

RESEARCH BRIEFS

Brief #11

Fall 2011

The Problem with Ranked Comparisons

Large scale assessment programs allow us to make comparison—of students, of districts, schools, states, and even countries. Having said this, there is little evidence to show that these types of comparisons serve any useful purpose. For example, knowing that students in an affluent suburb have higher average test scores than students attending a poor, inner city school in another district is not particularly helpful.

Comparisons based on rank

Ranking schools, districts, states, or countries based on test scores is a misleading use of test results. In 2005, Bert Stoneberg from the Idaho Department of Education warned against this practice.¹ What he says about the ranking of states also can be applied to schools, districts, or nations.

As an example of how rankings can misrepresent, consider the following: If we were to have a race among the ten fastest runners in the world and then rank them from fastest to slowest, one person would finish first and another last. Yet, there would be less than a fraction of a second difference between first and last place.

The author cited two problems associated with ranked comparisons of states:

1. **The average score for each state is an estimate of what students in that state actually know and are able to do.** If we had a perfect measure—with no random error—we could determine the students' "true" score. However, this is an impossible goal because we can never eliminate all sources of error (e.g., students not understanding the questions that are asked; a student is not feeling well at the time of testing; test directions are not understood; a student accidentally marks the wrong answer choice; a student had a bad experience at home the prior evening that makes it difficult for her to concentrate, etc.).

RESEARCH BRIEFS

Because there is error on all tests, those who report the results always should include a confidence interval (similar to a margin of error on a public opinion poll). Confidence intervals are reported in terms of probabilities, typically at the 95 percent level.

In 2003, the average scale score for Wisconsin on the 8th grade reading test administered by the National Assessment of Educational Progress (NAEP) was 266.47.² The 95 percent confidence interval is +/- 2.69 (ranging from 263.98 to 268.98). This tells us that there is a 95 percent probability that the “true” score for Wisconsin’s students would fall within this range.

Wisconsin was ranked 17th on this test, but given the confidence interval, it could have ranked as high as 2 or as low as 29.

2. When states are compared, the differences among their average scale scores are often very small, yet the ranks make those differences appear greater than they actually are. The table below shows the average scale score for each of eight Midwestern states on the 2003 NAEP reading test. Also shown is the reported rank, the 95 percent confidence interval, and the possible range of ranks (from the highest to lowest possible).

Among the eight states, the difference in test scores is only 5.59 points, yet the reported ranks range from 4th in the nation (South Dakota) to 27th nationally (Michigan).

When the 95 percent confidence interval is used (as it should be), South Dakota could have ranked as high as 1st nationally or as low as 18th. Michigan could have ranked as high as 6th nationally or as low as 36th.

[see table next page]

RESEARCH BRIEFS

2003 NAEP State Results for Eighth Grade Reading National Assessment of Educational Progress (Scale Scores Shown)³

State	Ave. Scale Score	Rank	95 percent Confidence Interval	Range of Ranks	
				Highest	Lowest
Iowa	267.50	12	265.952 - 269.048	2	29
Michigan	264.38	27	260.774 - 267.986	6	36
Minnesota	267.71	10	265.593 - 269.827	2	30
Illinois	266.41	18	264.430 - 268.390	5	31
North Dakota	269.73	6	268.201 - 271.259	1	19
Ohio	266.57	16	263.983 - 269.157	2	31
Wisconsin	266.47	17	263.981 - 268.959	2	31
South Dakota	269.97	4	268.461 - 271.479	1	18

The next time you see ranked comparisons, look for the confidence interval. If there is none, then the rankings are probably distorting the actual results.

Written by: Russ Allen PhD and Jeffrey Leverich PhD

Endnotes

¹ "Please Don't Use NAEP Scores to Rank Order the 50 States." Retrieved on August 26, 2010:
<http://pareonline.net/pdf/v10n9.pdf>.

² There are numerous examples of scales. For example, weight is measured using a scale based on pounds or kilograms; height is measured using a scale based on inches, feet, meters, centimeters, etc.

³ The author of this table reports scores to the third decimal point. Using three decimal places suggests a level of precision that is not justified. We didn't round it off to a single decimal point because we wanted to be true to the author. However, one decimal point would be fine.