibility that random errors will influence the appearance of progress. The occurrence of such errors is dependent in large part on the numbers of observations upon which statistics are based. When such errors result in the judgment of failure, but the event they are related to has been in truth a success, the judgment is called a "false negative". This paper presents an analysis of the rates of false negatives to be expected for different subgroup sizes in the evaluation of adequate yearly progress for California's schools.

The focus here is on the rates of student proficiency in ELA and mathematics in the year 2003. The NCLB law requires schools achieve 100% proficiency rates for students by the year 2014 in these subjects, with rates increasing from present levels until then. The median proficiency rates for elementary and middle schools overall in 2002 was 0.28 for ELA and 0.30 for mathematics. The median is the mid-point value in a listing of proficiency rates that has been rank-ordered from smallest to largest, and can be referred to as the 50th percentile. The target rates for schools at the outset of the 12-year development period were 0.136 for ELA and 0.160 for mathematics, the proficiency rates that defined the 20th percentiles for schools in these subjects.

While expectations of chance-failures for schools having rates of proficiency that are above median levels cannot be ruled out, they are low; so attention has been restricted here to schools with true proficiency rates that are above the minimum 20th and below the

median, 50th percentile values. True score rates in equal steps between these values were modeled by means of Monte Carlo simulations, producing samples of 2000 cases at a time based on them and tabulating the frequencies at which generated proficiency rates were below the targets. The results, for samples sizes of 30, 50, and 100 students, are presented below in the Table. For purposes of simplification, all students are assumed to have a probability of scoring proficient equal to the school's overall proficiency rate.

Simulation of False Negatives in the Attainment of an ELA Target of 13.6% and a Mathematics Target of 16.0% Based on 2000 Replications					
p[Actual Proficiency] for Mathematics	p[Actual Proficiency] for ELA				
	0.14	0.17	0.21	0.25	0.28 (Median)
For N = 30					
0.16	0.79	0.68	0.60	0.53	0.52
0.20	0.73	0.58	0.40	0.34	0.33
0.23	0.66	0.48	0.30	0.23	0.18
0.27	0.63	0.45	0.27	0.17	0.10
0.30 (Median)	0.62	0.39	0.24	0.13	0.07
For N = 50					
0.16	0.71	0.58	0.47	0.45	0.43
0.20	0.59	0.39	0.27	0.23	0.23
0.23	0.51	0.28	0.16	0.10	0.09
0.27	0.49	0.24	0.10	0.06	0.03
0.30 (Median)	0.46	0.24	0.08	0.03	0.01
For N = 100					
0.16	0.74	0.55	0.48	0.46	0.47
0.20	0.59	0.29	0.18	0.14	0.14
0.23	0.53	0.20	0.06	0.03	0.03
0.27	0.51	0.15	0.03	0.01	0.00
0.30 (Median)	0.49	0.17	0.03	0.00	0.00

As the table shows, the maximal risk for false negatives occurs when either or both of the true rates of proficiency are close to the target values – that is, when the school's ELA proficiency rate is near 14% and its mathematics rate is near 16%. Risk declines as the true values of each increase. The shaded areas in the bottom-right of the table, for subgroup sizes of 50 and 100, highlight risks that are low enough, at less than 5%, that the chances of false-negatives would be small. As can be seen, there are few of

these shaded areas. For the smallest sample size, 30, the false negative risk remains at a fairly substantial level (e.g., greater than 20%) for all but the largest true score values.

It is important to emphasize that the false negative rate for a school is significantly higher than what is shown in the table. That is because of the cumulative nature of AYP decisions: the average school will have two to four significant subgroups in addition to its school-wide group. Even for a minimum group size of 100, many schools between the 20th and 50th percentiles of the statewide distribution will face a significant risk of a false negative AYP determination.