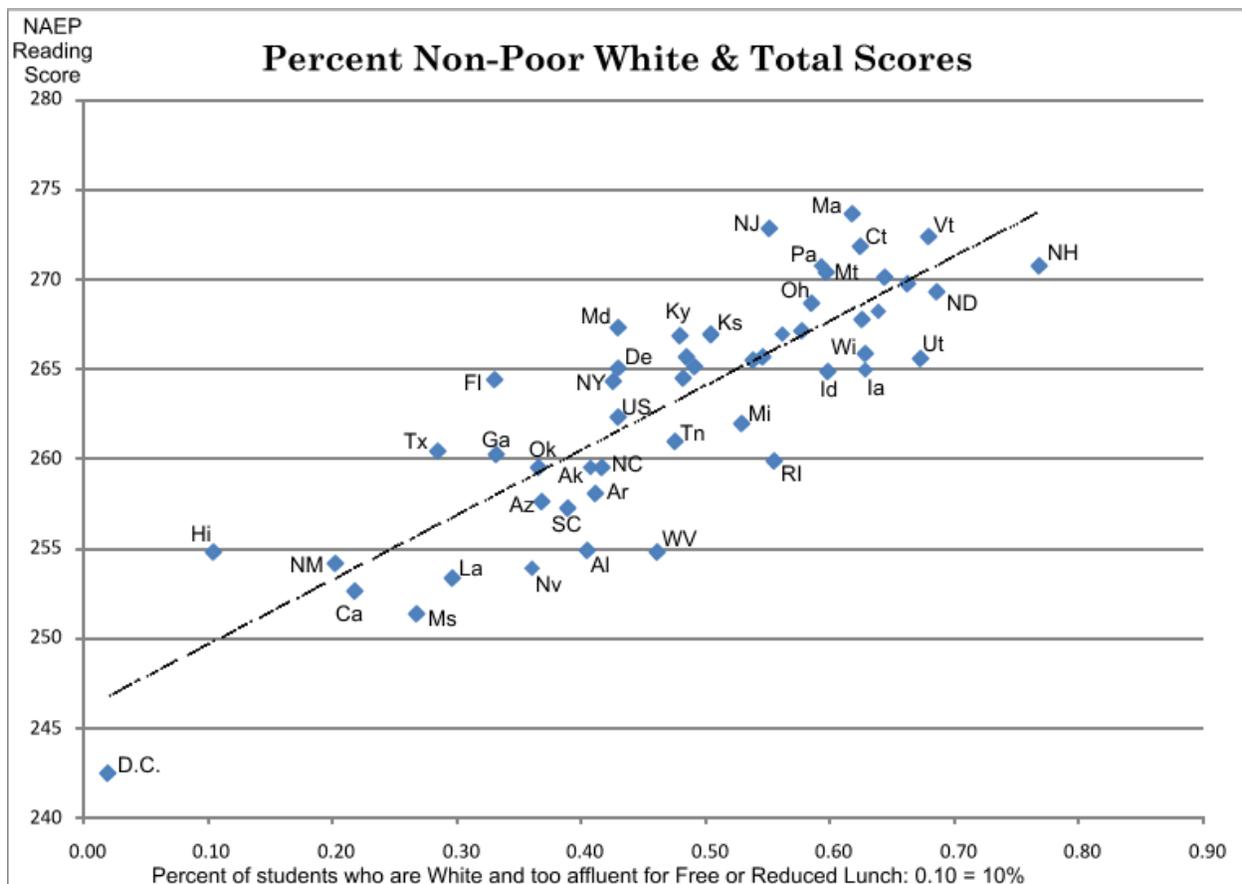


## Demographics And NAEP Eighth-Grade Reading Scores

By Michael T. Martin, Research Analyst, Arizona School Boards Association

The National Assessment of Educational Progress (NAEP) test of eighth-grade reading, administered in 2009, reaffirms that student socioeconomic status (SES) heavily influences state test scores. The test ranked Arizona 41 out of the 51 states and the District of Columbia with a score of 258 which was significantly below the national average of 262. However, that ranking is mostly due to the demographics of Arizona in comparison to other states, rather than differences in education or learning.

For example, Arizona's 2009 8<sup>th</sup> grade NAEP reading score could be closely predicted simply on the percentage of students in the state who are white and not poor enough to qualify for the federal free and reduced lunch program. The NAEP does not make any adjustments for differences in populations in each state. States with mostly non-poor White students are compared to states with mostly poor minority students. The rankings of the states therefore mostly represent their individual population demographics rather than their educational practices. How much is "mostly"?



Source: NAEP Data Explorer data for public schools using ethnic and Free and Reduced Lunch variables  
X axis data equals percent of students who are White multiplied by percent of White students ineligible for FRL

The above plot of state NAEP eighth-grade reading scores and the percentage of students in the state who are white and not poor enough to qualify for the federal free and reduced lunch program shows that Arizona's score lies very close to the trendline of all states. The correlation has a Pearson R of 85.2 percent, and thus the Pearson R-squared, known as the coefficient of determination, is 72.5 percent.

The coefficient of determination measures the amount of the variation in one variable, such as test scores, that is determined (often cited as "explained") by variation in another variable. In other words, 72.5 percent of the variation in the state NAEP eighth-grade reading scores was entirely accounted for by the variation in the percentage of non-poor White students in each state, a demographic variable.

All of the other factors that could possibly affect state NAEP reading scores, such as teacher competence, curriculum, number of English Language Learners, and other differences such as mother's education can, combined, only account for the remaining about one-fourth of the differences in the state test scores. Some other variables tend to be highly correlated with being non-poor, thus when you measure the percent of non-poor White students, you are also likely measuring smaller class sizes, greater per-pupil spending, and higher teacher salaries that attract better teachers.

In other words, the official NAEP eighth-grade reading test score that Arizona received is almost exactly the score that would be predicted just from the percentage of students in Arizona who are White and not poor enough to qualify for the federal Free and Reduced Lunch program. This "prediction" is the regression line of the same variables for every other state in the nation. Very few states have scores that differ significantly from this prediction, and that difference may be from differences in their educational programs, or differences in their social services (such as the Massachusetts universal health program) that affect their socioeconomic status, which is what achievement tests primarily measure.

What is true of the NAEP test is true of all achievement tests: one should be very reticent to make decisions based on test scores as measures of student knowledge. Those test scores more likely reflect the socioeconomic demographics of the student population tested. On the other hand, differences in state scores may also reflect programs that influence the socioeconomic status of students, such as healthcare and early childhood education, rather than the students' classroom experiences.

Experts in achievement testing repeatedly caution that using test scores to evaluate schools or teachers is a misuse of the tests. Harvard University Professor Daniel Koretz, a national expert on testing, wrote in his 2008 book "Measuring Up" that there is "a single principle" that should guide the use of tests: "Don't treat 'her score on the test' as a synonym for 'what she has learned.'"

Indeed, the National Research Council in March, 2001, released a report on its proposed evaluation of the Washington, D.C., school system in which they cautioned:

For this discussion, it is perhaps most important to underscore that most tests are not designed to support inferences about related questions, such as how well students were taught, what effects their teachers had on their learning, why students in some schools or classrooms succeed while those in similar schools and classrooms do not, whether conditions in the schools have improved as a result of a policy change, or what policy makers should do to solidify gains or

reverse declines. Answering those sorts of questions requires more and different kinds of evidence than test scores.

Writing about demographic circumstances in Washington, D.C., the National Research Council cautioned: “naïve aggregate comparison of test scores among race-ethnic groups in the District should be interpreted critically and cautiously. Thus, analysts need to carefully consider student backgrounds when comparing average scores, for example, by disaggregating by socioeconomic background.”

Arizona’s student population, many of whom are recent Hispanic immigrants and many more who live on the border with Mexico where Spanish is the primary language, plus Arizona’s large Native American reservations that seek to retain their traditional culture and language, represent a different population than that of New Hampshire, Vermont, North Dakota and Utah where the tested population is largely non-poor White students. Their test scores will reflect those differences.

Still, it is quite phenomenal that a single characteristic of the student population in each state accounts for such a large proportion of the variation in test scores. In essence, this means that the NAEP test, although portrayed as measuring reading skill, is primarily measuring student status. This is not a new revelation. This has been a known characteristic of achievement testing for decades.

In 1966 the “Coleman Report” requested by the U.S. Congress found the primary predictor of test scores is the socioeconomic status (SES) of those taking the test. It was often illogically interpreted at the time to mean that schools had little effect on student achievement, as if the tests themselves were sacrosanct measures of learning. Despite teachers repeatedly denying the validity of test results for students that they knew intimately from day-to-day experience, the existence of a number from a test score has continued to be interpreted as a measure of student knowledge.

Yet this scientific knowledge of the limitations of achievement test scores has not affected the practice of characterizing NAEP tests, and other achievement tests, as a valid “Report Card” on academic proficiency. The NAEP testing program is highly regarded. It is difficult to fault the tests as poorly constructed or misadministered. The NAEP tests are the best that can be implemented. Yet they founder on the same reality that has been repeatedly demonstrated since the Coleman Report: They primarily measure the socioeconomic status (SES) of the students rather than what they purport to measure.

Copyright March 25, 2011