School of Education, University of Colorado at Boulder
Boulder, CO 80309-0249
Telephone: 802-383-0058

NEPC@colorado.edu
http://nepc.colorado.edu

Responding By Not Responding: A Reply to Paul E. Peterson

Jeremy Kilpatrick, University of Georgia

My recent review of a study by Hanushek, Peterson, and Woessmann[1] has prompted a response by Paul E. Peterson[2] on the *EdNext Blog* that seems contrived to avoid responding to the review I actually wrote. Because readers of the blog may not have seen the review (find it at http://nepc.colorado.edu/thinktank/review-us-math),[3] this reply attempts to set some of the record straight.

The Claim that Correlations Mean Comparability

A major point I make in the review is that NAEP and PISA (Program for International Student Assessment) mathematics scores, like TIMSS (Trends in International Mathematics and Science Study) and PISA mathematics scores, should not be put on the same scale, because the tests measure different aspects of mathematical proficiency. I point out that correlation does not imply equivalence. Peterson argues that a high between-country correlation (.93) between PISA mathematics and TIMSS mathematics makes them comparable. But high between-country correlations do not at all imply that the same aspects of learning are being measured; instead, the measures are quite likely to be linked to similar learning assets (e.g., countries' wealth, prior educational level, and investment in schools). If high correlations were all it took to ensure comparability, then the PISA scores for mathematical literacy, scientific literacy, and reading literacy would all be comparable because their correlations at the country level in PISA are above .90.[4] Would anyone want to claim that the PISA tests of scientific and reading literacy could therefore be used as measures of mathematical literacy?

Deceptive, Misleading Exaggerations

One of Peterson's most egregious charges is the following: "To deny that the U. S. ranks number 31 in math performance requires the claim that PISA is a 'deceptive', 'misleading' test that 'exaggerates small differences' something Kilpatrick wants to imply but never dares to argue." My comments using those words were not at all about the PISA mathematics test, which I have worked on and hold in reasonably high regard. Rather, I used those words to describe the study by Peterson and his colleagues. What I said was that to rank the percentages of advanced-level students in countries, states, and districts on the same scale was deceptive, was misleading, and exaggerated small differences on the outer reaches of a score distribution.

Ranking the Insignificantly Different

Although in the review I nowhere deny that the U.S. ranked 31st in the percentage of its 15-year-olds scoring at or above a particular cut-score on the PISA 2006 mathematics test, I do criticize the practice of ranking in the Peterson study. As I wrote in the report, "When standard errors are accounted for, it becomes clear how meaningless it is to provide a ranked list of countries and states; no difference is found between a state like Minnesota ranked 20th in the list and a country like Sweden ranked as 24th." The ranking of states on NAEP performance has long been criticized.[5] Peterson claims that even if percentages reaching NAEP and PISA cut-scores cannot be placed on the same scale, the ranking of states would remain unchanged. Massachusetts and Minnesota would still come first, and Mississippi last. "No other state's ranking vis a vis one another changes by so much as a rabbit's hair." That is correct—the state and district rankings are based entirely on NAEP data and gain nothing from being linked to PISA—but it misses the point. Those rankings might not change, but not all are meaningful to begin with. For example, Table 1 in the report shows Maryland, South Carolina, Wisconsin, Ohio, New Hampshire, South Dakota, Colorado, New York, and Texas ranked from 10th to 18th in the percentages of students at the advanced level in NAEP 2005. The percentages range from 6.2 to 6.8, and the standard errors range from 0.5 to 0.8. All the standard error bands overlap considerably, so the nine states are not significantly different in their rankings.

Mixing and Matching Scales

In addition to the problems raised by ranking, there is a very serious reason for not putting percentages of "high achievers" on NAEP 2005 on the same scale as corresponding percentages of "high achievers" on PISA 2006: the NAEP cohort consisted of U.S. eighth graders, whereas the PISA cohort consisted of 15-year-olds. Peterson considers that a "trivial technical concern," arguing that I fail to show that mathematics "performance fluctuates noticeably from one student cohort to the next." His argument is that "the percentage of U.S. 8th graders performing at the advanced level [on NAEP] changes only very modestly over time." But that contention is irrelevant. The key question is instead whether there is any good reason to believe that if PISA had sampled U.S. 15-year-olds in 2006 as thoroughly as NAEP sampled eighth graders in 2005, the percentage of PISA high achievers in each state and urban district would have been identical to the percentage of NAEP high achievers. As Gary Phillips has pointed out,

> NAEP and TIMSS can be linked statistically because they have administered to equivalent representative samples of students, in the same subject (mathematics), in the same year (spring of 2007), and in the same grades (Grades 4 and 8). No similar basis exists to create a linkage between PISA and NAEP.[6]

When one cohort is sampled by grade and the other by age, putting their performance on the same scale is hardly a trivial technical matter.

No Class

Peterson claims it is not significant "that the international students who took the PISA were not always from the same graduating class as the U. S. students taking the National Assessment of Educational Progress (NAEP)." Does it really not matter at all that a figure in the report that gives each country's percentages of high achievers is labeled "Class of 2009," when the percentages of PISA 15-year-olds who were actually in the 10th grade in 2006 were 71% for the U.S., 0% for Finland, 28% for Germany, 100% for Japan, and 1% for the U.K.? It is certainly misleading to claim that one is portraying the performance of 12th graders internationally in 2009 when such widely varying numbers of students in the PISA sample were actually in that grade and when unknown numbers have dropped out of school in the preceding years.

Thankful Absence of Recommendations

Finally, Peterson says that I "fault" him and his colleagues for failing to make policy recommendations. That is not true. Although his report was prepared under the auspices of a program on education policy and a journal that concerns education policy, I found no fault in the absence of recommendations. In fact, the final comment in my review was one of relief, not fault finding: "It is just as well that the report offers no recommendations for 'policy changes that might foster excellence' (p. 24). Any such recommendations might have been as untenable as the study itself."

Notes and References

[1] Hanushek, E.A., Peterson, P.E., & Woessmann, L. (2010, November). *U.S. math performance in global perspective: How well does each state do at producing high-achieving students?* (PEPG Report No. 10–19). Cambridge, MA: Harvard's Program on Education Policy and Governance and *Education Next*. Retrieved December 2, 2010, from http://www.hks.harvard.edu/pepg/PDF/Papers/PEPG10-19_HanushekPetersonWoessmann.pdf.

[2] Peterson, P. E. (2011, January 12). No matter how hard you try, you cannot deny U. S. math performance is terrible. *The EdNext Blog*. Retrieved January 12, 2011, from http://educationnext.org/no-matter-how-hard-you-try-you-cannot-deny-u-s-math-performance-is-terrible.

[3] Kilpatrick, J. (2011). *Review of "U.S. Math Performance in Global Perspective: How Well Does Each State Do at Producing High-Achieving Students?"* Boulder, CO: National Education Policy Center.

[4] Organization for Economic Co-operation and Development. (2009). *PISA data analysis manual: SPSS* (2nd ed.). Paris: Author, 26. Retrieved December 20, 2010, from http://www.oecd.org/document/31/0,3746,en_2649_35845621_42592351_1_1_1_1,00.html.

[5] Stoneberg, B. D. (2005). Please don't use NAEP scores to rank order the 50 states. *Practical Assessment Research & Evaluation, 10*(9). Retrieved January 17, 2011, from http://pareonline.net/getvn.asp?v=10&n=9

[6] Phillips, G.W. (2009, June). *The second derivative: International benchmarks in mathematics for U.S. states and school districts*. Washington, DC: American Institutes for Research, 4. Retrieved December 20, 2010, from http://www.air.org/files/International_Benchmarks1.pdf.

*The **Think Twice** think tank review project (http://thinktankreview.org), a project of the National Education Policy Center, provides the public, policy makers, and the press with timely, academically sound, reviews of selected think tank publications. The project is made possible in part by the generous support of the Great Lakes Center for Education Research and Practice.*

*The mission of the **National Education Policy Center** is to produce and disseminate high-quality, peer-reviewed research to inform education policy discussions. We are guided by the belief that the democratic governance of public education is strengthened when policies are based on sound evidence. For more information on NEPC, please visit http://nepc.colorado.edu/.*