Friday, November 1, 2002

Student Data Merits Scrutiny

By M. Kim Johnson and Walt Murfin
Coalition for Excellence In Science and Math Education

Over the last several years, there have been a number of letters and opinion articles from public officials, citizens and business leaders concerning education reform. A recent letter in the Journal from the Greater Albuquerque Chamber of Commerce insisted that academic expectations be raised. We applaud the involvement of the business community to raise New Mexico's educational performance. However, we are saddened at the lack of basic understanding demonstrated by many of these people — especially after most have been shown how educational measures should be interpreted.

To be fair, many do understand. But many do not and plunge ahead anyway, increasing the probability that real attempts at positive change will yield counterproductive results. The following describes what data measures used to make academic performance judgments really mean.

All children do not have equal ability. Some do exceptionally well. Some do not do well. Most are between these extremes. If the scores of all the students in a school are arranged in order from lowest to highest, the score in the middle is the median. If all the scores are added together and divided by the number of students, the result is the average, or mean. The median and average are usually close to each other, but are seldom exactly the same.

Standardized tests, like the TerraNova used in New Mexico, are first administered to a large sample of students (the "norming sample"). Test developers try hard to make the sample representative of the nation's school population. The average score of the sample is often called a "national average." Actually, it is really the average of a group intended to be representative of the nation's student population. The sample is usually a good match, but might differ in some particulars.

The TerraNova is called a norm-referenced test because each student's performance is compared to the performance of the norming sample. One student's or school's performance can be compared to another's and to the national standard. This is an important value of norm-referenced tests. They tell us the relative standing of schools and districts and their standing relative to the national sample, and do so quite reliably.

If a student is at the 40th percentile, it means 40 percent of students in the norming sample scored the same as, or lower than that student. At the 50th percentile, half scored lower and half scored higher. This has absolutely nothing to do with the percent of correct answers! Suppose three children take a test with 10 questions — one gets two correct, one gets three and one gets four. The median, or 50th percentile, is the middle score of 2, 3, and 4 — 3 out of 10, or 30 percent correct. Now, suppose another three children take a different test with 10 questions, and get 7, 8 and 9 correct. The middle score is 8, so the 50th percentile is 80 percent correct. Which children are smarter? Unless we know something about the relative difficulty of both tests, it's impossible to say. Perhaps the first test was extremely difficult and the second was very easy. It seems some business leaders and others fail to understand that percentiles have nothing whatever to do with the "passing grades" they remember from their school

days.

Scores are also given as grade equivalents. Grade equivalents do not show how children ought to perform, but how the average child does perform. A grade equivalent of 5.6 indicates performance equivalent to the average of fifth-grade students in the sixth month of the school year. It is not a criterion that all fifth-graders ought to be able to pass. If a school performs right at average, half the students will perform below the expected grade equivalent and half will perform above. Alarmists sometimes trumpet that "half of our children are reading below grade level!" What it really means is that they are about average. The statement "our schools must be doing badly because a third of the children are performing below grade level" is actually cause for congratulation, not alarm. It means two-thirds are performing above average. We should always strive to raise the average, but all children cannot be above average!

Given these definitions and numbers, one should question other people's conclusions until it is clear they understand what they're saying. Be skeptical, especially if it seems someone is deliberately misusing the data.

If you blindly accept the strong opinions of uninformed people, and actions result (legislation, for example), then it is likely that more harm than good will be done. For example, over the past several years a sizeable majority of New Mexico's students have been in districts whose median scores were considerably higher than the national average. However, many other school districts had much lower scores, and those are the ones that need help.

So, what did the Greater Albuquerque Chamber of Commerce letter really say and mean? It asked, "When did 70 percent come to mean that achievement was exemplary?"

It appears that the chamber has confused "percent answered correctly" with percentile. The impact of the chamber's requirement (as stated in the letter) of setting standards such that "meets standards at the very least mean our children are achieving at the 70th percentile" can be simply illustrated. (Note that the chamber's letter appears to use percent and percentile interchangeably.)

If New Mexico were to do what the chamber asks, then roughly 70 percent of the schools could be placed on probation. Even worse, this standard implies that 70 percent of the schools in the national sample would be probationary. Of course, this is totally absurd.

How would raising the academic achievement percentile bar for normed tests to some unrealistic number help extremely low-scoring schools? It would not. It simply would divert needed resources to solve a problem that could be better solved by targeting underperforming schools.

And what about the schools that are doing very well? They are already near the top; a drastic change to the entire school system will almost certainly not bring them higher and has a good chance of harming   them.

By all means, let us seek to improve academic performance, both overall and, in particular, for the lower performers. But first, understand the data that tell us where we need to improve, rather than misusing   the   data   and   looking   for   people   to   blame.

M. Kim Johnson is past president of the Coalition for Excellence in Science and Math Education and president-elect of the New Mexico Academy of Science. Walt Murfin is a statistician and analyst for CESE and a member of the New Mexico Academy of Science.