Practical Assessment, Research & Evaluation
A peer-reviewed electronic journal.

Volume 10 Number 9, August 2005 ISSN 1531-7714

Please Don't Use NAEP Scores to Rank Order the 50 States

Bert D. Stoneberg
Idaho Department of Education

Results from the National Assessment of Educational Progress (NAEP) and other large scale tests are often reported as rank ordered lists showing mean values for each of the 50 states. Using data from the 2003 State NAEP Assessment, this paper examines the standard errors associated with State NAEP scores and explains why the use of rank order statistics is inappropriate. An alternate approach anchored to a given state is offered.

Results from the 2003 and earlier administrations of the National Assessment of Educational Progress (NAEP) often have been reported as rank ordered lists of the 50 states. The reports come from a variety of sources including government agencies (Bourque, Champagne and Crissman, 1997), corporate and foundation "think tanks" (Grissmer, Flanagan, Kawata and Williamson, 2000), and news outlets (Roberts, 2003). This paper was born in the hope that this practice might be avoided for the 2005 and later administrations. It will explain why the use of rank order statistics is inappropriate and will recommend a better way to report how a state performed relative to the other states.

Rank order reporting rests on two flawed assumptions about NAEP scores. The first is that each state's score is absolute. NAEP scores, however, are only estimates of state performance determined through a statistical sampling of students and subject matter. Not all students in a state are tested, and the students who are assessed don't do the whole test. In Idaho, for example, only one in seven eighth grade students were in the 2003 reading sample, and each of them completed only 50 minutes of the 200 minute reading test. NAEP used this sample to estimate a score of 264.44 for Idaho with a standard error of 0.89. The National Center for Education Statistics (NCES) always publishes both the estimated scale score and its standard error. NAEP scores are not absolute.

The second assumption is that even the smallest difference between two NAEP scores justifies ranking one state higher than another. A ranking of states based on estimates from the 2003 eighth grade reading assessment, for example, lists Idaho as 26th and Michigan as 27th. For sure, Idaho's score was 264.44 while Michigan's was only 264.38. The difference between the two states was six one-hundredths (0.06) of a point. On the other hand, the standard errors were 0.89 for Idaho and 1.84 for Michigan. The combined measurement error (2.73 points) was more than 45 times larger than the difference between their scores. There was simply no justification in the 2003 eighth grade NAEP

reading results to rank Idaho above Michigan. The difference was too small to have meaning.

When standard errors are included in the analysis of NAEP scores a different understanding of each state's ranking among the states emerges. One technique, for example, uses the standard error to define confidence intervals for the state estimates from which a "range of ranks" for each state can be identified. Table 1 presents a range of ranks for each state from the 2003 eighth grade reading assessment using the 95 percent confidence interval. The information in columns labeled *State*, *Estimated Score* and *Standard Error (SE)* were downloaded from the web using the NAEP Data Tool (U.S. Department of Education, 2004a). NCES does not include sample sizes in the NAEP Data Tool. NCES, however, did list the number of students tested by state in its participation (U.S. Department of Education, 2003). The sample size for each state is listed in the column labeled *N* on Table 1, but it can be misleading. NAEP assigns different weights to state sample sizes to obtain valid inferences about the state populations of interest. A state's estimated scores and standard errors are based on its weighted sample, not on the number of students tested.

The data in the Table 1 columns labeled *Rank*, *Confidence Interval (95%)*, R*ange of Ranks High* and *Low* were generated for this paper as follows.

Rank: A "1" was assigned to state with highest estimate, a "2" to the state with the next highest, and on down to "50" for the state with the lowest estimate. The assignments did not account for standard error.

Table 1: Rank Order Scores for NAEP Reading 2003, All Students, Grade 8

State	N	Average	Rank	Standard Error	Confidence Interval (95 percent)	Range of Ranks	
						High	Low
Alabama	2,585	253.17	46	1.51	250.210 - 256.130	37	50
Alaska	2,498	256.41	42	1.10	254.254 - 258.566	35	46
Arizona	2,625	255.32	43	1.36	252.654 - 257.986	35	50
Arkansas	2,575	258.00	39	1.29	255.472 - 260.528	31	46
California	5,510	251.01	50	1.28	248.501 - 253.519	43	50
Colorado	2,710	267.59	11	1.20	265.238 - 269.942	2	30
Connecticut	2,725	267.22	14	1.08	265.103 - 269.337	2	30
Delaware	2,496	264.53	24	0.74	263.080 - 265.980	8	33
Florida	2,443	257.30	41	1.33	254.693 - 259.907	31	46
Georgia	4,219	257.71	40	1.14	255.476 - 259.944	31	46
Hawaii	2,768	251.28	49	0.87	249.575 - 252.985	43	50
Idaho	2,642	264.44	26	0.89	262.696 - 266.184	8	33
Illinois	4,039	266.41	18	1.01	264.430 - 268.390	5	31
Indiana	2,642	264.83	23	1.04	262.792 - 266.868	7	33
Iowa	2,823	267.50	12	0.79	265.952 - 269.048	2	29
Kansas	2,916	266.01	21	1.48	263.109 - 268.911	3	33
Kentucky	2,800	266.19	20	1.25	263.740 - 268.640	4	31
Louisiana	2,305	253.45	45	1.58	250.353 - 256.547	37	50
Maine	2,882	268.32	7	0.98	266.399 - 270.241	2	26
Maryland	2,449	261.60	33	1.45	258.758 - 264.442	17	41
Massachusetts	3,770	272.91	1	0.96	271.028 - 274.792	1	6
Michigan	2,625	264.38	27	1.84	260.774 - 267.986	6	36
Minnesota	2,605	267.71	10	1.08	265.593 - 269.827	2	30
Mississippi	2,694	255.01	44	1.38	252.305 - 257.715	35	50
Missouri	2,651	267.36	13	1.01	265.380 - 269.340	2	30
Montana	2,581	269.83	5	1.04	267.792 - 271.868	1	23
Nebraska	2,476	266.31	19	0.91	264.526 - 268.094	6	30

Table 1: Rank Order Scores for NAEP Reading 2003, All Students, Grade 8

State	N	Average	Rank	Standard Error	Confidence Interval (95 percent)	Range of Ranks High	Range of Ranks Low
Nevada	2,651	252.31	47	0.82	250.703 - 253.917	43	50
New Hampshire	2,868	270.73	2	0.93	268.907 - 272.553	1	17
New Jersey	2,866	267.79	9	1.21	265.418 - 270.162	2	30
New Mexico	3,061	251.60	48	0.87	249.895 - 253.305	43	50
New York	3,424	265.33	22	1.33	262.723 - 267.937	6	33
North Carolina	4,057	261.71	32	0.98	259.789 - 263.631	21	41
North Dakota	2,612	269.73	6	0.78	268.201 - 271.259	1	19
Ohio	3,414	266.57	16	1.32	263.983 - 269.157	2	31
Oklahoma	2,839	261.72	31	0.95	259.858 - 263.582	21	41
Oregon	2,561	264.03	30	1.23	261.619 - 266.441	7	34
Pennsylvania	2,792	264.27	29	1.18	261.957 - 266.583	7	34
Rhode Island	2,643	260.88	34	0.71	259.488 - 262.272	28	41
South Carolina	2,446	258.09	38	1.26	255.620 - 260.560	31	46
South Dakota	2,770	269.97	4	0.77	268.461 - 271.479	1	18
Tennessee	2,655	258.11	37	1.17	255.817 - 260.403	31	46
Texas	4,378	258.78	36	1.12	256.585 - 260.975	30	44
Utah	2,732	264.30	28	0.84	262.654 - 265.946	10	33
Vermont	2,682	270.52	3	0.82	268.913 - 272.127	1	16
Virginia	2,733	268.00	8	1.05	265.942 - 270.058	2	30
Washington	2,625	264.49	25	0.88	262.765 - 266.215	8	33
West Virginia	2,234	259.56	35	1.00	257.600 - 261.520	30	44
Wisconsin	2,566	266.47	17	1.27	263.981 - 268.959	2	31
Wyoming	2,763	267.00	15	0.53	265.961 - 268.039	6	29

Confidence Interval (95%): The lower limit of the 95 percent confidence interval for each state is its score minus 1.96 times its standard error. The upper limit is its score plus 1.96 times its standard error. Idaho's score was 264.44 with a standard error of 0.89, so its lower limit is 264.44 – (1.96 x 0.89) = 262.696. Idaho's upper limit was 264.44 + (1.96 x 0.89) = 266.184. Sometimes sampling error works for a state, sometimes against it. NAEP draws a student sample to represent a state for the assessment; and all scores and standard errors for the state were based on students in that sample. If it were possible to draw a large number of representative samples from a state, then a state could have a large number of estimated scores. A confidence interval identifies a percentage of such scores that would fall between its lower and upper limits, based on the given sample. Thus, if Idaho students could have been repeatedly sampled with replacement for the 2003 eighth grade reading assessment, then based on the given sample one might expect that 95 percent of Idaho's estimates would have fallen somewhere between 262.696 and 266.184.

Range of Ranks High/Low: A state's highest rank is identified when the upper limit of its confidence interval is compared with the lower limits of the other 49 states. A state's lowest rank is found when the lower limit of its confidence interval is compared with the upper limits of the other 49 states.

Below are three pairs of rank order statements side-by-side that describe results from the 2003 NAEP eighth grade reading assessment based on the data in Table 1. The first makes use of the estimated score only; the second brings into play the estimated score and the standard error.

Rank Using Estimate Only	*Rank Using Estimate and Standard Error*
Massachusetts ranked first among the states.	Six states shared a claim to the top rank: MA, MT, NH, ND, SD, and VT.
Idaho ranked 26th among the states.	Idaho's rank among the states was somewhere between 8th and 33rd place.
California ranked last among the states.	Eight states flirted with the lowest rank: AL, AR, CA, HI, LA, MS, NV, and NM.

These statements describe state-level rank order results based on all students in a state sample. NAEP also provides state-level estimated average scores and standard errors disaggregated by gender, ethnicity, poverty, location, and student eligibility for certain educational programs (e.g., educationally disadvantaged students, students with disabilities, and limited English proficient students). Due to their smaller size, the estimated scores for these subgroups typically have larger standard errors.

The discussion thus far has focused on using rank order statistics based on average scale scores and their associated standard errors of measurement. NCES, however, also estimates and reports the percentages of students scoring at each of the NAEP achievement levels (i.e., *Basic*, *Proficient* and *Advanced*) and their standard errors. Rank order lists based on percentage scores, whether for all students or disaggregated groups, have the same difficulties as those based on average scale scores.

The public typically prefers reports where test scores and state ranks are precise and absolute, but NAEP data lack the precision to support such claims. Some are likely be wary of "loosy goosy" reporting based on a range of ranks because there are no clear winners or losers. It would not take long before the public begins mistakenly to question the validity of the test rather than the usefulness of rank order statistics for reporting the test results. For reporting NAEP results, it is best to "just say no" to rank ordering the 50 states.

A WAY TO COMPARE YOUR STATE TO OTHER STATES

Figure 1 illustrates an alternative way to compare one state with other states using NAEP estimated scores. The process, which accounts for measurement error and small differences between state scores, uses three categories rather than 50 ranks. After a state is picked as the focal state (e.g., Idaho for this example), its estimated score is compared via a *t*-test with the scores from each of the other 49 other states. States with scores that are statistically significantly higher than the focal state form one group. States with scores that are statistically significantly lower than the focal group are in a second group. The remaining states with scores than are not statistically different from the focal group make up a third group. The results can be reported graphically as in Figure 1 or in narrative form as follows.

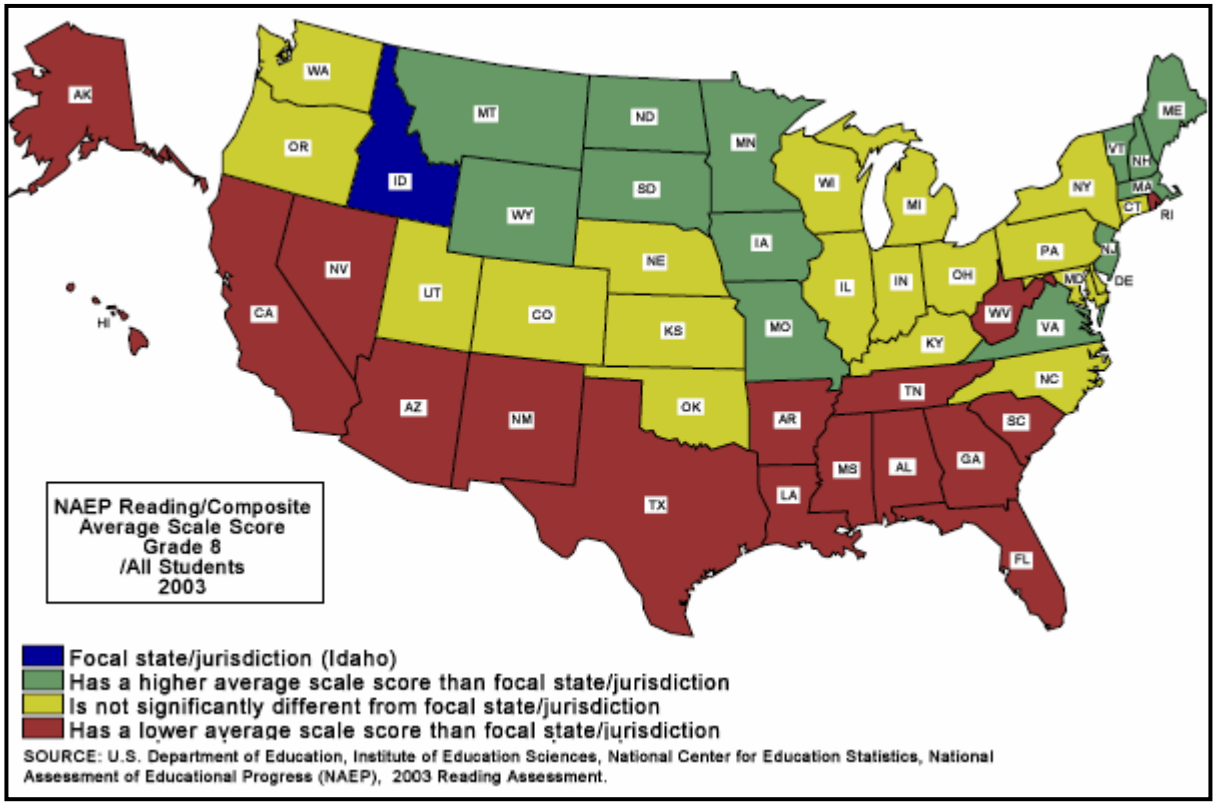

Figure 1. States with NAEP estimated scores that were significantly higher than Idaho, not different from Idaho, or significantly lower than Idaho on the 2003 reading composite for all grade 8 students.

On the NAEP 2003 eight grade reading assessment for all students:

- ✓ 13 states scored higher than Idaho (IA, ME, MA, MN, MO, MT, NH, NJ, ND, SD, VT, VA and WY),
- ✓ 17 states scored lower than Idaho (AL, AK, AZ, AR, CA, FL, GA, HI, LA, MS, NV, NM, RI, SC, TN, TX and WV), and
- ✓ 19 states were not different from Idaho (CO, CT, DE, IL, IN, KS, KY, MD, MI, NE, NY, NC, OH, OK, OR, PA, UT, WA and WI).

The best thing about this way of making cross-state comparisons is that NCES has already run all the *t*-tests and prepared the graphs for all of the states. Graphs showing state NAEP results for all students since 1990 on the reading, mathematics, writing, and science assessments are available on the web from the NCES "State Profile" page (U.S. Department of Education, 2004b):

1. Go to http://nces.ed.gov/nationsreportcard/states/
2. Click on a state, and scroll down to the history section
3. Look under "Graphics" column in the history section
4. Click on any of link for *Cross-State Comparison Maps ○ Scale Scores*
5. The check marks in the top table indicate graphs that are available for each subject and year. Click on any check to call up the graph

Enjoy using the graphs. The hard work here has been done for you, and it is statistically defensible. It is also possible to use other on-line NCES tools such as the NAEP Data Explorer to generate similar graphs for all students and for disaggregated groups using either average scale scores or achievement level percentage scores. Most important, though, you will find that your

audience will correctly understand and appreciate this kind of cross state comparisons.

DISCUSSION

NAEP has generated fourth- and eighth grade state level results for reading, mathematics, science and writing at irregular intervals since 1990. It has scrupulously estimated scores and standard errors for each state assessment regardless of grade, subject or year. The data used for this rank order analysis and the alternate cross-state comparison method were only from one grade (eighth) on one subject (reading) at one point in time (2003). Nonetheless, the need to take error into account when using NAEP to rank order the states has been clearly illustrated, and one method of doing this has been presented. It would be helpful if a system of analysis existed that integrates NAEP estimated scores and standard errors from multiple measures across grades or subjects or time or a combination thereof to compare performance levels among the states. Someday, maybe.

Nothing in this paper should be construed as a criticism of the National Assessment of Educational Progress. If only all large-scale assessments were as well designed and executed. The criticism is directed at the failure to take error into consideration when developing rank order lists of the states using NAEP scores. Unfortunately this failure is not unique to NAEP, but seems common to many sets of rankings.

REFERENCES

Bourque, M.L., Champagne, A.B., & Crissman, S. (1997, October). *1996 science performance standards: achievement results for the nation and the states*. Washington, D.C.: U.S. Department of Education, National Assessment Governing Board. Retrieved June 11, 2005, from http://www.nagb.org/pubs/1996science/index.html

Grissmer, D.W., Flanagan, A., Kawata, J., & Williamson, S. (2002). *Improving student achievement: what state NAEP test scores tell us*. Santa Monica, CA: Rand Corporation. Retrieved June 11, 2005, from http://www.rand.org/publications/MR/MR924/

Roberts, B. (2003, November 11). Idaho kids good, but not great in math, reading: Idaho Hispanics rank lower in skills nationwide. *The Idaho Statesman,* pp. 1A, 6A.

U.S. Department of Education, Institute for Education Sciences, National Center for Education Statistics. (2003). *School and student participation rates, grade 8, public schools: By state, 2003*. Retrieved July 16, 2005, from http://www.nces.ed.gov/nationsreportcard/naepdata/reading/stateparticgr8.asp

U.S. Department of Education, Institute for Education Sciences, National Center for Education Statistics. (2004a). *NAEP Data, 2004* [Data file]. Available from the National Center for Education Statistics web site, http://www.nces.ed.gov/nationsreportcard/naepdata/

U.S. Department of Education, Institute for Education Sciences, National Center for Education Statistics. (2004b). *State Profiles, 2004* [Data file]. Available from the National Center for Education Statistics web site, http://www.nces.ed.gov/nationsreportcard/states/

Citation

Stoneberg, Bert D. (2005). Please Don't Use NAEP Scores to Rank Order the 50 States. *Practical Assessment Research & Evaluation*, 10(9). Available online: http://pareonline.net/getvn.asp?v=10&n=9

Author

Dr. Bert D. Stoneberg is the NAEP State Coordinator for the Idaho Department of Education. Correspondence regarding this article should be sent to him at the Idaho Department of Education, P.O. Box 83720, Boise, ID 83720-0027 or BDStoneberg @sde.idaho.gov.