API's error margin leaves a lot to chance

Mathematical imprecision could lead to inaccurate interpretations.

August 11, 2002

By RONALD CAMPBELL
The Orange County Register

Like all statistics, California's Academic Performance Index is imprecise.

It contains a margin of error, meaning the score is considered reliable within a certain number of points above or below the reported score. Unlike most published statistics, the API's margin of error never was openly disclosed.

The Register used a handful of technical reports, including some independent reports recommended by California Department of Education experts, to determine the size and consequences of the API's statistical error.

There are two reasons why a statistical error exists in the API. The first and most important is sampling error - the students tested differ from school to school and year to year. The second is measurement error, related to the way questions are asked -- for example, the role guessing plays on multiple-choice questions. The smaller the school, the greater the error.

The Register relied primarily on an October 2000 study by Richard Hill of the National Center for the Improvement of Educational Assessment, which he prepared for the San Francisco-based Stuart Foundation. The Hill study was the most comprehensive look at the issue of statistical error in the API. Testing experts at the California Department of Education cited Hill's work in interviews with the Register and agreed with his conclusions about the size of the API's statistical error. Hill calculated separate errors by school type (elementary, middle and high school), number of students and relative performance on the API.

The newspaper also reviewed a July 2002 study by David Rogosa of Stanford University, the California Department of Education's top consultant on API issues. The department posted his report on its Web site July 3 -- the agency's first public acknowledgement that the API has a built-in statistical error.

Both men calculated the "standard error," the range above or below a school's reported API in which its true API would be found about two-thirds of the time -- a 68 percent confidence interval.

The Register reported errors at the 95 percent confidence interval. This is the benchmark used by the College Board, which publishes the Scholastic Aptitude Test.

To get from the standard error to the 95 percent confidence interval, the standard error is multiplied by 1.96.

The Register applied Hill's formulas to each of the 7,300 schools participating in the API and checked the steps in our calculations with him. California Department of Education testing experts have not done that but have agreed that as a rule, the margin of error is about 20 points, the median value under the Hill formula for California schools.

Six testing experts reviewed the paper's methods.

Comparing school scores from year to year required an additional step because there is a margin of error for each year's score.

At Hill's suggestion, the Register multiplied the error for one year by 1.3 to get the error of the difference between the two years.

Suppose an elementary school with 900 students reported APIs of 620 in 2000 and 640 in 2001. How do you know if the score really increased?

The error for each year is 16.3, and the error for a year-to-year comparison is 1.3 times that, or 21.2. The next part is simple math: Take the 2000 API (620), add 21.2 and you get 641.

That's how much the score could have risen by chance alone - and it's more than the reported increase.

Bottom line: Chance, not improved academics, might be behind that 20-point increase.